ISSN 2710-0766
DOI 10.31891/CSIT

THE INTERNATIONAL SCIENTIFIC JOURNAL

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

No 2-2022

МІЖНАРОДНИЙ НАУКОВИЙ ЖУРНАЛ

КОМП'ЮТЕРНІ СИСТЕМИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

2022

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC JOURNAL

Published since 2020 year *Four time a year*

Khmelnitskyi, 2022, № 2 (7)

Establishers: **Khmelnytskyi National University** (Ukraine)
Associated establisher: **Institute of Information Technologies** (Slovakia)
National Library og Ukraine named after V.I.Vernadsky http://nbuv.gov.ua/j-tit/csit

The journal is included in scientometric databases:
Index Copernicus https://journals.indexcopernicus.com/search/details?id=69998&lang=en
Google Scholar https://scholar.google.com.ua/citations?hl=uk&user=HW1XpMsAAAAJ
CrossRef http://doi.org/10.31891/CSIT

Recommended for publication by the decision of the Academic Council of Khmelnytskyi National University, protocol № 18 from 30.06.2022

Editorial board address: International scientific journal "Computer Systems and Information Technologies", Khmelnytskyi National University, Institutska str. 11, Khmelnytskyi, 29016, Ukraine
☎ **(0382) 67-51-08**
e-mail: csit.khnu@gmail.com
web: http://csitjournal.khmnu.edu.ua/index.php/csit
http://lib.khnu.km.ua/csit_khnu.htm

Registered by the Ministry of Justice of Ukraine
Certificate of state registration of the print media
Series KB № 24924-14864PR dated 12.07.2021

КОМП'ЮТЕРНІ СИСТЕМИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВИЙ ЖУРНАЛ

Засновано в 2020 р. *Виходить 4 рази на рік*

Хмельницький, 2022, № 2 (7)

Засновник і видавець: **Хмельницький національний університет** (Україна)
Асоційований співзасновник: **Інститут інформаційних технологій** (Словаччина)

Наукова бібліотека України ім. В.І. Вернадського http://nbuv.gov.ua/j-tit/csit

Журнал включено до наукометричних баз:
Index Copernicus https://journals.indexcopernicus.com/search/details?id=69998&lang=en
Google Scholar https://scholar.google.com.ua/citations?hl=uk&user=HW1XpMsAAAAJ
CrossRef http://doi.org/10.31891/CSIT

Рекомендовано до друку рішенням Вченої ради Хмельницького національного університету, протокол № 18 від 30.06.2022

Адреса редакції: **Україна, 29016,**
м. Хмельницький, вул. Інститутська, 11,
Хмельницький національний університет
редакція журналу "Комп'ютерні системи та інформаційні технології"
☎ **(0382) 67-51-08**
e-mail: csit.khnu@gmail.com
web: http://csitjournal.khmnu.edu.ua/index.php/csit
http://lib.khnu.km.ua/csit_khnu.htm

Зареєстровано Міністерством юстиції України
Свідоцтво про державну реєстрацію друкованого засобу масової інформації
Серія КВ № 24924-14864ПР від 12 липня 2021 року

CONTENTS

SERGII OREKHOV, HENADII MALYHON
METHOD OF SOLVING THE PROBLEM OF SITUATIONAL MANAGEMENT OF THE SEMANTIC KERNEL OF WEB CONTENT 6

BOHDAN SAVENKO, ANTONINA KASHTALIAN
A METHOD FOR DETERMINING THE EFFECTIVENESS OF A DISTRIBUTED SYSTEM FOR DETECTING ABNORMAL MANIFESTATIONS 14

ANDRII KOPP, DMYTRO ORLOVSKYI, DORUKHAN ERSOYLEYEN
AN APPROACH TO APPLICATIONS ARCHITECTURE MODELS ANALYSIS 23

BOHDAN HUNKO
HARDWARE AND SOFTWARE SYSTEM OF LIGHT VISUALIZATION OF SOUND SIGNALS 33

VASYL YATSKIV, VOLODYMYR BODNAROVSKYI
RESEARCH OF METHODS OF ENERGY EFFICIENCY MANAGEMENT IN THE "SMART HOUSE" SYSTEM 42

VOLODYMYR KHOROSHKO, VADYM KUDINOV, MARIIA KAPUSTIAN
EVALUATION OF QUALITY INDICATORS OF FUNCTIONING CYBER PROTECTION MANAGEMENT SYSTEMS OF INFORMATION SYSTEMS 47

LUBOMYR SIKORA, NATALIIA LYSA, OLGA FEDEVYCH, ROSTYSLAV TKACHUK
LASER AND INFORMATION TECHNOLOGIES FOR CONTROLLING DYNAMIC DISPLACEMENTS SPATIAL STRUCTURES OF OBJECTS UNDER THE INFLUENCE OF ACTIVE MAN-MADE AND NATURAL RISK FACTORS FOR ACCIDENTS 57

VIRA SHENDRYK, YULIIA PARFENENKO,
VALENTYN MAIKOVSKYI, DENYS YURCHENKO, SERGII SHENDRYK
SUBSYSTEM OF COLLECTION, STORAGE AND VISUALIZATION OF OPERATING DATA OF THE DECISION SUPPORT SYSTEM FOR MICROGRID MANAGEMENT 69

IRYNA VECHIRSKA, OLEG KOBYLIN, STEPAN PROKOPIEV,
ANNA VECHIRSKA, MAKSYM KUCHERENKO
BUILDING A LOGICAL NETWORK FOR SOLVING THE PROBLEM OF CAR RENTAL BY MEANS ALGEBRA OF FINITE PREDICATES 78

VASYL PRYIMAK, SVITLANA PRYIMA, OLGA HOLUBNYK
FUZZY TECHNOLOGIES IN THE MANAGEMENT OF SOCIO-ECONOMIC PROCESSES 88

YELYZAVETA HNATCHUK, VITALII BASHUK, DENYS KVASNITSKYI
RESEARCH OF METHODS AND MEANS OF ENSURING THE RELIABILITY OF A SPECIALIZED COMPUTER VOICE VEHICLE CONTROL SYSTEM 95

ANDRII POPOVYCH, VITALII YAKOVYNA
COVID-19 MORTALITY PREDICTION USING MACHINE LEARNING METHODS 104

YELYZAVETA HNATCHUK, ANITA BOIKO, ALINA HNATCHUK
RESEARCH OF METHODS OF SEAT DISTRIBUTION IN PUBLIC TRANSPORT 112

ЗМІСТ

СЕРГІЙ ОРЄХОВ, ГЕНАДІЙ МАЛИГОН
МЕТОД ВИРІШЕННЯ ЗАДАЧІ СИТУАЦІЙНОГО УПРАВЛІННЯ СЕМАНТИЧНИМ ЯДРОМ ВЕБ КОНТЕНТУ 6

БОГДАН САВЕНКО, АНТОНІНА КАШТАЛЬЯН
МЕТОД ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ РОЗПОДІЛЕНОЇ СИСТЕМИ ВИЯВЛЕННЯ АНОМАЛЬНИХ ПРОЯВІВ 14

АНДРІЙ КОПП, ДМИТРО ОРЛОВСЬКИЙ, ДОРУХАН ЕРСОЙЛЕЄН
ПІДХІД ДО АНАЛІЗУ МОДЕЛЕЙ АРХІТЕКТУРИ ЗАСТОСУНКІВ 23

БОГДАН ГУНЬКО
АПАРАТНО-ПРОГРАМНА СИСТЕМА СВІТЛОВОЇ ВІЗУАЛІЗАЦІЇ ЗВУКОВИХ СИГНАЛІВ 33

ВАСИЛЬ ЯЦКІВ, ВОЛОДИМИР БОДНАРОВСЬКИЙ
ДОСЛІДЖЕННЯ МЕТОДІВ УПРАВЛІННЯ ЕНЕРГОЕФЕКТИВНІСТЮ В СИСТЕМІ «РОЗУМНИЙ БУДИНОК» 42

ВОЛОДИМИР ХОРОШКО, ВАДИМ КУДІНОВ, МАРІЯ КАПУСТЯН
ОЦІНКА ПОКАЗНИКІВ ЯКОСТІ ФУНКЦІУНУВАННЯ СИСТЕМ УПРАВЛІННЯ КІБЕРЗАХИСТОМ ІНФОРМАЦІЙНИХ СИСТЕМ 47

ЛЮБОМИР СІКОРА, НАТАЛЯ ЛИСА, ОЛЬГА ФЕДЕВИЧ, РОСТИСЛАВ ТКАЧУК
ЛАЗЕРНІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ КОНТРОЛЮ ДИНАМІЧНИХ ЗМІЩЕНЬ ПРОСТОРОВИХ СТРУКТУР ОБ'ЄКТІВ ПІД ВПЛИВОМ АКТИВНИХ ТЕХНОГЕННИХ ТА ПРИРОДНИХ ФАКТОРІВ 57

ВІРА ШЕНДРИК, ЮЛІЯ ПАРФЕНЕНКО,
ВАЛЕНТИН МАЙКОВСЬКИЙ, ДЕНИС ЮРЧЕНКО, СЕРГІЙ ШЕНДРИК
ПІДСИСТЕМА ЗБОРУ, ЗБЕРІГАННЯ ТА ВІЗУАЛІЗАЦІЇ ОПЕРАТИВНИХ ДАНИХ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ УПРАВЛІННЯ МІКРОГІД 69

ІРИНА ВЕЧІРСЬКА, ОЛЕГ КОБИЛІН,
СТЕПАН ПРОКОП'ЄВ, АННА ВЕЧІРСЬКА, МАКСИМ КУЧЕРЕНКО
ПОБУДОВА ЛОГІЧНОЇ МЕРЕЖІ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧІ ОРЕНДИ АВТОМОБІЛІВ ЗАСОБАМИ АЛГЕБРИ СКІНЧЕННИХ ПРЕДИКАТІВ 78

ВАСИЛЬ ПРИЙМАК, СВІТЛАНА ПРИЙМА, ОЛЬГА ГОЛУБНИК
НЕЧІТКІ ТЕХНОЛОГІЇ В УПРАВЛІННІ СОЦІАЛЬНО-ЕКОНОМІЧНИМИ ПРОЦЕСАМИ 88

ЄЛИЗАВЕТА ГНАТЧУК, ВІТАЛІЙ БАШУК, ДЕНИС КВАСНІЦЬКИЙ
ДОСЛІДЖЕННЯ МЕТОДІВ ТА ЗАСОБІВ ЗАБЕЗПЕЧЕННЯ НАДІЙНОСТІ СПЕЦІАЛІЗОВАНОЇ КОМПЮТЕРНОЇ СИСТЕМИ ГОЛОСОВОГО КЕРУВАННЯ АВТОМОБІЛЕМ 95

АНДРІЙ ПОПОВИЧ, ВІТАЛІЙ ЯКОВИНА
ПРОГНОЗУВАННЯ СМЕРТНОСТІ ВІД COVID-19 МЕТОДАМИ МАШИННОГО НАВЧАННЯ 104

ЄЛИЗАВЕТА ГНАТЧУК, АНІТА БОЙКО, АЛІНА ГНАТЧУК
ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПОДІЛУ МІСЦЬ В ГРОМАДСЬКОМУ ТРАНСПОРТІ 112

https://doi.org/10.31891/csit-2022-2-1
UDC 004.9

Sergii OREKHOV, Henadii MALYHON
National Technical University "Kharkov Polytechnical Institute"

METHOD OF SOLVING THE PROBLEM OF SITUATIONAL MANAGEMENT OF THE SEMANTIC KERNEL OF WEB CONTENT

Research conducted over the past ten years in the field of search engine optimization on the Internet shows the creation of a new phenomenon - virtual promotion [1]. Its main goal is to increase the level of sales of goods through technologies that exist in virtual space. In terms of properties, virtual promotion is similar to a logistics channel, but with the characteristics of a marketing channel. That is, virtual promotion can be described as the establishment of a channel for the transmission of a special message from the company to a potential buyer of its goods or services. This article proposes to consider the problem of creating and managing such a message in cyberspace. To do this, it is proposed to solve the problem of situational management of the semantic kernel of web content. At the first stage the review of properties of a semantic kernel is carried out. This analysis allows us to formulate the problem of situational management of the semantic kernel. The main idea of this problem is that the current situation is a variant of the semantic kernel of web content. Then the paper proposes a mathematical model for estimating the current situation and choosing the target situation. Conditions are also proposed to determine the need for transition between situations. The article also describes the algorithmic support for the implementation of the method of solving the problem of situational management. The result of the algorithm is K-applicant, ie the semantic kernel that describes the class of needs, which was identified at the stage of clustering. The prospect for further research is the software implementation of this algorithm on the NodeJS platform.

Keywords: Virtual promotion of a product, situational management, semantic kernel, web content.

Сергій ОРЄХОВ, Генадій МАЛИГОН
Національний технічний університет „Харківський політехнічний інститут"

МЕТОД ВИРІШЕННЯ ЗАДАЧІ СИТУАЦІЙНОГО УПРАВЛІННЯ СЕМАНТИЧНИМ ЯДРОМ ВЕБ КОНТЕНТУ

Дослідження, що було проведено за останні десять років у сфері пошукової оптимізації в мережі Інтернет, свідчать про створення нового феномену – віртуальне просування [1]. Його основна ціль полягає у підвищенні рівня продажу товару завдяки технологіям, що існують у віртуальному простору. За властивостями віртуальне просування схоже з логістичним каналом, але з характеристиками, що має маркетинговий канал. Тобто віртуальне просування можна описати як встановлення каналу передачі спеціального повідомлення від підприємства до потенційного покупця його товару чи послуги. В статті запропоновано розглянути задачу створення та управління таким повідомленням у віртуальному просторі. Для цього пропонується вирішувати задачу ситуаційного управління семантичним ядром веб контенту. На першому етапі проведено огляд властивостей семантичного ядра. Цей аналіз дозволів сформулювати постановку задачі ситуаційного управління семантичним ядром. Головна ідея цієї задачі полягає в тому, що в якості поточної ситуації розглядається варіант семантичного ядра веб контенту. Тоді в роботі пропонується математична модель для оцінювання поточної ситуації та вибору цільової ситуації. Також пропонуються умови для визначення необхідності переходу між ситуаціями. Також в статті описано алгоритмічне забезпечення для реалізації методу вирішення задачі ситуаційного управління. Результатом виконання алгоритму є К-претендент, тобто семантичне ядро, що описує клас потреби, який було виявлено на етапі кластерізації. Перспективою для подальших досліджень є програмна реалізація даного алгоритму на платформі NodeJS.

Ключові слова: віртуальне просування продукту, ситуаційне управління, семантичне ядро, ВЕБ контент.

Introduction

The process of virtual promotion is similar to the logistics channel, where there are several concepts of work. From our point of view, the concept of "continuous replenishment" [1] is suitable for virtual promotion. This means that goods, and in our case knowledge about goods, should be concentrated in the main nodes of the logistics network. That is, it is necessary to place knowledge about the product or the need it covers on the Internet, where it will be read or downloaded by users - potential buyers of goods or services. In other words, you need to place knowledge about the product on the Internet and wait for the search engine database to be updated. A network of cross-references between network nodes is also formed to confirm both the originality of information about the product and the primary source of this data in the form of: a) a corporate website, b) a profile on a social network, or c) a profile on a trading platform.

The main idea of virtual promotion is that we need to form a channel and a special message in it. This message is called the semantic core. This approach is completely new and alternative. This is a new look at the problem of search engine optimization on the Internet [1].

Thus, understanding the physical content of virtual promotion means that a powerful promotion channel on the Internet is being created, called thematic search [2]. To create other channels, it is also necessary to unify the message so that it is distributed in the same content and format on the Internet with a link to the corporate website. In our work [3], for the first time, it was proposed to use as a message the semantic core of web content generated by a company that is interested in selling a given product or in satisfying a given need on its own using Internet technologies [3, 4].

This paper proposes an alternative approach to solving the problem of search engine optimization on the Internet. The essence of the approach is to change the semantic core of web content step by step in the promotion channel in order to improve the main values of WEB metrics. The step-by-step change of the core is based on the method of situational control [5].

Thus, the article proposes a new look at the problem of search engine optimization on the Internet by generating special web content and a strategy for changing it in the promotion channel.

Problem statement

In [5] the first attempt to solve the problem of search engine optimization on the Internet as a problem of situational management was described. According to the classical theory of situational management [6], we have the following statement of the problem:

$$S_i : E_{i1} \underset{U_i}{\Rightarrow} E_{i2} \ , \tag{1}$$

where S_i - the full situation, E_{i1}, E_{i2} - current situations, U_i - the rule of logical transformation from one current situation to another.

We will assume that the current situation corresponds to the current version of the semantic kernel. We have the following process that implements the marketing channel of virtual promotion - Figure 1. That is, we have the current situation. In which case should this kernel be replaced with a new one?

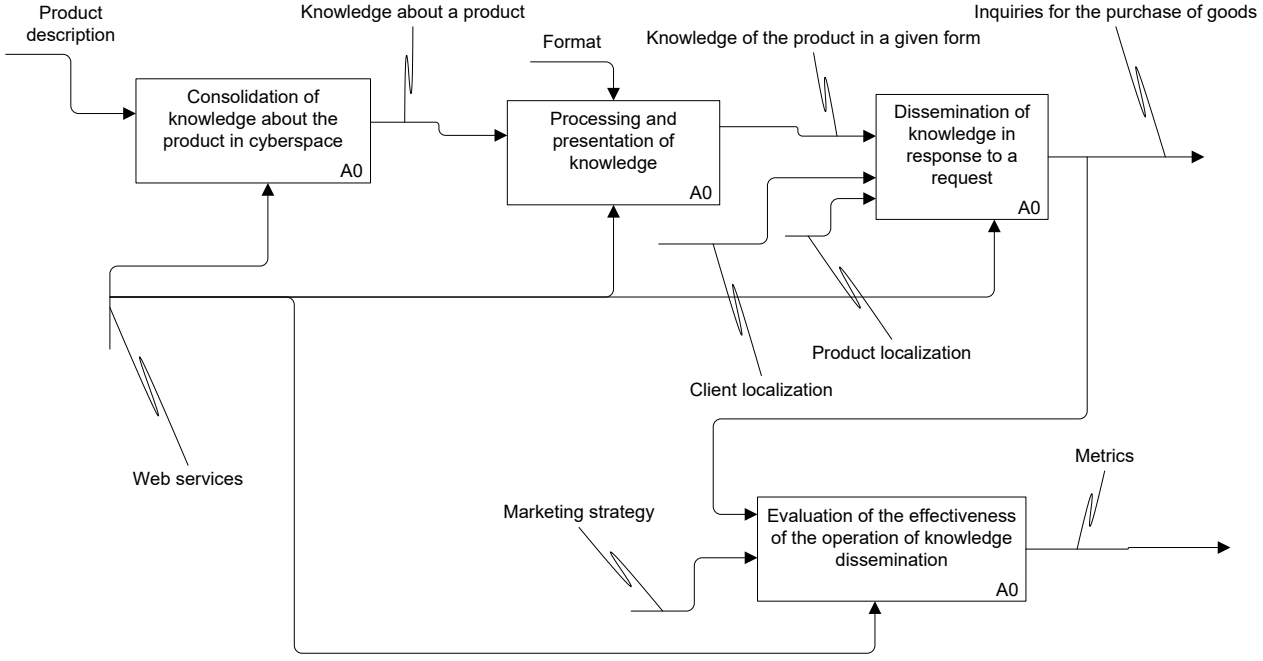

Fig. 1. Modern scheme of virtual promotion marketing channel

The first case. Whether there is a criterion of usefulness or a criterion of effectiveness of virtual promotion. This criterion can assess the profitability of sales of goods through a given channel of promotion on the Internet. Then you can evaluate the promotion complex consisting of the semantic kernel and the promotion channel. We will assume that such an estimate is presented in the form of a graph of the relationship between the value of profitability and time. If the value of the profitability of this complex decreases over a period of time, then such a semantic kernel must be changed. The first attempt to write such a criterion for assessing the efficiency of the channel and the kernel is presented in [7].

The second case. The main metrics for search engine optimization on the Internet [8] show a steady decline (or increase) over time. Such metrics are traffic, number of leads, bounce rate, conversion, number of repeat visits. These metrics belong to the class of WEB metrics. All these metrics show the conditional efficiency of the promotion channel with a given semantic kernel.

The third case. An enterprise decides to change product characteristics or marketing segment for your sale. In this way, the entire text complex *TD* is changed, and that needs to repeat the task of synthesizing the kernel again.

In the paper it is suggested to look at the first aspect of the formation of the efficiency criterion for the complex (the channel and the kernel), and also to improve the metrics that indicate the mental efficiency. We will

mean that, like $j=\overline{1,J}$ - the number of nodes in the channels of passing through the Internet, where the semantic kernel will be located. B_{tj} – budget, which can be accepted for the placement of the kernel in the *j*–th node for the period of time *t*. u_{tj} - the number of leads, which was taken for a period of time *t*. The period of time $t \in T$ may be 24 hours. We remind you that lead is the action of a potential buyer on a web resource that leads him to: 1) purchase on the web site, 2) pay for the product on the website or 3) enter your contact details at the special form for the remote purchase of the product in the office [9]. Then $u_{tj}=u_{tj}^1+u_{tj}^2+u_{tj}^3$. Our goal is to maximize the component u_{tj}^2. It is reasonable to [10-11] to modify the criterion of effectiveness for offensive

$$\sum_{t\in T}\sum_{j\in J}\frac{B_{tj}(tr_{tj})}{\left(u_{tj}^1(tr_{tj})+u_{tj}^2(tr_{tj})+u_{tj}^2(tr_{tj})\right)}\to \min, \tag{2}$$

where tr_{tj} is the traffic received by the *j*-th node for the time period *t*. The physical meaning of criterion (2) is to reduce the cost of attracting a buyer to purchase a product or service on the Internet. Then the index $u_{tj}=u_{tj}^1+u_{tj}^2+u_{tj}^3$ corresponds to the conversion of traffic into the appropriate form of lead. For us, the most interesting is the second type of lead, when the buyer pays for the goods or services directly online. We transform the formula (2):

$$\sum_{t\in T}\sum_{j\in J}\frac{1}{\frac{\left(u_{tj}^1(tr_{tj})+u_{tj}^2(tr_{tj})+u_{tj}^2(tr_{tj})\right)}{B_{tj}(tr_{tj})}}=\sum_{t\in T}\sum_{j\in J}\frac{1}{\frac{u_{tj}^1(tr_{tj})}{B_{tj}(tr_{tj})}+\frac{u_{tj}^2(tr_{tj})}{B_{tj}(tr_{tj})}+\frac{u_{tj}^3(tr_{tj})}{B_{tj}(tr_{tj})}}=,$$

$$\sum_{t\in T}\sum_{j\in J}\frac{1}{\left(O^1(tr_{tj})+O^2(tr_{tj})+O^2(tr_{tj})\right)}\to \min,$$

$$\text{or } \sum_{t\in T}\sum_{j\in J}P(tr_{tj})\to \min, \tag{3}$$

where $O^1(tr_{tj})$ is the payback of the channel on the first type of lead, which guarantees the placement of the semantic kernel in the *j*-th node for a period of time *t*. The higher the payback of the channel and the message, the lower the price of attracting a buyer $P(tr_{tj})$ to a given node for a given period of time.

Let's take as a basis the classic problem of linear programming to describe the mathematical model of the channel with the kernel from the standpoint of the cost of attraction. Then formula (3) has the following modification:

$$\sum_{t\in T}\sum_{j\in J}w_j P(tr_{tj})\to \min, \sum_{j\in J}w_j=1,$$

where w_j is the weighting factor that takes into account the importance of the *j*-th node in the promotion channel. The higher the payback of the channel and the message, the lower the price of attracting a buyer $P(tr_{tj})$ to a given node for a given period of time.

The first block of task conditions will mean that the traffic in the node is a positive value and an integer: $tr_{tj}\ge 0$, $tr_{tj}=\{0,1,2,...\}$.

For the promotion channel to be successful, it is necessary to enter the threshold of its profitability:

$$\sum_{t\in T}\sum_{j\in J}\frac{\mathrm{Pr}_{tj}}{P(tr_{tj})+Ex_{tj}}\ge\sum_{t\in T}R_t,$$

R_t - is the rate of return, which is set by the company for a period of time *t*. It shows the lowest value that guarantees the success of Internet promotion. Pr_{tj} - income received by the enterprise, if the buyer ordered a

product or service, reading the semantic kernel in the *j*-th node for a period of time *t*. Ex_{tj} - additional costs borne by the enterprise from the placement of the semantic kernel in the *j*-th node for a period of time *t*. $P(tr_{tj})$ - the price of attracting customers in the *j*-th node for a period of time *t*.
Then we have the following optimization problem:

$$\sum_{t\in T}\sum_{j\in J} w_j P(tr_{tj}) \to \min , \tag{4}$$

$$\sum_{t\in T}\sum_{j\in J} \frac{\mathrm{Pr}_{tj}}{P(tr_{tj}) + Ex_{tj}} \geq \sum_{t\in T} R_t , \tag{5}$$

$$\sum_{j\in J} w_j = 1 , \tag{6}$$

$$tr_{tj} \geq 0 , \; tr_{tj} = \{0,1,2,...\} . \tag{7}$$

The physical meaning of the problem (4)-(7) is that you need to estimate the minimum cost of attracting one buyer, having a given channel and semantic kernel. The problem with this task is that the traffic-dependent form of the function $P(tr_{tj})$ is unknown.

Then the problem of situational management in modern conditions is formulated as follows [6]. We denote the current situation $C = \langle K, F \rangle$, where *K* - the current version of the semantic kernel, and *F* - a function that expresses the evaluation of the efficiency of the channel and the kernel, i.e. the result of solving the problem (4)-(7). Then the full situation $S = \langle C, C_R \rangle$ is formed, as a combination of the current situation and the goal that the company sets for the promotion process. This goal primarily C_R applies to conditions (5), i.e. the rate of return on the channel and the kernel of the promotion. We will assume that the current situation belongs to the class $C \in Q'$. And the target situation of the class. Then you need to form the following management $U \in \Omega$ to achieve the target situation:

$$C \in Q' \underset{U\in\Omega}{\to} C_R \in Q'' \text{ or } (Q', Q'') \to U \in \Omega \tag{8}$$

where Ω - a set of control options, in other words, options for transforming the kernel or changing the semantic kernel to another. We believe that the structure of the promotion channel remains unchanged, because it is the subject of further research.

Thus, the problem of situational management of the semantic kernel has the following wording: we must choose $U \in \Omega$ the transformation of the semantic kernel of web content based on the classification of the current situation $C \in Q'$ and adequate assessment F of the promotion channel and the kernel as a whole, which should lead to a new target $C_R \in Q''$ (channel and kernel) by the marketing department of the enterprise.

So our task includes four subtasks: kernel synthesis (initial version), kernel evaluation - task (4)-(7), kernel classification and selection of a new class of kernel, ie changing the kernel to another taking into account the requirements of marketing services.

The advantages of this approach are significant. Because the problem under consideration is dynamic. A typical scheme for solving the problem of linear or nonlinear programming is as follows [12]. We believe that a mathematical model exists. Then you must first choose the basic parameters of the problem, for example, based on the Monte Carlo method or mathematical statistics. That is, you need to collect initial data and process them. Then, having a record of the problem and a standard solution method, the optimal Pareto alternative is sought. Finally, you can check the stability of the solution. In our case, this is not possible, because you can collect input data in the process of solving the problem. Therefore, the situational approach allows, so to speak, to adapt the probable solution of the problem in the process of its solution. This is similar to the task of opening a safe when you turn the knob and test the lock for opening.

Proposed approach

In the paper it is offered to solve a problem of situational management according to the following UML scheme [13-14] - figure 2.

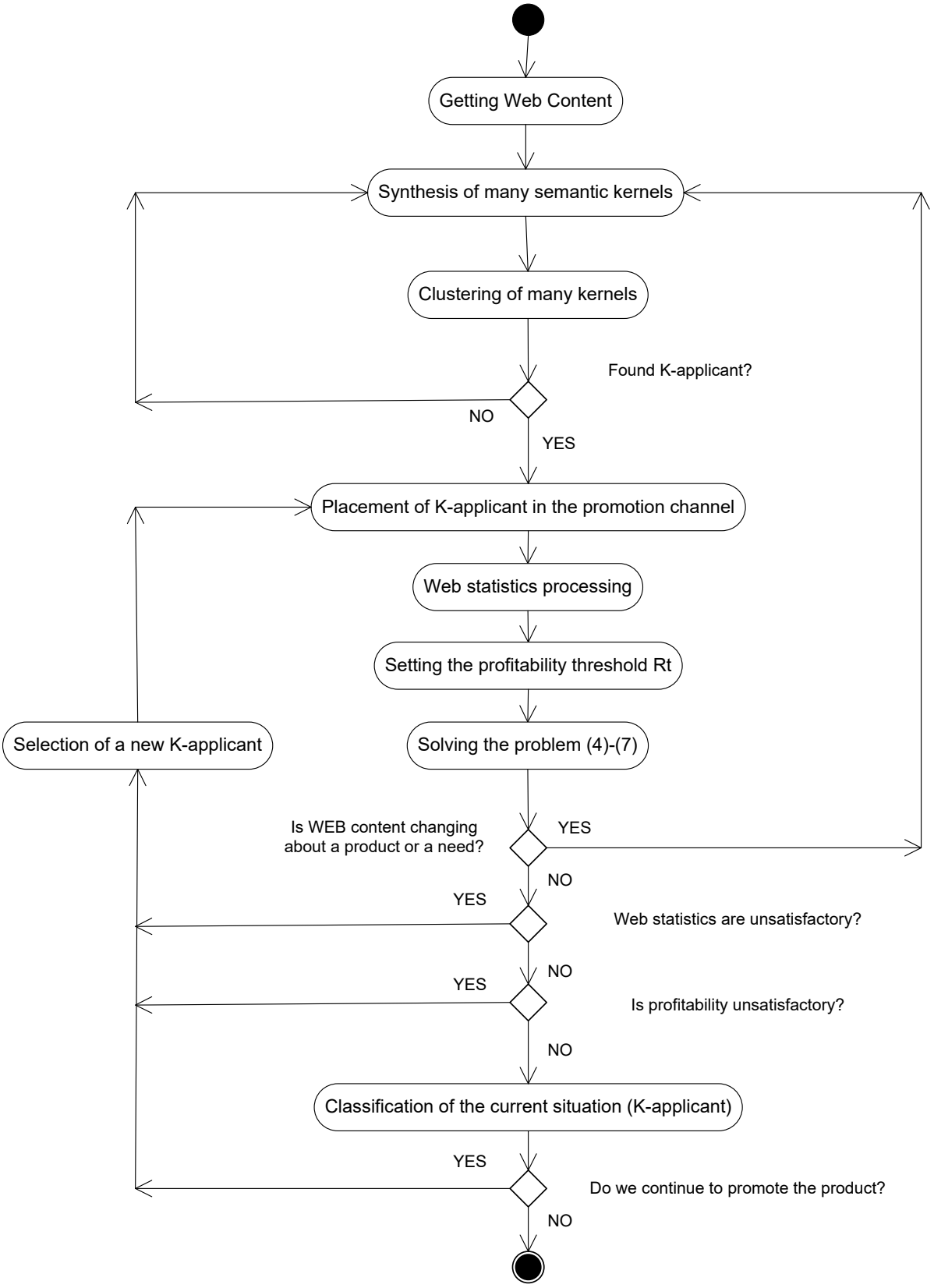

Fig. 2. The scheme of solving the problem

Consider the content of this scheme for solving the problem of situational management of the semantic kernel. At the first stage we will receive web content from the enterprise for synthesis of set of semantic kernels. Next, the clustering process is performed in order to create a set of needs classes, as well as centroids of each class. We believe that every centroid of needs class is a K-applicant. Place it in the promotion channel in each *j*-th node.

Next we follow the metrics of WEB statistics. We set the profitability threshold of the channel as a whole at a given period of time *t*. Based on these input data we solve problem (4)-(7) to evaluate the efficiency of the channel with the K-applicant installed in it.

We are checking the conditions for the transition to a new K-applicant. If all web content changes, we re-synthesize the whole set of semantic kernels and repeat the whole algorithm first. If the results of WEB statistics or profitability are unsatisfactory, then choose a new kernel (new current situation) and run it into the promotion channel. If these conditions are satisfactory, it is necessary to classify the K-applicant and continue the process of placing it in the channel. If the promotion of goods or needs stops, then the algorithm ends its execution.

Also uncertain is the step on the classification of the current situation (see figure 2). Based on the effect of aging of the semantic kernel, which was found in the research work [15], it is proposed to classify the current situation as one that corresponds to the present or as outdated. The outdated kernel needs to be changed to a new one anyway.

We offer a list of rules for the transition from one current situation to another, i.e. the rules for changing the current kernel to another. Let's divide the whole set of rules into two parts. The first part is the rules that affect the promotion process, i.e. management rules. The second group of rules is the rules that record the fact of aging of the nucleus or evolutionary rules.

The first group of rules is launched if the value of the profitability of the promotion is unsatisfactory. Let us denote this fact by the following formula:

$$\frac{\sum_{t \in T} R_t}{|T|} \geq \Delta^R \tag{9}$$

where the average profitability for a period of time T must not be lower than the specified value Δ^R. If condition (9) is not met, the kernel should be updated.

The second group of rules is formed on the basis of analysis of web statistics. As mentioned, we have four main indicators of the semantic kernel: the number of “isa” rules, the frequency of occurrence in web content, the frequency of occurrence in the search engine database and the number of keywords in the semantic kernel. But they do not correspond to the statistics that are actually available to us in the analysis of the promotion process. Available metrics in the promotion channel node are traffic, number of paid online orders, and number of leads. These are real indicators that give attention to the efficiency of the node with the semantic kernel placed in it.

Let's analyze each indicator and try to establish the rule of changing the kernel to another based on it. The first indicator is traffic. It shows the number of semantic kernel impressions in the promotion channel node. As practice shows, the dynamics of this indicator is the most important, but also the most uncertain. It all depends on the product itself and the factors around it. There are situations when the value of this indicator increases and profitability decreases, so it is extremely difficult to determine the direct relationship between these indicators. Let us denote this indicator as tr_{tj} - the traffic received by the *j*-th node for the time period t. Then we have the condition:

$$\frac{\sum_{t \in T} tr_{tj}}{|T|} \geq \Delta_j^{tr} \tag{10}$$

where Δ_j^{tr} is the rate of average traffic in the *j*-th node. If condition (10) is not met, the kernel should be updated. However, it should be noted that the fulfillment of condition (10) in some cases does not guarantee the need to change the kernel, because, for example, too small a budget for the placement of the semantic kernel in a given node does not increase traffic.

This is, so to speak, the first step in analyzing the need to update the kernel in the promotion channel. The second step is based on identifying the need to do so on the basis of other conventions that are not related to profitability or web statistics.

The above four cases of the need to update the kernel should also be considered through the prism of existing quantitative indicators. The first case when we discuss the evolution of the product on the market, and therefore its reflection on the Internet, can be recorded in terms of the following indicators, which can be monitored over time, and in a given node of the promotion channel. Such indicators may be:

1) the volume of sales of goods for a period of time in a given node;

2) conversion in a given node over time $t \in T$;

3) profit received through a given node over a period of time $t \in T$.

These three indicators can capture the process of product evolution in the market, in particular in cyberspace. In this case, if the profit schedule has only one maximum, it is likely that the product goes into the stage of extinction. Therefore, let there be a graph of the dependence of profits on time: $y = \Pr(t)$. If this graph has one extremum per maximum and the current value of this function is less than the extremum and newer in time, then the semantic kernel update needs to be updated:

$$\max_{t \in T} \Pr(t) \geq \Pr(t+1), (t+1) \notin T$$

Unfortunately, it is not possible to quantify other cases of kernel upgrades at present. Therefore, it is appropriate to introduce expert evaluation in the algorithm execution scheme. If the value is one, then an update is required. Values are set expertly by the company or team that performs the process of virtual promotion. Unfortunately, in this case, the execution of the scheme (Figure 3) will be in automated mode.

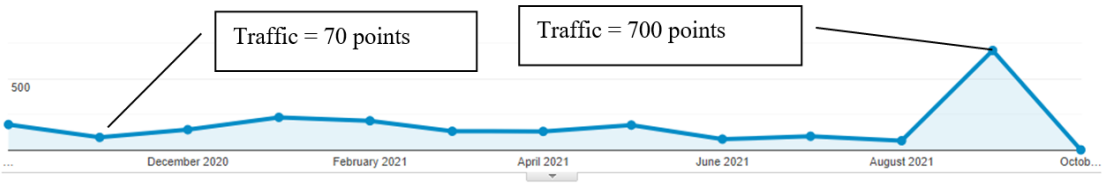

Fig. 2. The scheme of solving the problem

Results

In 2020-2021, a study was conducted for a web resource in the American service market, in which the cycle of situational management of the semantic kernel was applied. This web resource provides a unique service for building a psychological portrait of a person based on personal information. The results of this study are reflected in the graph of web statistics - Figure 3. This graph reflects the change in the main web metric - traffic or the number of visits. Figure 3 shows that the result of applying the cycle was an increase in traffic over a given period of time by ten times. This graph was obtained in the environment Google Analytics.

The data in Figure 3 show the existence of the very fact of the existence of the semantic kernel. That is, this is a confirmation of the efficiency of the proposed new method for implementing search engine optimization technology on the Internet.

Conclusions

The article considers new scientific results, namely:

1) for the first time the formulation of the problem of situational management of the semantic kernel and the method of its solution are formulated;

2) the method of search engine optimization on the Internet based on the concept of the semantic kernel was further developed;

3) for the first time the rules of transition between semantic kernels in the process of product promotion on the Internet are proposed, which is aimed at improving the efficiency of the search engine optimization process on the Internet;

4) for the first time proposed metrics and criteria for evaluating the effectiveness of the semantic kernel both individually and in the promotion channel on the Internet.

Further areas of research will be a deeper testing of the proposed method of controlling the kernels. You should test the transition rules and the strategy for increasing or decreasing the size of the semantic kernel in the Internet promotion channel.

References

1. Orekhov S. Method of Building Virtual Promotion Map // Computer Systems and Information Technologies. – 2021. – № 3. – pp. 52–58.
2. Sharma U., Thakur K.S. A Study on Digital Marketing and its Impact on Consumers Purchase. // International Journal of Advanced Science and Technology. – 2020. – №29(3). – pp. 13096 – 13110.
3. Orekhov S., Malyhon H. Virtual promotion knowledge management technology. // Bulletin of the National Technical University "KhPI". Series: System analysis, control and information technology. – Kharkiv: NTU «KhPI». – 2020. – № 1 (3) 2020. – pp. 74-79.
4. Orekhov S., Malyhon H., Goncharenko T., Liutenko I. Using Internet News Flows as Marketing Data Component. // CEUR-WS. – 2020. – Volume 2604. – pp. 358–373.
5. Borisov V., Avramenko D. Fuzzy Situational Control of Complex Technical Systems based on Composite Hybrid Models. // CEUR-WS. – 2021. – Volume 3044. – pp. 8-17.
6. Godlevsky M., Orekhov S., Orekhova E. Theoretical Fundamentals of Search Engine Optimization Based on Machine Learning. // CEUR-WS. – 2017. – № 1844. – pp. 23–32.
7. Orekhov S., Malyhon H. Metrics of virtual promotion of a product. // Bulletin of the National Technical University "KhPI". Series: System analysis, control and information technology. – Kharkiv: NTU «KhPI». – 2021. – № 2 (6). – pp.23–26.

8. Hashimova K.K. Analysis Method of Internet Advertising-Marketing Information's Dynamic Changes. I.J. Information Engineering and Electronic Business. – USA. – 2017. – № 5. – pp. 28–33.
9. García J., Lizcano D., Ramos C., Matos N. Digital Marketing Actions That Achieve a Better Attraction and Loyalty of Users: An Analytical Study. // Future Internet. – 2019. – №11(130). – pp. 1-16.
10. Bendle N., Farris P., Pfeifer P., Reibstein D. Marketing metrics. The manager's guide to measuring marketing performance. USA: Pearson Education, Inc., 2016. – 456 p.
11. Hashimova K.K. Analysis Method of Internet Advertising-Marketing Information's Dynamic Changes. I.J. Information Engineering and Electronic Business. – USA. – 2017. – № 5. – pp. 28–33.
12. Charu C. Aggarwal. Data Mining. The Textbook. USA: Springer Publ., 2016. – 746 p.
13. Rumpe B. Agile modeling with UML. Springer, Germany, 2017.
14. Keeling M. Design It! From programmer to software architect. USA: The Pragmatic Programmers LLC, 2017. – 345 p.
15. Orekhov S., Malyhon H., Stratienko N., Goncharenko T. Software Development for Semantic Kernel Forming. // CEUR-WS. – 2021. – Volume 2870. – pp. 1312–1322.

https://doi.org/10.31891/csit-2022-2-2
UDC 004.75:004.49

Bohdan SAVENKO, Antonina KASHTALIAN
Khmelnytskyi National University

A METHOD FOR DETERMINING THE EFFECTIVENESS OF A DISTRIBUTED SYSTEM FOR DETECTING ABNORMAL MANIFESTATIONS

Studying the effectiveness of self-organized distributed anomaly detection systems in computer systems is an important mandatory step, which is carried out to confirm the correctness, feasibility and feasibility of the developed solutions, including architecture, method of maintaining system integrity.

In order to conduct a study on the effectiveness of the use of self-organized distributed systems, anomalies in computer systems were identified in relation to the evaluation criteria. Determining the specifics of the application of the system also affects the choice of criteria for evaluating effectiveness.

The method of determining the effectiveness of the proposed solutions for the developed self-organized distributed system for detecting anomalies in computer systems has been developed. Software has been developed to ensure the functioning of a self-organized distributed anomaly detection system in computer systems to confirm the feasibility of the proposed solutions.

Experimental studies with the developed implementation of a self-organized distributed system for detecting anomalies in computer systems according to the obtained coefficients confirmed the effectiveness of the proposed solutions and the developed distributed system for its operation in the computer network.

Keywords: efficiency, method, criterion, distributed systems, intrusion detection systems.

Богдан САВЕНКО, Антоніна КАШТАЛЬЯН
Хмельницький національний університет

МЕТОД ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ РОЗПОДІЛЕНОЇ СИСТЕМИ ВИЯВЛЕННЯ АНОМАЛЬНИХ ПРОЯВІВ

Дослідження ефективності застосування самоорганізованих розподілених систем виявлення аномалії в комп'ютерних системах є важливим обов'язковим етапом, який проводиться з метою підтвердження коректності, доцільності та можливості реалізації розроблених рішень, зокрема щодо архітектури, методу підтримки цілісності системи.

Для проведення дослідження з ефективності застосування самоорганізованих розподілених систем виявлення аномалії в комп'ютерних системах було визначено щодо критеріїв для оцінювання. Визначення особливостей застосування системи впливає, також, і на вибір критеріїв для оцінювання ефективності.

Здійснено розробку методу визначення ефективності запропонованих рішень для розробленої самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах. Розроблено програмне забезпечення для забезпечення функціонування самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах для підтвердження можливості реалізації запропонованих рішень.

Проведені експериментальні дослідження з розробленою реалізацією самоорганізованої розподіленої системи виявлення аномалій в комп'ютерних системах згідно отриманих коефіцієнтів підтвердили ефективність запропонованих рішень і розробленої розподіленої системи щодо її функціонування в комп'ютерній мережі.

Ключові слова: ефективність, метод, критерій, розподілені системи, системи виявлення вторгень.

Introduction

Distributed computing, which is organized and implemented with the support of distributed systems, depends significantly in terms of their efficiency on the time spent on delivering tasks to a particular component of the distributed system.

The influence of time on the efficiency of distributed systems is very significant for such a class of them as intrusion detection systems, bait systems, distributed processor emulators, and so on. Therefore, an important stage in their development is to establish efficiency, the factors that affect it and study the possibility of improving efficiency.

The development of appropriate methods for calculating the effectiveness of distributed systems for the tasks of detecting computer attacks and malicious software is relevant. Because not all generally accepted methods are suitable for the specifics of this direction, and not all factors of influence may be significant in them.

Therefore, the development of a method for determining the efficiency of distributed systems is important and relevant. It will make it possible to compare the obtained result with the results of known methods.

Subject area analysis and related decisions

Distributed computing and the systems that support them are given a lot in scientific papers [1-5] and this area of research remains promising. Consider some features of distributed systems, which are presented in scientific articles.

In [6], the authors investigated the processes that exchange data through shared memory in distributed systems, in order to establish the possibility of randomly obtaining them from the main scheduler. They present a general method of calculating these values by classifying distributed algorithms according to their scheme of access to shared memory. In [7-9] the same problem is investigated, which is related to the generation of random numbers.

In [10], the authors present an approach to managing computer network resources under the condition of establishing trust in system components.

Thus, the development of a distributed anomaly detection system in computer systems is based on two components: the development of a distributed system; use and improvement of anomaly detection methods. Analysis of scientific results in these areas is the basis for creating a distributed system for detecting anomalies in computer systems. At the same time, determining the effectiveness of the created systems by comparing them with known solutions remains an urgent issue that needs to be investigated. To achieve this goal requires information about the architecture of known distributed systems, which are analogous to the developed distributed systems. Therefore, we will analyze the existing systems for detecting anomalies and intrusions into computer systems.

The closest software solutions to the developed system of detection of anomalies in local computer networks according to the set task are network systems of detection of intrusions, network anti-virus programs and non-commercial developments of the corresponding purpose.

The network system for detecting malicious software or computer attacks mainly has a centralized management module [11-14]. This allows the system administrator to manage updates and settings of all network settings from a single console. Network anti-virus tools are usually used in conjunction with anti-virus protection tools for network nodes as the second level of protection [15]. With such an architecture of network systems, it is recommended to use network and host parts from different manufacturers. For example, the system of protection of hosts in corporate networks is used as a security technology in network systems [12, 16]. Its elements allow you to control applications, web traffic and devices that connect. System functions are controlled from a single console. The network system from Dr.Web [17, 18] is the application "Dr.Web CureNet!". It is built according to a centralized architecture. Symantec Endpoint Protection is a network antivirus system. It provides the network administrator with the necessary tools to deploy antivirus tools on the network [13]. A hardware-software system for detecting malicious software and computer attacks was developed by Palo Alto Networks [19, 20]. The Malwarebytes Endpoint Security network system [21] provides an extended set of local tools for detecting and removing threats to computers on the network. Cisco® Network Admission Control (NAC) technology was developed to protect all hosts on the network [17, 18]. The Kaspersky Administration Kit network system implements independent work without intervention at the stage of research or identification of the administrator [22].

Host tools designed to detect malware or computer attacks: Avast! [23], AVG Antivirus [24], AntiVir (Avira) [25], BitDefender [26], Clam AntiVirus [20], etc.

Thus, the results of the analysis, in particular in [8, 9], show that host tools and network systems to detect malicious software or computer attacks do not ensure its complete detection. Both manufacturers and users of these tools agree with these results. The development and use of network detection systems based on the detection of anomalies in nodes in the network is promising for improving the reliability of detection. But at the same time there is a problem to prove the effectiveness of the developed distributed system. Therefore, a separate study requires the study of parameters that affect efficiency, and the direct development of a method for calculating the efficiency of the designed distributed systems.

The method for determining the efficiency of a distributed system

Research on the effectiveness of self-organized distributed anomaly detection systems in computer systems is an important mandatory step, which is carried out to confirm the correctness, feasibility and feasibility of developed solutions, including improved architecture, system integrity, anomaly detection method and implemented system. .

To conduct a study on the effectiveness of the use of self-organized distributed systems to detect anomalies in computer systems, it is necessary to determine the criteria for evaluation. Determining the specifics of the application of the system also affects the choice of criteria for evaluating effectiveness.

Since the system in question is self-organized, it will function in virtually all functions without user intervention and, therefore, the criterion that satisfies the user's requirements is to maximize the time in the system when it does not involve the user to decide on further steps. work or its component, as well as changes in its architecture. As a variable, ie an argument, in the function that will describe and set the requirements in this criterion will be time, because it can be measured during the operation of the system. The time can be determined to determine the entire period of operation of the system from the beginning of the system startup, during the working day, as well as the time spent on the user's processing of certain system states. The time spent processing certain states of the system by the user (system or network administrator) can be divided by the time determined when the system is completely shut down and when part of it is running. This division allows you to take into account the peculiarities of the processes that will occur during the period of operation of the system, and provides an opportunity to correlate different time intervals.

Let t_1^1 – the time during which the self-organized distributed system functioned. Time will be determined in hours. This is all the time spent on the work of the entire system, which includes the functioning of its individual modules offline during the absence of the center and the functioning of the system in a disabled state, that is, the time in a passive state between the periods of operation of the system. Thus, this is the time of the full cycle of

operation of the system. We will introduce $t_{2,i}^{1}$. as the time of functioning of the system or its individual components in the active state during i – the day where $i = 1, 2, \dots, d$, d – the number of days of functioning of the system, then the fair ratio $t_1^1 =< 24 * d$. Since, during the day, the system can be active or be inactive, when all the computer stations in which it is installed are disabled, it is advisable to divide the time between these cases, and also, we can consider the frequency for it, if you take a day as a time interval. As the time of stay of the system or its individual components in an inactive state, when the computer stations in which they are installed are disabled, we will enter the $t_{3,i}^{1}$. during i – the day where $i = 1, 2, \dots, d$, d – the number of days of operation of the system, then the following ratios will be fair:

$$t_{2,i}^{1} + t_{3,i}^{1} = 24, (t_{2,i}^{1} + t_{3,i}^{1}) * d \geq t_1^1, \quad (1)$$

consider the time t_1^2 - the time during which the self-organized distributed system functioned without the intervention of the user or system administrator. Then, time t_1^3 - the time during which the self-organized distributed system functioned with user or system administrator intervention at the request of the system. The time when the system functioned and there was interference of the user or system administrator will not be considered and will not be taken into account in this case, since it refers to the time during which the anomalies related to the specifics of the system will be investigated. Generally investigated time intervals are designated as t_1^1, t_1^2, t_1^3, $t_{2,i}^{1}$, $t_{3,i}^{1}$. can be attributed to the external characteristics of the system, such as the time when a user or system administrator intervened without asking for such an action by the system's internal system characteristic. Given time values associated with the ratio:

$$t_1^2 + t_1^3 = t_1^1. \quad (2)$$

If we consider the time value t_1^3, as taking into account the time during which the system or its component was serviced by the user or system administrator and was completely inactive, we denote the value of $t_1^{3,1}$, and the time during which the system continued to work, and were served by the user or system administrator on its request components, which is denoted by the value $t_1^{3,2}$. In the case of determining the value $t_1^{3,2}$, the presence of a working component with a decision-making center of the highest level of hierarchy in the system is mandatory. Divide the value $t_1^{3,2}$ at different time values that will characterize the time intervals in different cases: $t_1^{3,2,1}$ – system uptime or components with a decision-making center of the highest level of hierarchy in the system, when the user or system administrator serves at the request of the system part of its components; $t_1^{3,2,2}$ – maintenance time components that do not contain a top-level decision center in the system, user, or system administrator.

Using the entered numeric characteristics of the system, we determine the value r_1^1, which will characterize the share of time spent by the user or system administrator to maintain components that do not contain a higher-level decision center in the system, using the formula:

$$r_1^1 = \frac{t_1^{3,2,2}}{t_1^{3,2,1} - t_1^{3,2,2}}. \quad (3)$$

Value r_1^1 characterizes the case when a part of the system with a decision center of the highest level of the hierarchy works and at the same time part of the system is served by the system administrator or user, but it is considered that the time spent on maintenance affects the system and, therefore, the time of full independent functioning of the system is considered. If it is not taken into account during the independent functioning of the system, then we will introduce the appropriate value r_1^2 and calculate it by the formula:

$$r_1^2 = \frac{t_1^{3,2,2}}{t_1^{3,2,1}}. \quad (4)$$

We will introduce a criterion for evaluating the effectiveness of work according to the coefficient K_1, based on its ability to independently make decisions without involving a user or a system administrator. Values r_1^1 and r_1^2 will be convergence to zero if time is minimized $t_1^{3,2,2}$, but the convergence rate depends very much on the value $t_1^{3,2,1}$, so given that they are close in value between them we will take them into account in the final representations to determine the coefficient K_1 criterion for evaluating the effectiveness of work as follows:

$$K_1 = \frac{r_1^1 + r_1^2}{2}. \quad (5)$$

The value of the coefficient K_1 calculated by the formula (5) will be averaged and more accurately describes the given time value as a characteristic in the system. The criterion for assessing the effectiveness of the system is as follows:

$$K_1 = f\left(t_1^{3,2,1}, t_1^{3,2,2}\right) = \frac{\frac{t_1^{3,2,2}}{t_1^{3,2,1} - t_1^{3,2,2}} + \frac{t_1^{3,2,2}}{t_1^{3,2,1}}}{2} = 2 \cdot \frac{t_1^{3,2,1} \cdot t_1^{3,2,2} - t_1^{3,2,2} \cdot t_1^{3,2,2}}{t_1^{3,2,1} \cdot (t_1^{3,2,1} - t_1^{3,2,2})},$$

$$K_1 \to 0, \text{при} \min\left(t_1^{3,2,2}\right), \max\left(t_1^{3,2,1}\right). \quad (6)$$

So, determining the coefficient K_1 the formula (6) allows you to evaluate the effectiveness of the system. For example, if $t_1^{3,2,1} \leq t_1^{3,2,2}$, then it is obvious that $K_1 > 1$ and the functioning of the system during the study was ineffective.

Results obtained for efficiency criterion according to coefficient K_1 relate to the case for the total time of the system functioning and can be clarified taking into account each component of the system. That is, for each component of the system we will calculate the time of its functioning in different cases and present in the criteria for evaluating the effectiveness of the system according to the characteristics in each of the components. The component of a self-organized distributed system in a computer station in a network when analyzing its functioning according to the time indicator can be attributed to one of the cases: functioning simultaneously with the component in which the decision-making center of the highest level of hierarchy is located; functioning compatible with other components of the system without the component in which the decision-making center of the higher level of hierarchy is located; non-functioning and located in a computer system that is disabled; is non-functioning because it is being processed by a user or system administrator. Taking into account that the components of the system create parallel executable processes in different nodes in the network and the time intervals in which they are active are different, the study of the functioning of the system according to the time indicator must be carried out throughout the time of the system functioning as a whole or on average data during the day, since the active functioning of the system is periodic.

We will introduce for each j – the components of the system such values that will characterize its functioning according to time indicators, and $j = 1, 2, \dots, N$, N – the number of components of the system. For the case of $j = 0$, that is, the components that contain a decision center of a higher level of the hierarchy, we will define the time by its different states separately. Let $t_{1,j}^1$ – the time during which j functioned – that component of the self-organized distributed system. That is, it is all the time spent on the work of j – that component of the system, which includes the functioning in the following cases: offline in the presence of components with a decision center of the highest level of hierarchy; in the absence of the center; functioning of the system components in a disabled state, that is, the time in a passive state between the periods of operation of the system; A time period when a component was not functioning and was maintained by a user or system administrator. Thus, we consider these four time intervals as describing the time of the full cycle of operation j – that component of the system. Then, let's introduce $t_{2,i,j}^1$ as the time of functioning of j – the component of the system in the active state in the presence of components with a decision-making center of the highest level of hierarchy during i – the day where $i = 1, 2, \dots, d$, d – the number of days of operation of the system. To characterize the period of absence of the center, but the functioning of the system components, we will enter the value $t_{3,i,j}^1$ as the time of functioning of j – the component of the system in an active state in the absence of components with a decision-making center of the higher level of the hierarchy during i – the day where $i = 1, 2, \dots, d$, d – the number of days of operation of the system. The functioning of j is the component of the system in the disabled state, that is, the time in a passive state between the periods of operation of the system to mean the characteristics of the value $t_{4,i,j}^1$. The time interval when j – and the component did not function, but was maintained by the user or system administrator, we set the value of $t_{5,i,j}^1$, in which this characteristic is defined during i – the day when $i = 1, 2, \dots, d$, d – the number of days of operation of the system.

Similarly, we will introduce time characteristics for components with a higher level decision center of the hierarchy, that is, when $j = 0$. Enter the value $t_{2,i,0}^1$ as the time of functioning of the components with the decision-making center of the higher level of the hierarchy during i – the day where $i = 1, 2, \dots, d$, d – the number of days of operation of the system. To characterize the absence period, components with a higher-level decision center of the hierarchy due to its inaccessibility to the rest of the system components due to the definition of performing functions to determine the further actions or steps of the system, we will enter the value $t_{3,i,0}^1$ during i – the day where $i = 1, 2, \dots, d$, d – the number of days of operation of the system. Functioning of the system components with a decision center of the higher level of hierarchy in the disabled state, that is, the time in a passive state between the periods of functioning of the system to mean the characteristics of the value $t_{4,i,0}^1$. The time interval when the system component with the decision center of the higher level of the hierarchy did not function, but was served by the user or system administrator, we set the value of $t_{5,i,0}^1$, in which this characteristic is defined during i – the day when

$i = 1, 2, \dots, d$, d – the number of days of operation of the system. We will establish the following ratio between the time characteristics of the system components:

$$\frac{\sum_{i=1}^{d}\sum_{j=0}^{N}\sum_{k=2}^{5} t_{k,i,j}^{1}}{N} = t_{1,j}^{1}. \quad (7)$$

All components of the system are given time characteristics in such a way that they are parts of the system for the entire functioning of the system and this is specified by the formula (7). But in the process of prolonged functioning of the system, some of its components can be removed from it and for this case the following ratios are fair:

$$\sum_{j=0}^{N} t_{1,j}^{1} = \sum_{i=1}^{d}\sum_{j=0}^{N}\sum_{k=2}^{5} t_{k,i,j}^{1}, \quad (8)$$

$$t_{1}^{1,c} = \frac{\sum_{i=1}^{d}\sum_{j=0}^{N}\sum_{k=2}^{5} t_{k,i,j}^{1}}{N} \geq t_{1,j}^{1}, \quad (9)$$

where $t_1^{1,c}$ – the average arithmetic value of the time components in the system throughout the entire period of its functioning.

Determine the coefficient K_2 for the criterion of estimation of the effectiveness of the system, based on the entered characteristics and the obtained ratios specified by the formulas (7)-(9), to take into account the characteristics that depend on the components of the system.

Using the entered numeric characteristics of the system components, we determine the value $r_{1,j}^{1}$, which will characterize the share of time spent by the user or system administrator to serve the j-th component by the formula:

$$r_{1,j}^{1} = \frac{t_{1,j}^{3,2,2}}{t_{1,j}^{3,2,1} - t_{1,j}^{3,2,2}}. \quad (10)$$

where $t_{1,j}^{3,2,1}$ – the time of operation of j-th components with a decision-making center of the highest level of the hierarchy in the system, when the user or system administrator serves at the request of the system part of its components, including the j-th component; $t_{1,j}^{3,2,2}$ – maintenance time for j-part that does not contain a top-level decision center in the system, user, or system administrator.

Value $r_{1,j}^{1}$ characterizes the case when a part of the system with a decision center of the highest level of the hierarchy works and at the same time part of the system is served by the system administrator or user, but it is considered that the time spent on maintenance affects the system and, therefore, the time of full independent functioning of the system is considered. If it is not taken into account during the time of self-operation of the system, then we will introduce the appropriate value $r_{1,j}^{2}$. For j-part and calculate it by formula:

$$r_{1,j}^{2} = \frac{t_{1,j}^{3,2,2}}{t_{1,j}^{3,2,1}}. \quad (11)$$

We will introduce a criterion for evaluating the effectiveness of work for j-th component according to the coefficient $K_{2,j}$, based on the possibility of part of the system to independently make decisions without involving the user or system administrator. Values $r_{1,j}^{1}$ and $r_{1,j}^{2}$ will be convergent to zero if time is minimized $t_{1,j}^{3,2,2}$, but the convergence rate is very dependent on the value $t_{1,j}^{3,2,1}$, so given that they are close in value between them we will take them into account in the final representations to determine the coefficient $K_{2,j}$. criterion for evaluating the effectiveness of the work of j-th component as follows:

$$K_{2,j} = \frac{r_{1,j}^{1} + r_{1,j}^{2}}{2}. \quad (12)$$

The value of the coefficient $K_{2,j}$, calculated by the formula (12) will be averaged and more accurately describe the given time value as a characteristic of j-of-that component in the system. The criterion for assessing the efficiency of the j-part of the system is set as follows:

$$K_{2,j} \to 0, \text{при } \min\left(t_{1,j}^{3,2,2}\right), \max\left(t_{1,j}^{3,2,1}\right). \quad (13)$$

So, determining the coefficient $K_{2,j}$ the formula (13) allows you to evaluate the effectiveness of the functioning of the j-th component of the system. The general efficiency factor of the system, taking into account the

time characteristics of the system components and the values of coefficients obtained from the formula (13), is found by the formula:

$$K_2 = \max\left(K_{2,0}, K_{2,1}, \dots, K_{2,N}\right). \quad (14)$$

The system performance criterion, which is set using the coefficient value K_2, similar to the case of K_1, will display the best result when minimizing its value. Thus, the value of coefficients K_1 and K_2 will coincide to zero, provided that there are no long-term failures in the system and its components, as well as in the absence of destructive effects of malware and computer attacks on nodes on the network in which the system components are located, and short-term maintenance work on the part of the user or system administrator. In addition, the obtained values of coefficients that are calculated for a certain period of time will be taken into account when determining the further actions of the system.

Experimental research on the use of a self-organized distributed system

The purpose of experimental research is to study the effectiveness of the functioning of the self-organized distributed system and the reliability of anomaly detection in computer systems. The study of the effectiveness of the functioning of the self-organized distributed system is necessary to establish the implementation of its functions to maintain the integrity and coordination of the work of the system components. The reliability of the detection of an anomaly in computer systems requires research to establish the possibility of its use in real conditions. The part of the distributed system responsible for detecting an anomaly is implemented as an appropriate method in it, and its testing will allow to evaluate the reliability of the detection of an anomaly in computer systems.

Therefore, consider testing the distributed system first with the disabled module responsible for detecting an anomaly. Indicators that will be investigated are grouped by the following characteristics: communication time between individual components; time of communication between the components of the system and the component in which the decision-making center of the higher level of the hierarchy is located; communication time between components depending on the number of active components that form the system; the time spent on the work when the center distribution was carried out in the system, i.e. distributed center testing, and the time spent when the center is not divided between the levels of the hierarchy, but is completely in one component.

Experimental studies with the developed self-organized distributed system were carried out in a local computer network created using Ethernet technology with a data transfer rate of 1 GB/s between nodes in the network. Since the messages for transmission between the components of the system contain a very small amount of information, they are formed in short packages and we assume that each of them was transmitted in one package of 64 bytes.

The self-organized distributed system studied contained eight components, one of which was a decision-making center of the highest level of hierarchy. Experimental studies were carried out within 50 days separately for the case when the decision-making center was distributed among all components taking into account two levels of hierarchy and also 50 days for the case when the decision-making center was placed only in one component. During the entire time of the experiment, the time of sending messages, their number and recording of the time of receipt were recorded. This was carried out in order to conduct a study of the time spent on communication in the middle of the system itself. The results of experimental studies are presented in Table 1.

Table 1

Results of experimental studies on the effectiveness of the functioning of the self-organized distributed system

№	Time value characteristics	Time interval for the case when the time spent on the work of the system with the distribution of the center, i.e. testing with a distributed center, s	Time interval for the time spent on the system without distributing the center between the components, s
1	The time between individual components	1,42 – 2,61	1,41 – 2,56
2	Communication time between the system components and the component that hosts the top-level decision-making center of the hierarchy.	1,81 – 2,87	1,67 – 2,23
3	Communication time between components depending on the number of active components that form the system. Cases 3.1. Number of components 2-4. 3.2. Number of components 5-8.	1,41 – 2,51 1,83 – 2,81	1,27 – 2,39 1,72 – 2,43
4	Number of packages moved for the event under study 1.	7546	5788
5	Number of packages moved for the event under investigation 2.	12458	8386

The results of experimental studies presented in Table 1 confirm the increase in the number of messages transmitted for the case when the time spent on the work of the system with the distribution of the center, that is, testing with a distributed center, compared to the case when the time spent on the system without the distribution of the center among the components. In addition, the processing time of transmitted message packets increases because the time is wasted.

To the work of the system with the distribution of the center, messages from the system components in parallel at one time are sent to the components, which houses the decision-making center and is sent to all components, unlike when the decision-making center, which is represented only in one component, independently decides on communication with the rest of the component, which reduces significantly the number of messages between components. But at the same time, the time costs for the transmission of messages are significantly small in both cases, so the use of architecture with the distribution of the center between the levels of hierarchy in two different types of system components is effective, which is confirmed by the results of experiments.

The next step in the processing of experimental studies is to determine the ratio of processing time and functioning of the self-organized distributed system according to the introduced coefficients of its efficiency of work K_1 and K_2. Since the time characteristics of the processes that took place in the self-organized distributed system under study are fixed, in addition to the fact that they were used by the system itself to determine further actions, we will use them to build function graphs according to its efficiency coefficients K_1 and K_2. For the experiment, a variant of the system was used with the distribution of the center between two levels of the hierarchy. Initially, the case when the user or system administrator did not intervene was analyzed, then the time spent processing was zero and, indeed, there were no costs associated with servicing part of the components in the system. The second case in the experiment involved and was realized when the time of maintenance of system components by the user or system administrator was significantly shorter than the entire system and when it was longer. The results of such experiments are presented in Fig. 1. They refer to two parameters that characterize the system. These settings are the system components' time in the site and the time it took the user or system administrator to process the event in a computer station. Determining such time parameters will allow calculating the coefficients K_1 and K_2. The time intervals in the graphs are represented by time charts. Of the eight components, only three (4,5,7) required user or system administrator interventions. The rest were not needed. The last ninth graph displays a summary time chart of the entire system.

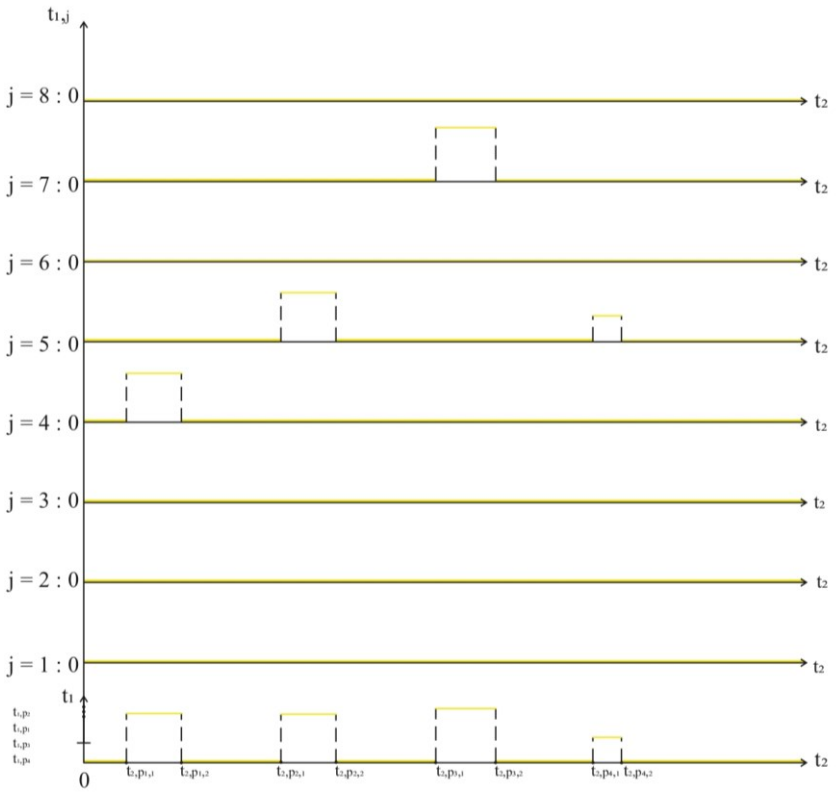

Fig. 1. Time charts based on the results of experimental studies

The value of coefficients K_1 and K_2 after the experiments are presented in Table. 2.

Table 2

Coefficient values K_1 and K_2

№	System components maintenance time by the user or system administrator was significantly shorter than the entire system	System components maintenance time by the user or system administrator was significantly longer than the entire system
	Case 1	Case 2
K_1	0,00987567	0,04873418
K_2	0,00986972	0,04873326

The obtained coefficients confirm the effectiveness of the proposed solutions and the developed distributed system for its functioning in the computer network. In the second case, not large values of coefficients are justified by the fact that for a long time only certain components of the system were served, and most components of the systems functioned. This proves the effectiveness of using distributed systems of this type.

Also, according to the results of the experiment, it was found that when the decision-making center is divided between the levels of the hierarchy, the efficiency in time is better, because processing at the lower level of the hierarchy reduces the time of event handling compared to using one center. Reliability in the processing of an anomaly in computer systems, which were artificially entered, is 0.8356, which is a satisfactory result and confirms the sufficient effectiveness of the proposed solutions.

Conclusions

To evaluate the effectiveness of the proposed solutions, a method for calculating the effectiveness of the use of a self-organized distributed anomalies detection system in computer systems has been developed. It uses two entered criteria, the results of which are determined by the appropriate coefficients. In addition, the obtained values of coefficients that are calculated for a certain period of time are taken into account when determining further actions of the system.

Experimental studies with the developed implementation of a self-organized distributed system for detecting anomalies in computer systems according to the obtained coefficients confirm the effectiveness of the proposed solutions and the developed distributed system for its functioning in the computer network.

References

1. Savenko, O., Nicheporuk, A., Hurman, I., Lysenko, S. - CEUR-WS. – 2019. – Vol. 2393. – P.633-643, ISSN: 1613-0073.

2. Savenko O. Metamorphic Viruses' Detection Technique Based on the Equivalent Functional Block Search / O. Savenko, S. Lysenko, A. Nicheporuk, B. Savenko // CEUR-WS, ISSN: 1613–0073. – 2017. – Vol. 1844. – Pp. 555–569.

3. Lysenko S., Bobrovnikova K., Matiukh S., Hurman I., Savenko O. Detection of the botnets' low-rate DDoS attacks based on self-similarity. *International Journal of Electrical and Computer Engineering*, Vol. 10, Issue 4, 2020, Pages 3651-3659. DOI: http://doi.org/10.11591/ijece.v10i4.pp3651-3659.

4. Pomorova O. Metamorphic Viruses Detection Technique based on the the Modified Emulators [Text] / O. Pomorova, O. Savenko, S. Lysenko, A. Nicheporuk // CEUR-WS. – 2016. – Vol. 1614. – PP.375-383, ISSN: 1613-0073

5. Sergii Lysenko, Kira Bobrovnikova and Oleg Savenko. A Botnet Detection Approach Based on The Clonal Selection Algorithm // Proceedings of 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DeSSerT-2018, Kyiv, Ukraine, May 24-27, 2018) – Pp. 424-428.

6. Antoniadis, K., Blanchard, P., Guerraoui, R. *et al.* The entropy of a distributed computation random number generation from memory interleaving. *Distrib. Comput.* **31,** 389–417 (2018). https://doi.org/10.1007/s00446-017-0311-5

7. Alistarh, D., Sauerwald, T., Vojnovic, M.: Lock-free algorithms under stochastic schedulers. In: Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing, PODC 2015, Donostia-San Sebastián, Spain, July 21–23, 2015, pp. 251–260 (2015). doi:10.1145/2767386.2767430

8. Barker, E., Kelsley, J.: Recommendation for random bit generator (rbg) constructions. SP 800-90C (2012)

9. Zhou, H., Bruck, J.: Generalizing the Blum-Elias method for generating random bits from markov chains. In: Proceedings of IEEE International Symposium on Information Theory (ISIT) (2010)

10. Pomorova O. Multi-Agent Based Approach for Botnet Detection in a Corporate Area Network Using Fuzzy Logic [Text] / Oksana Pomorova, Oleg Savenko, Sergii Lysenko, and Andrii Kryshchuk // Communications in Computer and Information Science. – 2013. – Vol. 370. - PP.243-254, ISSN: 1865-0929.

11. B. Savenko, S. Lysenko, K. Bobrovnikova, O. Savenko, G. Markowsky. Detection DNS Tunneling Botnets // Proceedings of the 2021 IEEE 11th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IDAACS'2021, Cracow, Poland, September 22-25, 2021.

12. Corporate Endpoint Protection Products Group Test: Socially-Engineered Malware Q2 [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.nsslabs.com/ research/ endpoint-security/ anti-malware/q2-2010-endpoint-protection-product-group-test.html (Viewed on April 3, 2021). – Title from the screen.

13. ESET Endpoint Security [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.eset.com/ (Viewed on April 3, 2021). – Title from the screen.

14. Symantec Endpoint Protection [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: https://www.anti-malware.ru/reviews/ Symantec_Endpoint_Protection (Viewed on April 3, 2021). – Title from the screen.

15. Branitskiy A., Kotenko I. Hybridization of computational intelligence methods for attack detection in computer networks. Journal of Computational Science. 2017. №23. P. 145–156.

16. Stetsyuk M., Stetsyuk V., Savenko B., Savenko O., Dobrowolski M. Implementation of Control by Parameters of Client Automated Workplaces of Specialized Information Systems for Neutralization malware. CEUR-WS. 2021. Vol. 2853. P. 340-352. URL: http://ceur-ws.org/Vol-2853/paper40.pdf.Securelist. FAQ: Disabling the new Hlux / Kelihos Botnet [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: https://securelist.com/blog/research/32634/faq-disabling-the-new-hluxkelihos-botnet-13/ (Viewed on April 3, 2021). – Title from the screen.

17. COMSS1 [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.comss.ru/page.php?id=2758 (Viewed on April 3, 2021). – Title from the screen.

18. Enterprise End Point Protection Comparative Analysis - Socially Engineered Malware: Report Overview [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: www.nsslabs.com/reports (Viewed on April 3, 2021). – Title from the screen.

19. Bakotech [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: https://bakotech.ua/ (Viewed on April 3, 2021). – Title from the screen.

20. ClamAV [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: https://www.clamav.net/ (Viewed on April 3, 2021). – Title from the screen.

21. Malwarebytes Endpoint Security [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: https://ru.malware bytes. com/business/endpoint security (Viewed on April 3, 2021). – Title from the screen.

22. Kaspersky Lab [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.kaspersky.ru (Viewed on April 2, 2019). – Title from the screen.

23. Avast! [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: https://www.avast.com/index (Viewed on April 21, 2020). – Title from the screen.

24. AVG [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.avg.com (Viewed on April 21, 2020). – Title from the screen.

25. Avira [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.avira.com (Viewed on April 21, 2020). – Title from the screen.

26. Bitdefender [Electronic resource]: [Web-site]. – Electronic data. – Mode of access: http://www.bitdefender.com/(Viewed on April 21, 2020). – Title from the screen.

https://doi.org/10.31891/csit-2022-2-3
UDC 004.051

Andrii KOPP, Dmytro ORLOVSKYI, Dorukhan ERSOYLEYEN
National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

AN APPROACH TO APPLICATIONS ARCHITECTURE MODELS ANALYSIS

A relevant problem of applications architecture model analysis was considered in this paper. Its significance is defined by the fact that designed blueprints of software systems should be thoroughly checked for all potential inefficiencies in order to avoid additional effort and costs for defect correction in later project stages. As a result, the research goal was defined as detecting strong and weak points in software design solutions via applications architecture model analysis. The research objective was set at the process of analyzing applications architecture models, and the research subject was set at the software solution for analyzing applications architecture models. Existing software tools for applications architecture modeling and analysis were defined based on an examination of general software development problems for applications architecture model analysis. The ArchiMate enterprise architecture modeling language was chosen as the standard representation of applications architecture models to be analyzed because there are nearly no alternatives to ArchiMate language for architectural description of enterprise application models that are standardized, supported by most diagramming software, and exchangeable. The domain of applications architecture models analysis was discovered, an approach to analyzing applications architecture models was proposed, a software solution for analyzing applications architecture models was designed and developed, and it was used to evaluate applications architecture models that represent web development patterns. The analysis results could be used by system or software architects to estimate the suitability of applications architecture solutions for ongoing projects, detect flaws in specific architectural patterns, and reduce effort and costs in later project stages.

Keywords: applications architecture, application model, model analysis, enterprise architecture, software solution.

Андрій КОПП, Дмитро ОРЛОВСЬКИЙ, Дорухан ЕРСОЙЛЕЄН
Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

ПІДХІД ДО АНАЛІЗУ МОДЕЛЕЙ АРХІТЕКТУРИ ЗАСТОСУНКІВ

У цій роботі було розглянуто актуальну проблему аналізу моделей архітектури застосунків. Її значення визначається тим, що розроблені проєкти програмних систем повинні бути ретельно перевірені на наявність усіх потенційних недоліків, щоб уникнути додаткових зусиль і витрат на виправлення дефектів на наступних етапах проєкту. Отже, метою дослідження є визначення сильних і слабких сторін проєктів за допомогою аналізу моделей архітектури застосунків. Метою дослідження є процес аналізу моделей архітектури застосунків, а предметом дослідження – програмне рішення для аналізу моделей архітектури застосунків. Існуючі програмні засоби для моделювання та аналізу архітектури застосунків були визначені на основі огляду загальних проблем розробки програмного забезпечення для аналізу моделей архітектури застосунків. Мова моделювання архітектури підприємства ArchiMate була обрана як стандартне представлення моделей архітектури застосунків, які підлягають аналізу, оскільки майже немає альтернатив мові ArchiMate для архітектурного опису моделей корпоративних застосунків, які стандартизовані, підтримуються більшістю програмного забезпечення для створення діаграм, а також є придатними для обміну. Розглянуто предметну область аналізу моделей архітектури застосунків, запропоновано підхід до аналізу моделей архітектури застосунків, спроєктовано та розроблено програмне рішення для аналізу моделей архітектури застосунків, за допомогою якого були проаналізовані моделі архітектури застосунків, що представляють собою шаблони веб-розробки. Результати аналізу можуть бути використані архітекторами систем або програмного забезпечення для оцінки придатності рішень щодо архітектури застосунків для поточних проєктів, виявлення недоліків у конкретних архітектурних шаблонах та зменшення зусиль і витрат на наступних етапах проєкту.

Ключові слова: архітектура застосунку, модель застосунку, аналіз моделі, архітектура підприємства, програмне рішення.

Introduction

According to the IEEE standard (1471-2000), architecture is the fundamental organization of a system, embodied in its components, their connections with one another, and the environment, as well as the rules controlling its design and development [1]. Because the modern enterprise is a complex system comprised of several interconnected areas, the enterprise architecture describes the components, their relationships, and the principles underlying this system [2]. Enterprise architecture development aims to identify explicit links and dependencies between various organizational domains such as business architecture, information architecture, applications architecture, and technical infrastructure. Architectural development is concerned with describing the parts that comprise an enterprise and how they interact in order to improve the understanding and vision required to successfully design a business architecture and information technology. According to industry practitioners, the most significant but also most complex organizational area is the relationship between business architecture and application, and IT (Information Technology) architecture [1].

Leading organizations create application and IT architecture with industry-proven technologies that decrease architectural documentation duplication, shorten development cycle times, and provide consistent vocabulary that promotes consistency in architectural descriptions. Some of these technology providers have entered the Gartner Magic Quadrant for Enterprise Architecture Tools' Leadership [3].

One of the first tools to support the business architecture technique was ARIS Business Architect (Software AG). The ARIS product has its own methodology and does not support TOGAF, the Zachman Framework, or similar frameworks. The ARIS (Architecture of Integrated Information Systems) methodology is a comprehensive approach to developing and analyzing business process models, as well as modeling the overall enterprise architecture. The ARIS technique (also referred as "ARIS House") distinguishes five sorts of representations: organizational, functional, process, data description, and outputs. ARIS includes integration modules for the SAP system. Among the significant flaws is the ARIS closed metamodel, which prevents changes to the approach to business architectural management and the introduction of new types of objects [4]. ARIS prioritizes business process management and business architecture over application and IT architecture, which are viewed as just supporting the execution of business operations and are not at the heart of ARIS methodology and its software solutions.

The MEGA Suite (MEGA International) is a comprehensive enterprise architectural management solution that includes tools for modeling, control, transformation, communication, project planning, and strategic migration from as-is to as-should. Unlike ARIS, MEGA is a highly configurable meta-modeling tool that supports international TOGAF, DoDAF, and Zachman Framework standards, as well as modeling in ArchiMate, BPMN (Business Process Model and Notation), and UML (Unified Modeling Language) notations, and contains libraries of industry standards eTOM, ITIL, and APQC [4]. By our opinion, the one advantage of MEGA over ARIS is support for ArchiMate and UML modeling standards; nonetheless, this software solution is still focused on the management prospect of EA rather than the application and IT domains in particular.

Visual Paradigm, a significant participant in the software modeling area, has begun to support the ArchiMate modeling language [5]. The vendor of Visual Paradigm claims that their product is a powerful and intuitive diagramming tool for architecture modeling that supports: drag-and-drop editing interface; precise shape positioning with alignment guide; many formatting options for shapes and connectors; and color categorization of shapes [5]. Visual Paradigm has debuted its online ArchiMate modeling software, which is marketed as the most capable and user-friendly corporate architecture modeling software on the market [6]. It has an easy-to-use user interface and drag-and-drop capabilities, which reduces modelers' learning curve when creating ArchiMate diagrams. This online program also supports Microsoft Visio import and interaction with Microsoft Office. PDF and graphic formats are among the export possibilities. The most significant advantage of Visual Paradigm Online for ArchiMate modeling is that it does not require downloading or user registration and login – diagrams may be built on the go; all that is required is to access the product's web page. Obviously, such a product is far better suited for application and IT architecture design than ARIS, because it not only supports ArchiMate (while ARIS only supports its own notation for application and technical enterprise modeling), but also UML diagramming capabilities to explode ArchiMate application components into detailed low-level software descriptions.

In addition to the commercial products discussed, Archi, an open source cross-platform tool that supports The Open Group TOGAF and ArchiMate approaches, is worth considering. The Archi enterprise modeling tool is a modular framework built on top of the Eclipse integrated development environment (IDE), allowing architects or software developers to write plugins in Java [4]. Because of its free availability, this program can be used in place of the paid ARIS and MEGA solutions for working on minor, particularly educational, projects. Archi's open source distribution enables practically anyone to improve its features at no additional expense [7].

Despite the ease of use of Visual Paradigm Online for ArchiMate modeling, this tool has a significant drawback: it only allows the user to make and publish designed application and IT architecture solutions in PDF or graphical image formats, which are fine for human reading but cannot be processed by computers. In turn, ArchiMate supports the exchange file format enabling diagram export and interoperability with other modeling applications. This format is based on XML (eXtensible Markup Language), but it only follows a specific schema [8]. In terms of analytical capabilities, the examined products are more concerned with diagramming than with analyzing developed architectural solutions. Archi includes built-in validation for EA models, whereas this technique just checks for inconsistent modeling element usage (missing connections, duplicated or unused elements).

Problem statement

In general, applications architecture modeling is a phase of requirements analysis in the software development life cycle (SDLC). As a result, the following relevant activity inputs and outputs have been discovered and displayed on the context IDEF0/SADT (Structured Analysis and Design Technique) diagram shown below (Fig. 1).

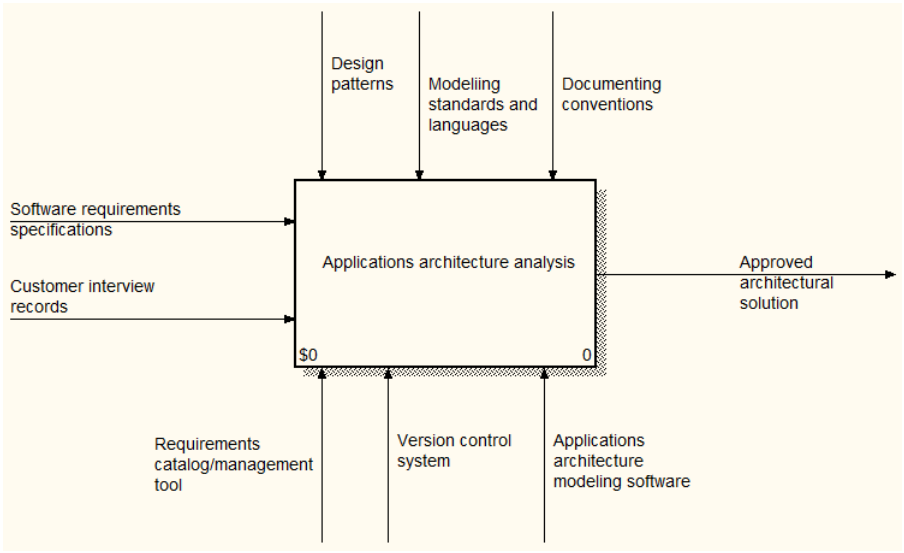

Fig. 1. The context diagram of the applications architecture models analysis process

The context diagram shown above provides a basic overview of the entire applications architecture models analysis and how the software system that should be implemented can support this activity. Clearly, the proposed activity is a complex business process comprised of the following sub-activities: analyze requirement specifications, create blueprints for applications architecture, model and analyze applications architecture, brainstorm to identify architectural flaws and bottlenecks; provide a solution to the product owner.

Fig. 2 shows the decomposed IDEF0 diagram of the applications architecture models analysis process.

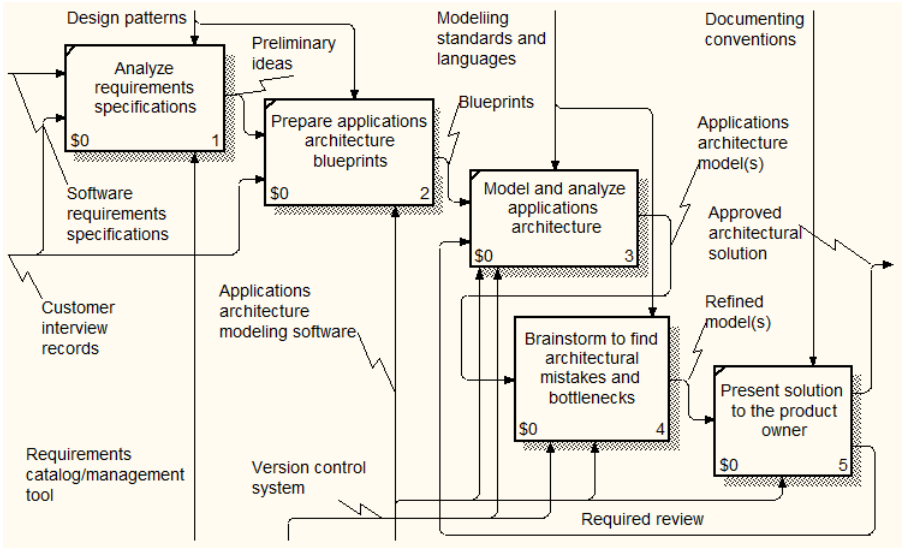

Fig. 2. The decomposition diagram of the applications architecture models analysis process

As shown on the context diagram (see Fig. 1) and the IDEF0 decomposition diagram (see Fig. 2), the following inputs and outputs are used by and produced by the considered business process:

– the software requirements specifications and other customer interview records are used as inputs for the preliminary design of the applications architecture;

– design patterns, modeling standards and languages, as well as documenting conventions for applications architecture modeling are used to retrieve the architectural solution;

– all of the considered activities are carried out using: a requirements catalog or even management system (e.g. Jira Software for large agile projects), applications architecture modeling software (e.g. ArchiMate or Visual Paradigm), and the version control system (usually decentralized are used now, such as Git).

Using the business process model (Fig. 2), it was discovered that the sub-process associated with brainstorming architectural mistakes and bottlenecks is a bottleneck of the applications architecture models analysis process. Following the completion of such an activity, the refined architectural solutions are demonstrated to the product owner or another customer's representative. This means that step 4 (see Fig. 2) will need to be improved with the addition of a specialized software solution for analyzing applications architecture models.

Fig. 3 depicts the IDEF0 decomposition diagram of the considered activity (step 4) with the introduced software solution to be developed.

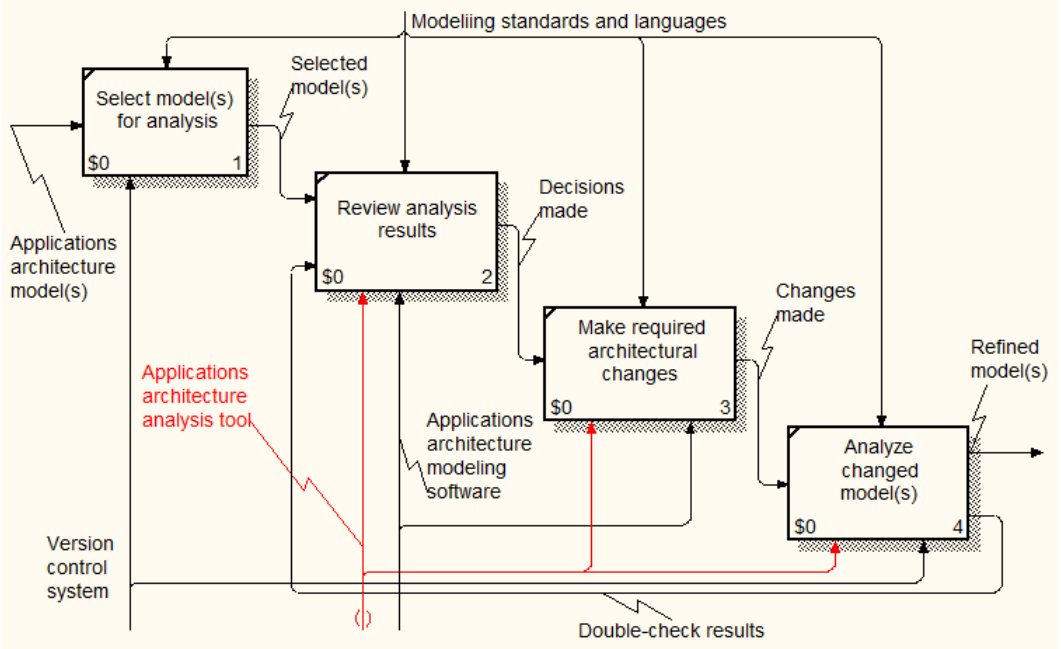

Fig. 3. Decomposition diagram of the brainstorming sub-process to find architectural mistakes and bottlenecks

The sub-process, as depicted in the decomposition diagram (Fig. 3), includes the following tasks:

1) choose an applications architecture model (or models) for analysis;

2) review analysis results (after the analysis procedure is completed);

3) make necessary architectural changes (based on decisions made when analysis results were obtained and taken into consideration);

4) analyze changed applications architecture models (in order to check whether they require additional analysis and changes).

The special arrow in red indicates that the software solution for applications architecture models analysis will be used to obtain initial recommendations after analyzing the original model, as well as to perform double-check analysis of models modified based on analysis results.

Proposed approach to applications architecture models analysis

The ArchiMate applications architecture model represents a graph data structure that is widely used in computer science and its applications in the domain of software engineering [9]. For graph-based data structures, such as ArchiMate applications architecture models, link analysis (or network analysis) methods commonly used in system analysis can be used to evaluate system structure and draw conclusions about its properties and features.

In mathematics (graph theory), a graph is defined as a structure that represents a collection of objects, with some pairs of objects interconnected. Objects that correspond to some terms, concepts, or other mathematical abstractions are referred to as “vertices” (also known as nodes or points), and each connected pair of vertices is referred to as a “edge” (also sometimes called links, arcs, or lines). A graph is typically represented in schematic (visual or rather graphical) form as a set of points or circles for vertices connected by arcs (if the links are directed) or edges [10].

For example, a segment of an applications architecture model created with the ArchiMate language (see the upper part of Fig. 4) could be represented using the directed graph shown below (see the bottom part of Fig. 4).

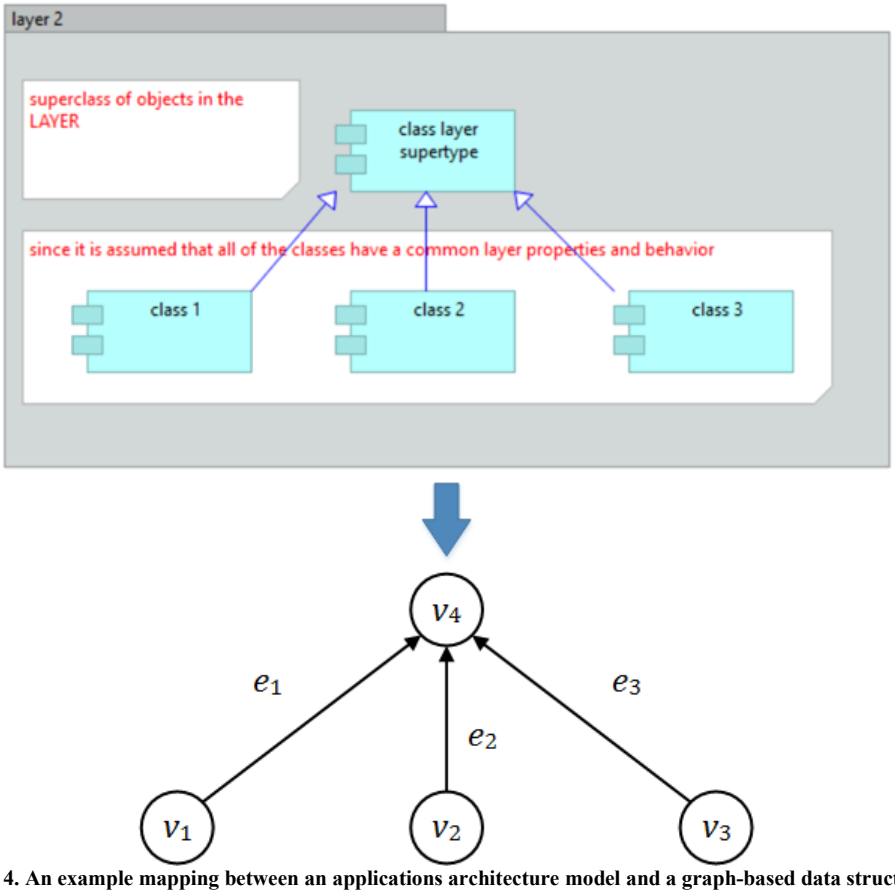

Fig. 4. An example mapping between an applications architecture model and a graph-based data structure

The most basic and elementary measures of directed graphs are the following [10]:

– $|V|$ – is the number of vertices in a graph G;

– $|E|$ – is the number of edges in a graph G;

– $d_G(v)$ – is the degree of a vertex in a graph G, which represents the number of incident edges (vertex degree is sometimes referred to as "valency");

– $d_G^{in}(v)$ – is the in-degree of a vertex in a graph G, which means the number of incoming edges;

– $d_G^{out}(v)$ – is the out-degree of a vertex in a graph G, which means the number of outgoing edges.

Using these elementary measures of a directed graph built on the basis of the ArchiMate applications architecture model, the following system analysis measures could be calculated:

1) connectivity:

$$A_G^{\Sigma} = \frac{1}{2} \cdot \sum_{v \in V} d_G(v); \quad (1)$$

2) resilience:

$$R_G = \frac{1}{2} \cdot \sum_{v \in V} d_G(v) \cdot \frac{1}{|V|-1} - 1; \quad (2)$$

3) centrality:

$$C_G = \frac{\sum_{v \in V} \left[\max_{v \in V}\{d_G(v)\} - d_G(v)\right]}{(|V|-1)\cdot(|V|-2)}; \quad (3)$$

4) irregularity:

$$\varepsilon_G = \sqrt{\sum_{v \in V} \left[d_G(v) - \frac{2 \cdot |E|}{|V|}\right]^2}. \quad (4)$$

Using these measures (1 – 4), the following n-space vector $X = (x_1, x_2, \dots, x_n)$ [11] could be obtained as a "footprint" or "image" of a specific ArchiMate applications architecture model. In our case, the elements of vector X are determined by the previously discussed measures, hence, $n = 4$. Consequently, applications architecture is based on design patterns. Obviously, those patterns, like any other pattern, have advantages and disadvantages. Patterns are used implicitly when designing applications architecture. They are mostly derived from the knowledge and experience of system engineers. However, architectures that appear to be faultless on "blueprints" may contain hidden threats during software implementation, testing, and even maintenance.

As a result, in order to avoid extra costs and efforts in the late SDLC phases, it is necessary to recognize design patterns within created applications architecture models [12]. When compared to a specific applications architecture model represented also using the n-space vector $S = (s_1, s_2, \dots, s_n)$ [11], design patterns described as

images X could be recognized using similarity measures. Since we need to check a given model for matching with multiple patterns, the set of design pattern images $P = \{X_1, X_2, \dots, X_m\}$, where m is the size of the design pattern collection used as "ideal" images, could be considered [11].

It is proposed to use the Euclidean distance to assess the similarity of design pattern images and applications architecture models under consideration [11]:

$$D(S, X_i) = \sqrt{\sum_{j=1}^{n} (x_{ij} - s_j)^2}. \quad (5)$$

Obviously, other similarity and distance measures could be used instead of (5), but for the current proof-of-concept implementation, stopping at the most well-known Euclidean distance measure is sufficient.

The pattern with the greatest similarity, i.e. the shortest Euclidean distance, should be used as the reference:

$$X^* = \arg\min_{X_i, i=\overline{1,m}} \{D(S, X_i)\}, \quad (6)$$

where X^* – is the pattern chosen to serve as a reference to provide suggestions.

Applications architecture design patterns discovered using (6) should be used to generate such recommendations based on the benefits and drawbacks of these patterns. Models of applications architecture describe software architecture as a system of interconnected application components [13].

Therefore, the following system design patterns [14], [13] can be considered:

1) if applications architecture model is the most similar to the sequential pattern, then following recommendation should be obtained "Development: easy, Cost: inexpensive, Flexible: yes, Reliability: moderate, Extension: easy, Robust: no";

2) if applications architecture model is the most similar to the ring pattern, then following recommendation should be obtained "Development: difficult, Cost: moderate, Flexible: no, Reliability: high, Extension: easy, Robust: no";

3) if applications architecture model is the most similar to the radial pattern, then following recommendation should be obtained "Development: easy, Cost: expensive, Flexible: yes, Reliability: high, Extension: easy, Robust: yes";

4) if applications architecture model is the most similar to the tree pattern, then following recommendation should be obtained "Development: easy, Cost: moderate, Flexible: yes, Reliability: high, Extension: easy, Robust: no";

5) if applications architecture model is the most similar to the mesh pattern, then following recommendation should be obtained "Development: difficult, Cost: expensive, Flexible: no, Reliability: moderate, Extension: difficult, Robust: yes".

All of the patterns under consideration should then be accompanied by their respective image vectors for use in applications architecture model analysis – $P = \{X_1, X_2, \dots, X_n\}$. Recognized design patterns should then be demonstrated with their obtained similarity measure values, each of which demonstrates the closeness or vice versa – incompatibility (for those patterns that were unexpectedly suggested as the analysis results) of the designed applications architecture model with the best practices. Such "suspicious" cases should then be displayed to a user (e.g., a system architect or analyst) to aid in the decision-making process for the architecture design of the software solution.

Design and development of the software solution

Let us now describe the proposed architectural solution's structure. We can begin with a description of IT infrastructure as the foundation for all proposed software systems. This layer includes the following structure elements:

– VCS (Version Control System) repository, which is a catalog in a file system that is under VCS control, and all changes in this catalog are tracked by VCS software, such as Git or Subversion;

– VCS file system, which is a built-over traditional file system that extends it with specific operations for version control management of stored files in a repository; (e.g. commit, push, pull, clone operations etc.).

The following behavior elements of the proposed architectural stack's IT infrastructure are:

– the process of software (and user) interaction with the VCS repository of applications architecture models;

– the services that VCS repository capabilities offer to software and users for accessing stored applications architecture ArchiMate models.

ArchiMate architecture models are stored as XML (eXtensible Markup Language) files in the OMG (Object Management Group) open exchange file format as information elements of the IT infrastructure of the depicted architectural solution.

The following structural architecture elements of the proposed architecture solution belong to the applications layer:

– Microsoft Power BI application (free desktop version) used to demonstrate analysis results as the user-interface solution for end users;

– software component (it is planned to create Java enterprise web application) that implements applications architecture models analysis approach considered before.

According to the proposed solution, behavioral software application elements are:

– the process of applications architecture models analysis;

– services of applications architecture models analysis provided to system analyst or system architect.

Java objects, also known as Java Beans, are information elements of the depicted architectural solution's applications layer. The proposed solution's business architecture layer depicts the activities of end users (system analysts or system architects), who should obtain recommendations for applications architecture model improvement via analytical reports to which Java objects are translated and displayed via Power BI. The developed solution's software architecture should be client-server in nature.

Spark Framework, a simple and expressive Java web framework built for rapid development, was used to implement web-based API (Application Programming Interface). Spark's goal is to provide an alternative for Java developers who want to create web applications that are as expressive as possible while using as little boilerplate as possible. Spark Framework is designed with a clear philosophy to not only make you more productive, but also to improve your code under the influence of Spark's sleek, declarative, and expressive syntax [15]. The Java application can pass JSON objects to the Power BI user interface using the Spark Framework.

Power BI is a Business Intelligence (BI) and Data Visualization tool that converts data from various sources into interactive dashboards and BI reports. Power BI suite includes a variety of software, connectors, and services, including Power BI desktop, SaaS-based Power BI service, and mobile Power BI apps for various platforms. Business users use this set of services to consume data and create BI reports [16]. Because of these features and benefits, Power BI was chosen to implement a user interface that could contain applications architecture models analysis reports that could be shown to software and system architects.

Fig. 5 depicts the software components UML diagram of the software solution for applications architecture models analysis.

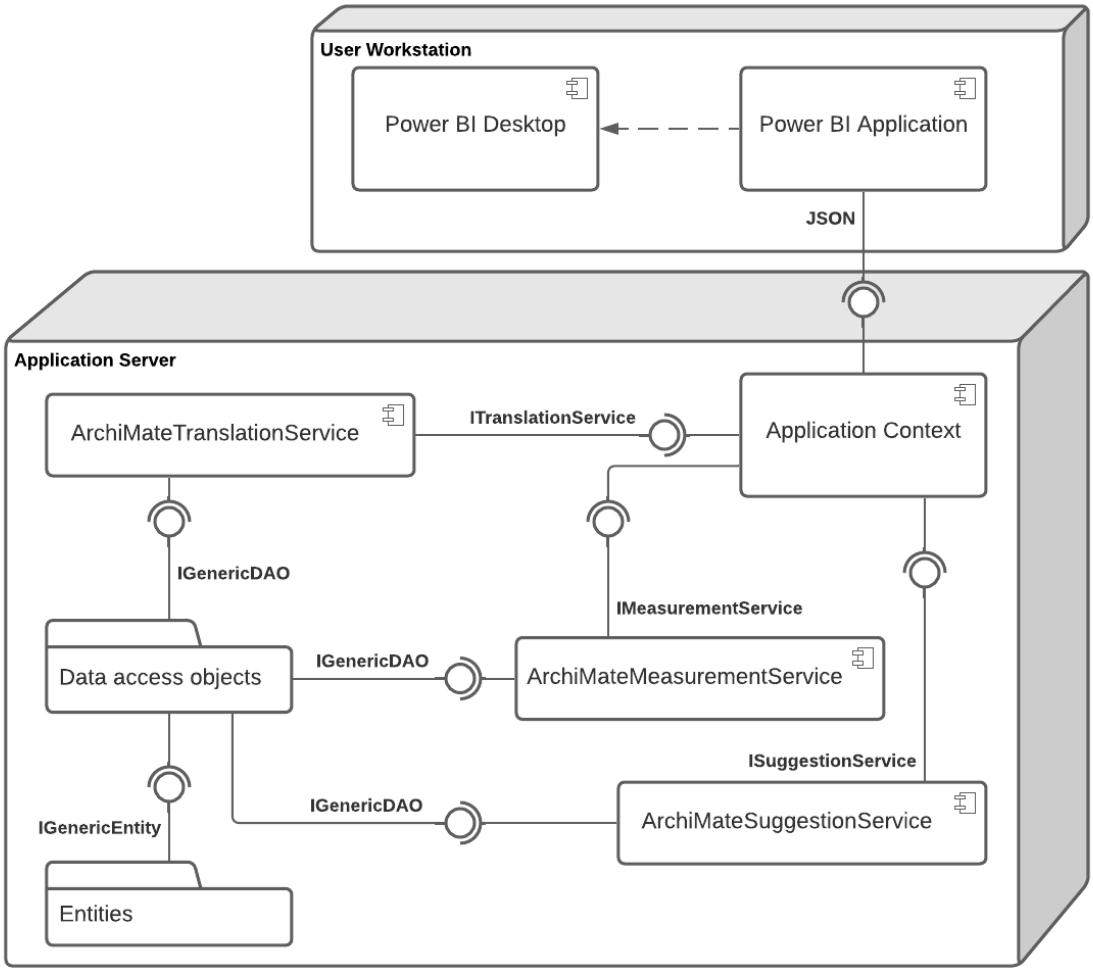

Fig. 5. The software solution's component diagram

The software solution is comprised of three major components, as shown in the diagram above (Fig. 5):

1) "ArchiMate Translation Service" component, which is in charge of extracting ArchiMate files from VCS repositories and translating them into Java objects for further processing;

2) "ArchiMate Measurement Service" component, which is in charge of calculating metrics (defined in section 2.1) for processed applications architecture models, the structure of which is already represented as Java objects;

3) "ArchiMate Suggestion Service" component, which is in charge of calculating the distance between applications architecture models and pre-defined structure patterns of systems analysis and producing appropriate recommendations.

As shown in Fig. 5, the considered software components rely on data access objects that are in charge of managing the data store and maintaining collections of Java objects that describe the structure of processed ArchiMate models. The user's workstation only contains the Power BI software and the respective analytical reporting application, which connects to the Java application via the JSON-based web API implemented with the Spark Framework.

Applications architecture models analysis using the software solution

To test the operability of the developed software solution, the following set of applications architecture models were chosen and presented in ArchiMate language, while originally belonging to the resource "Catalog of Patterns of Enterprise Applications architecture: Web Presentation Patters" [17]. Given the dominance of web applications today, it is important to detect the system design opportunities or threats of specific web development patterns.

Fig. 6 depicts the first part of the ArchiMate language's applications architecture models.

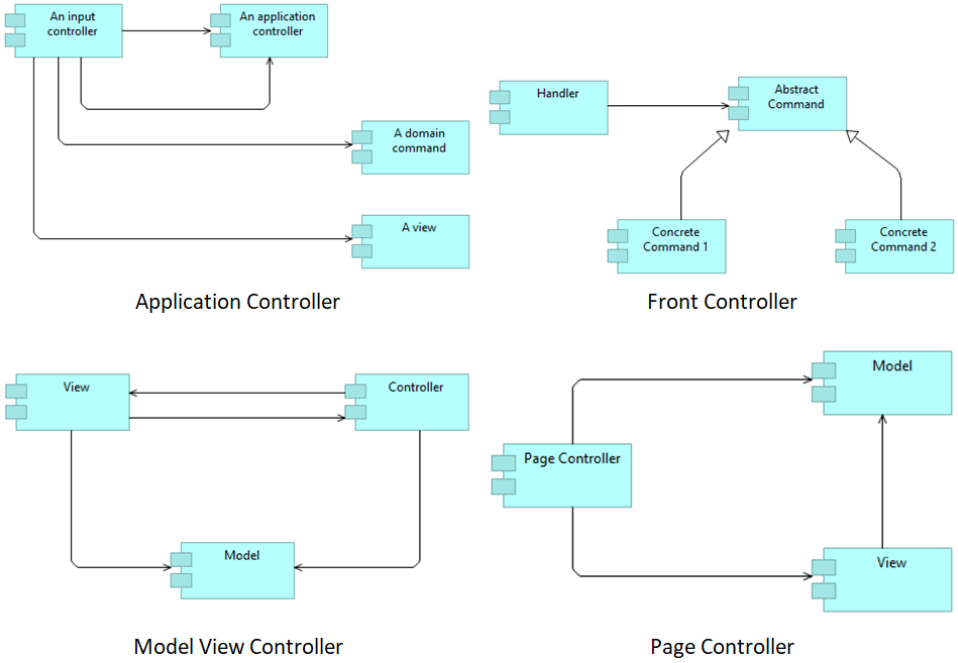

Fig. 6. The collection of applications architecture models that are used for analysis (first part)

Fig. 7 depicts the second part of the ArchiMate applications architecture models that have been prepared for analysis.

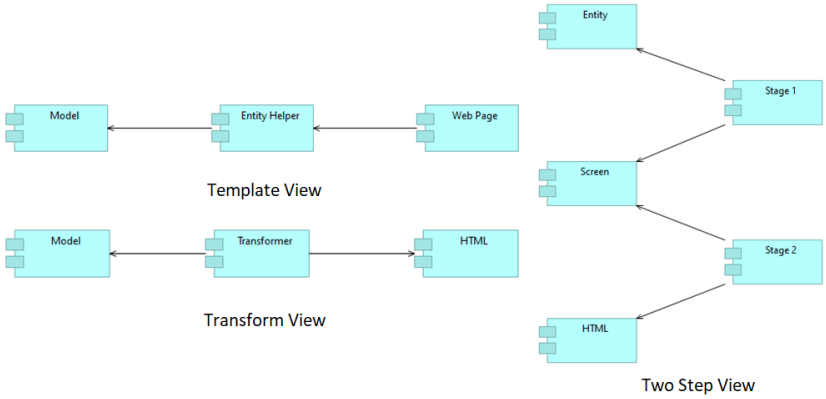

Fig. 7. The collection of applications architecture models that are used for analysis (second part)

Despite the fact that these models appear to be quite simple, they demonstrate essential web enterprise application development approaches, which analysis may be used by system or software architects to avoid defects and vulnerabilities in the future by making decisions during the design stage rather than fixing errors during the testing or even maintenance phases of the project's lifecycle.

The recommendations on web enterprise application solutions to use should be provided. Let us start with the models that have flaws and are not recommended for use:

– "Application Controller" is costly to maintain, despite its robustness;

– development of a "Page Controller" solution is difficult, and it is not flexible;

– "Template View", "Transform View", and "Two-Step View" are inexpensive to maintain, but they provide moderate reliability.

As a result, "Model View Controller" (see Fig. 8) and "Front Controller" (which are sometimes considered as part of the "Model View Controller" solution) could be recommended as better solutions for enterprise applications architecture design, where high reliability and flexibility, moderate cost of maintenance, and ease of development and extension are important system design features despite a lack of robustness.

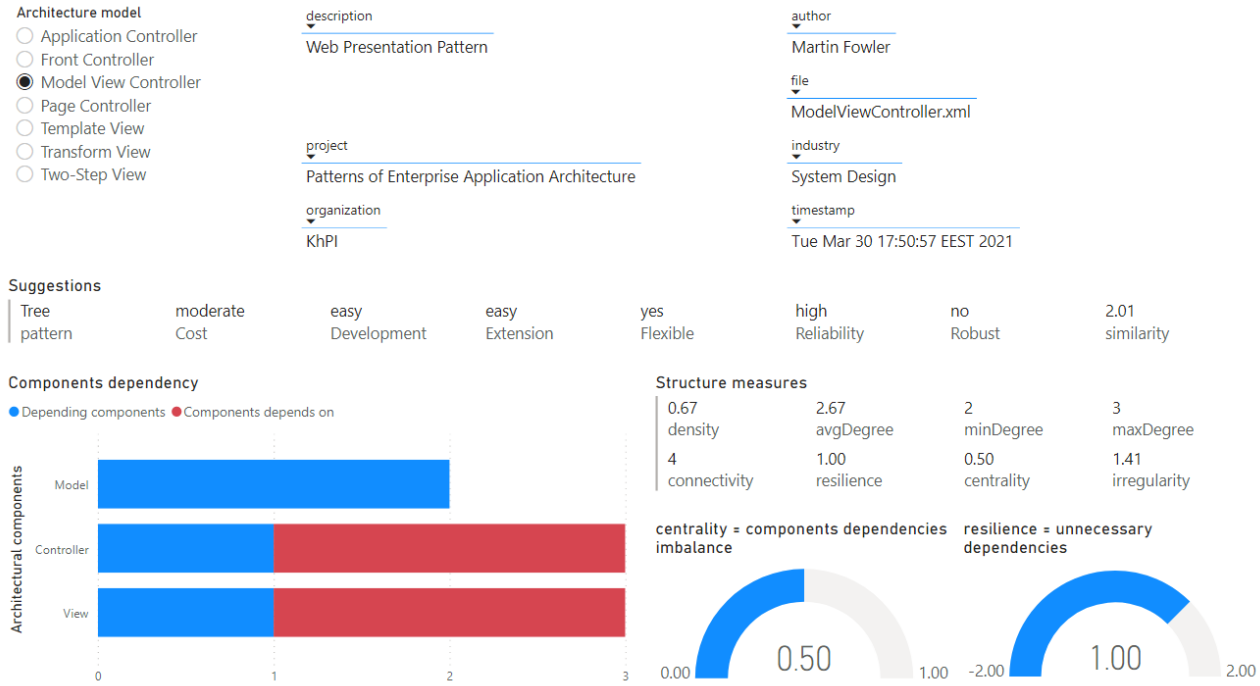

Fig. 8. Obtained results for the "Model View Controller" model

The obtained results (see Table 1) are supported by nearly two decades of "Model View Controller" (MVC) dominance in enterprise application development as a result of its concept of never combining data and presentation.

Table 1

Detailed results for the "Model View Controller" model

Pattern	Cost	Development	Extensibility	Flexibility	Reliability	Robustness	Distance
Tree	Moderate	Easy	Easy	Yes	High	No	2.01
Connectivity		Resilience		Centrality		Irregularity	
4		1.00		0.5		1.41	

A plethora of web enterprise frameworks are either completely based on or can support MVC principles of application design and development.

Conclusions

In this paper, a relevant problem of applications architecture model analysis was considered. Its significance is defined by the fact that designed blueprints of software systems should be carefully checked for all potential inefficiencies in order to avoid extra effort and costs for defect correction in later project stages. As a result, the research goal was defined as detecting strong and weak points in software design solutions through the analysis of applications architecture models. The process of analyzing applications architecture models was designated as the research objective, and the software solution for analyzing applications architecture models was designated as the research subject.

Existing software tools that support applications architecture modeling and analysis were defined based on an analysis of general software development problems for applications architecture models analysis. The ArchiMate enterprise architecture modeling language was chosen as the standard representation of applications architecture models to be analyzed. The following tasks were completed in order to achieve the study's goal:

1) the problem domain of applications architecture models analysis was discovered;

2) the approach to analyzing applications architecture models was proposed;

3) the software solution for analyzing applications architecture models was designed and developed;

4) the software solution was used to examine applications architecture models that represent web development patterns.

The results of the analysis could be used by system or software architects to estimate the suitability of applications architecture solutions for ongoing projects, detect weak points in certain architectural patterns, and reduce effort and costs in later project stages.

In the future, this approach and software tool should be expanded to analyze not only applications architecture, but also remaining domains of enterprise architecture, such as business architecture (organizational structure, business processes, etc.) and technological architecture (IT infrastructure including system software, hardware etc.). In addition, the set of system design measures could be expanded in the future, and alternative distance measures could be used.

References

1. Minoli D. Enterprise Architecture A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology / D. Minoli. – Boca Raton: Auerbach Publications, 2008. – 498 p.
2. Using ArchiMate and TOGAF to Understand the Enterprise Architecture and ITIL Relationship / M. Vicente et al. // Lecture Notes in Business Information Processing, 2013. – 148. – P. 134-145.
3. Products In Enterprise Architecture (EA) Tools Market // https://www.gartner.com/reviews/market/enterprise-architecture-tools, 10.03.2021.
4. Brocke J. Handbook on Business Process Management 2: Strategic Alignment, Governance, People and Culture / J. Brocke, M. Rosemann. – Springer, 2014. – 865 c.
5. ArchiMate Tool // https://www.visual-paradigm.com/features/archimate-tools/, 10.03.2021.
6. Online ArchiMate Software // https://ralpha-garcia.medium.com/online-archimate-software-93a29edaab4b, 12.03.2021.
7. Free ArchiMate Modeling Tool Archi // https://mikethearchitectblog.wordpress.com/2011/01/07/free-archimate-modeling-tool-archi/, 12.03.2021.
8. ArchiMate® Model Exchange File Format for the ArchiMate 3.1 Modeling Language // http://www.opengroup.org/xsd/archimate/, 20.03.2021.
9. Kopp A., Orlovskyi D., Ersoyleyen D. An approach to analysis of ArchiMate applications architecture models using the software coupling metric // Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies. – 2021. – No. 2 (6). – P. 67-72.
10. Trudeau R. Introduction to Graph Theory / R. Trudeau // Courier Corporation, 2013. – 224 p.
11. Sharma M. Analysis of Distance Measures in Content Based Image Retrieval / M. Sharma, A. Batra // Global Journal of Computer Science and Technology, 2014. – 14, No. 2. – P. 1-7.
12. Kopp A. M., Orlovskyi D. L., Ersoyleyen D. Applications architecture analysis based on design patterns and image recognition // MicroCAD-2021, 2021. – P. 14.
13. Kopp A., Orlovskyi D., Ersoyleyen D. General Issues of Applications Architecture Domain Design // Texas Journal of Multidisciplinary Studies, 2021. – No. 1 – P. 106-112.
14. Bisht N. Analytical study of different network topologies / N. Bisht, S. Singh // International Research Journal of Engineering and Technology, 2015. – 1, No. 2. – P. 88-90.
15. Spark // https://sparkjava.com/, 24.04.2021.
16. Power BI // https://www.tutorialspoint.com/power_bi/index.htm, 26.04.2021.
17. Catalog of Patterns of Enterprise Applications architecture // https://martinfowler.com/eaaCatalog/, 20.01.2021.

https://doi.org/10.31891/csit-2022-2-4
UDC 621

Bohdan HUNKO
Lviv Polytechnic National University

HARDWARE AND SOFTWARE SYSTEM OF LIGHT VISUALIZATION OF SOUND SIGNALS

The paper proposes a hardware and software system for light visualization of sound signals. The paper goes through performance requirements, system design process and practical solutions for audio visualization. The paper differs from the present solutions because it not only shows the end results, but also goes through design process, decision-making and performance measurements.

There are many methods that are practically used for audio visualization: amplitude visualization, spectral visualization, frequency visualization etc. One of the most interesting and common methods is spectral visualization of audio signals. This method is based on a mathematical model of obtaining the frequency spectrum of an audio signal using fast Fourier transform (FFT) and subsequent visualization of this spectrum.

The proposed solution is designed with high performance and low latency in mind and shows practical applications of hardware and software-based optimization techniques. Also, the paper describes several visualizations technics and gives an overview on possible visualization improvements.

Key words: hardware and software, performance, audio signal visualization, fast Fourier transform (FFT), analog signal processing.

Богдан ГУНЬКО
Національний університет «Львівська політехніка»

АПАРАТНО-ПРОГРАМНА СИСТЕМА СВІТЛОВОЇ ВІЗУАЛІЗАЦІЇ ЗВУКОВИХ СИГНАЛІВ

Системи опрацювання сигналів є надзвичайно важливою складовою сучасної електроніки. Саме аналогові сигнали є основним джерелом інформації про навколишній світ. Аудіо сигнали як різновид аналогових сигналів є надзвичайно поширеним способом передачі інформації.

Системи обробки та візуалізації аудіо сигналів займають значну частину сфери обробки сигналів. Візуалізація аудіо сигналів допомагає вирішувати багато практичних завдання та дозволяє людям краще усвідомлювати природу та властивості аудіо сигналів. Системи візуалізації аудіо сигналів можна зустріти як в професійних програмах, наприклад програмах для обробки аудіо треків, так і в програмах загального призначення, таких як наприклад аудіо плеєри.

Є багато способів які практично застосовуються для візуалізації аудіо: амплітудна візуалізація, спектральна візуалізація, частотна візуалізація та інші. Одним з найцікавіших та найбільш поширених способів є спектральна візуалізація аудіо сигналів. В основі цього методу лежить математична модель отримання частотного спектру аудіо сигналу за допомогою швидкого перетворення Фур'є (ШПФ) та подальша візуалізація цього спектру.

В цій статті розглядаються практичні методи обробки та візуалізації аудіо сигналів, способи оптимізації програмного забезпечення, критерії оцінювання продуктивності системи та способи досягнення необхідних системних параметрів. У процесі дослідження проведено аналіз існуючих методів обробки та візуалізації аудіо сигналів, визначено оптимальні параметри апаратного забезпечення та компонентів системи а також визначено технічні характеристики розробленої системи.

Результатом проведеної роботи стала програмно апаратна система для світлової візуалізації звукових сигналів. Створена система використовує мікроконтролер як апаратну базу для опрацювання сигналів та LED стрічку для світлової візуалізації спектральної характеристики сигналу.

Ключові слова: апаратне та програмне забезпечення, продуктивність програми, візуалізація звукового сигналу, швидке перетворення Фур'є (ШПФ), обробка аналогового сигналу.

Introduction

Demand on audio signal processing systems is rising each year. Systems that visually represent audio signal are very important part of current digital world because they expand the perception of audio signals which helps humans better understand the signal nature, simplify the process of audio editing and makes the exploration process easier.

There are several popular technics for audio visualization, the most popular being amplitude visualization and spectral visualization. This article goes into details of spectral visualization of audio signals. The project uses a microcontroller unit (MCU) to process audio signal. MCU resources are limited, so a high level of optimization is required to ensure high system performance can be reached. Signal spectrum is calculated using Fast Fourier transform (FFT) [1]. After FFT calculation, spectral characteristic of the signal is visualized on the LED strip.

There are many parameters which impact system performance, most significant being: system architecture, FFT size, audio sampling frequency, complexity of visualization function and number of LEDs in LED strip. It is critical to make the architecture as optimized as possible, so this article describes the investigation on which settings can be used to result in a system which satisfies real time visualization requirements.

Related works

There are several works related to applications of fast Furrier transformation (FFT) on microcontrollers:

- Basics of FFT algorithm on MCU [2] describes the basic ideas and shows the results of FFT algorithm performed on MCU.
- Develop FFT apps on low-power MCUs [3] goes into math behind the FFT and gives a general idea on audio sampling and signal processing.
- Practical FFT on microcontrollers using Common Microcontroller Software Interface Standard (CMSIS) Digital Signal Processing (DSP) [4] talks about usage of DSP library, its benefits and possible problems.

However, described works does not provide solution for light visualization of audio signals, instead they give general idea on FFT algorithm, its usage and expected results. Also, described works does not go into topic of audio sampling, design of visualization system and its characteristic.

Therefore, the purpose of this study is to design the high performance, low latency system which will combine the FFT algorithm with audio sampling to prepare the data which will be visualized on LED strip.

Proposed methodology

There are several approaches that are used to visualize the audio signal, the most popular being amplitude visualization and spectral visualization.

Amplitude visualization isn't the best solution when visualizing audio on LED strips because amplitude of a signal at a given moment of time does not carry much information, for example playing the same song on different volume levels will result in different amplitudes even though the song is the same.

The better approach is to sample audio during some period of time, then apply fast Fourier transformation (FFT) [1] to it, which will convert the signal from amplitude domain into frequency domain, and then visualize frequency domain of the signal.

Microcontrollers are very limited in terms of available system resources, hence choosing the right system parameters and making the design performance and latency optimized is critical.

Following chapters describe what the design of such a system might be, which limitations are present and how to overcome them.

Choosing the hardware

There are several requirements when choosing the microcontroller for this project, most curtail being: fast CPU core to calculate FFT, availability of fast peripheral interphases to output the data to the LED strip, presence of high-performance ADC to sample audio signal. Also, presence of the DMA channels is preferable so audio buffering operations can be offloaded from the CPU core.

PSoC 64 (CY8CKIT-064B0S2-4343W) satisfies all the requirements described above and also provides user-friendly APIs alongside with great documentation, therefore this microcontroller was chosen as hardware platform for this project.

Here is a list of PSoC 64 hardware that is used in this project:

- 100 MHz ARM Cortex-M4 (CM4) core
- 12-bit ADC capable of sampling in speeds up to 2 Giga samples per second
- SPI interphase
- DMA channels

When choosing an LED strip, the choice fell on the WS2812B, as this type of strip has a relatively simple control interface, provides full coverage of RGB spectrum and has a low price compared to other types of LED strips, which makes the WS2812B ideal for this project.

Generally, other types of LED strips can be used, but one thing to note is that it is preferable to have a LED strip with fast response time and rapid data transfer speeds to minimize data transition time.

Overall system architecture

To satisfy high performance and low latency requirements, there are several hardware and software optimizations that need to be done.

Fig. 1 shows the overall system architecture for this project.

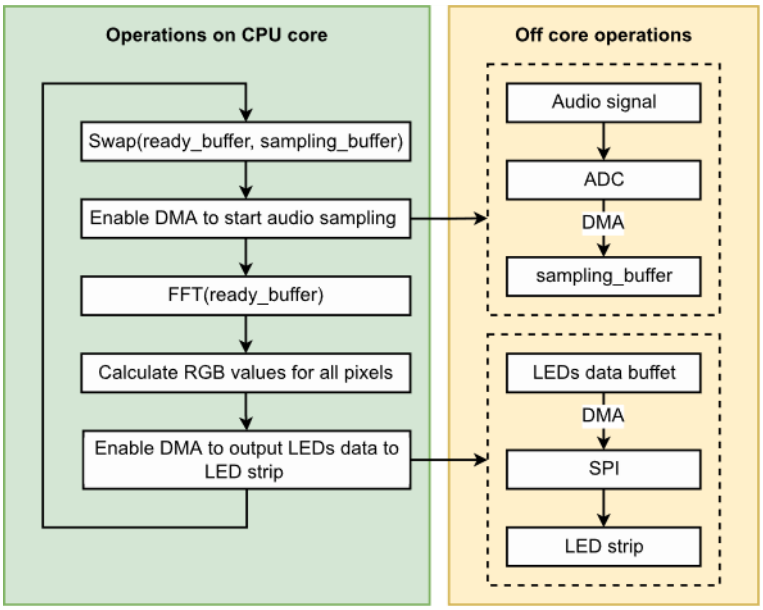

Fig. 1. Overall system architecture

As can be seen from the Fig. 1, a DMA channel is used to put ADC samples into audio buffer and another DMA channel is used to output LEDs data buffer through SPI to the LED strip. Such usage of DMA offloads a lot of work from CPU, which has noticeable positive impact on system performance.

Two buffers with same size are used for audio samples. While one buffer is getting processed by CPU, another buffer is being filled with ADC samples. This means that audio signal is sampled continuously and no data is lost as long as FFT and RGB values calculations are executed faster than new audio buffer is ready (later this article describes which system parameters will satisfy this requirement).

Utilization of DMA channels and usage of two buffers for audio signal effectively divides the system into 3 truly parallel tasks:

- FFT and RGB values calculation
- Audio sampling
- outputting LEDs data buffer to the LED strip

Fig. 2 shows execution timeline for these tasks and interactions between them.

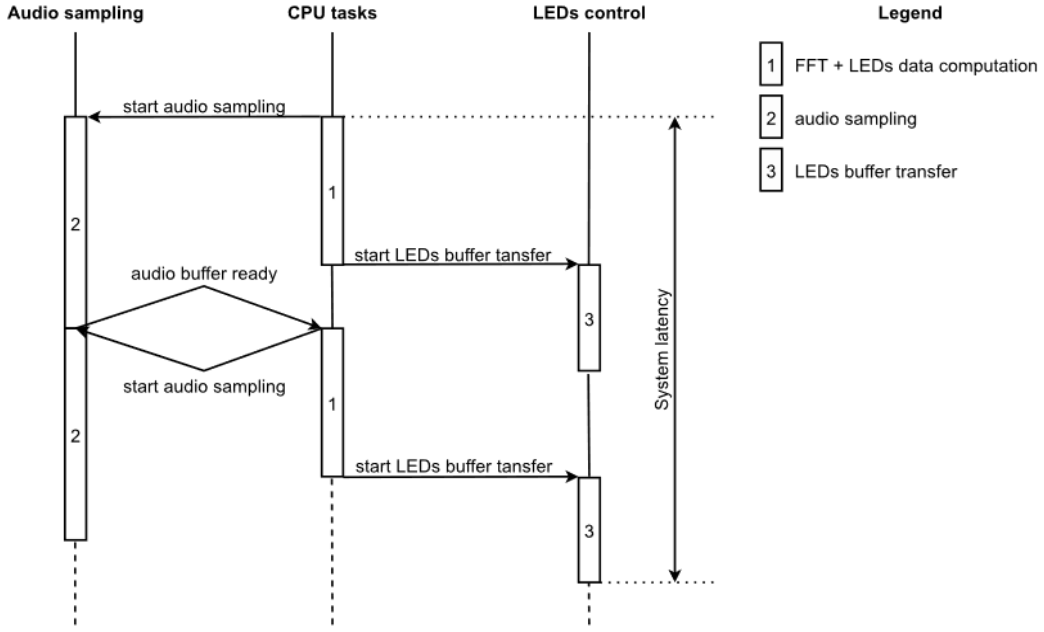

Fig. 2. Execution timeline for system tasks

Key system requirements can be formulated from Fig. 2:

- Requirement 1: Task 1 (FFT + LEDs data computation) must execute faster than Task 2 (Audio sampling). This requirement is needed to ensure that audio is sampled continuously without delays or wait time.

- Requirement 2: Task 3 (LEDs buffer transfer) must execute faster than Task 2 (Audio sampling). Again, this is needed to ensure that there are no delays in the system and audio is sampled continuously.
- Requirement 3: System latency should not exceed 45ms. Audio visualization can be treated as video signal, because visual image corresponds to audio (same as in video track), therefore video industry standards can be used to determine acceptable latency. Video industry standards [5] specify that audio should lead video by no more than 15 milliseconds and audio should lag video by no more than 45 milliseconds. Therefore, 45ms delay is used as largest acceptable latency.

Also, it is worth mentioning that semaphores are used to synchronize the tasks and ensure that the system behaves as expected.

System parameters optimization

As per Nyquist–Shannon sampling theorem [6] to capture full signal spectrum sampling rate must be at least twice the occupied bandwidth of the signal. From [7] it is known that humans can detect sounds in a frequency range from about 20 Hz to 20 kHz, so audio should be sampled at least at 40kHz frequency.

ISO/IEC 13818-3:1998 [8] define a set of frequencies for audio sampling: 44.1kHz, 48 kHz, 88.2 kHz, 96 kHz and 192 kHz. It is always a good idea to follow the industry standards, so this set of frequencies will be used in further calculations.

PSoC 64 ADC unit is capable of up to 2 GHz sampling rate in up to 12-bit resolution, so sampling audio fast enough should not be the problem.

Industry standards [9] shows that audio signal amplitude typically ranges between -1.228 V and +1.228 V. This amplitude range is not absolute because there is professional rated equipment (like studio microphones and audio systems) which is capable of outputting significantly higher voltage levels. Also, this range is maximum values at maximum volume, and if volume is lover the voltage will also be lower.

So, there is a tradeoff when configuring ADC:

- if ADC voltage boundaries are set high (e.g. from -3 V to +3 V) to account for professional equipment then discretization step will be quite big which will make ADC reading for similar amplitudes the same. So low or medium volume sound signal produced by consumer's electronic will be very limited in terms of values range which will result in poor samples quality.
- if ADC voltage boundaries are set low (e.g. from -1.2 V to +1.2 V) to only account for consumers rated electronics then discretization step will be smaller and low or medium volume sound will result in better samples quality. But in this case audio signal produced by professional rated equipment will produce voltage levels that are outside the range which will lead to poor samples quality.

This project is meant to be used with standard phone or computer (which is consumers rated electronic) therefore ADC is configured to be able to sample voltage range from -1.2 V to +1.2 V.

Having highly optimized implementation of FFT library is critical to ensure high system performance and low latency requirements are satisfied.

ARM provides FFT algorithm implementation as part of Common Microcontroller Software Interface Standard (CMSIS) Digital Signal Processing (DSP) library [10]. This library is highly optimized for Cortex-M processors and utilizes vector instructions to speed up FFT calculation, therefore this library perfectly suites the needs of this project, so it will be used to calculate the FFT.

The following input buffer sizes are supported by DSP FFT library: 32, 62, 128, 256, 512, 1024, 2048 and 4096 bytes.

FFT duration measurements are needed to choose the right buffer size.

Table 1

Performance measurements of FFT algorithm for each input buffer size

FFT size	**32**	**64**	**128**	**256**	**512**	**1024**	**2048**	**4096**
FFT duration [us]	44	89	164	351	731	1414	3061	6462

As previous mentioned PSoC 64 operates on 100MHz CM4 core therefore 1 us is 100 processor cycles. It should be noted that different CPU core might result in different cycles count as it may have different instructions set which may result in more/less optimized code, but I don’t believe that there will be a huge difference in cycles needed to calculate the FFT algorithm.

Table 1 will later be used to choose system parameters.

There are many ways to visualize the FFT, but a lot of them require 2-dimensional space for visualization. LED strip on the other hand is 1-dimensional. So graphing FFT like normal equalizers do will not work.

To solve this problem, this project uses LEDs color value to visualize the FFT. LED has 3 colors: red, green and blue, therefore FFT spectrum is divided into 3 intervals: low, medium and high frequency intervals. Each color then is assigned to frequency spectrum: red – low frequencies, green – medium frequencies and blue – high

frequencies. FFT values in each spectrum are averaged and mapped to value from 0 to 255, then these RGB values are shown on the LED strip.

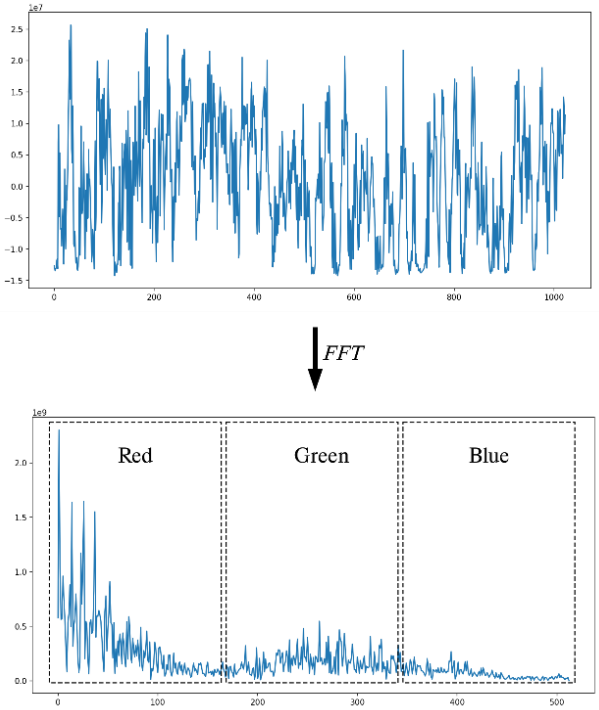

Fig. 3. FFT transformation results and their mapping to LEDs color

Visualization function utilizes functions from ARM DPS library to find the mean value of each of the intervals, this function uses vector instructions so overall visualization function does not take much time to execute.

Table 2

Performance measurements of visualization function

FFT size [bytes]	**32**	**64**	**128**	**256**	**512**	**1024**	**2048**	**4096**
FFT output size [bytes]	16	32	64	128	256	512	1024	2048
Visualization time [us]	17	18	21	27	38	61	107	199

Table 2 shows visualization function execution time for all possible FFT buffer sizes. Should be noted that an input buffer of size N produces FFT of size N/2.

Table 2 will later use these results to choose system parameters.

As mentioned previously, WS2812 is chosen as LED strip for this project, so this chapter describes the communication protocol of this LED strip.

WS2812 data sheet [11] shows that WS2812B LED strips are an almost arbitrary length string of pixels that can be cascaded together via a serial line.

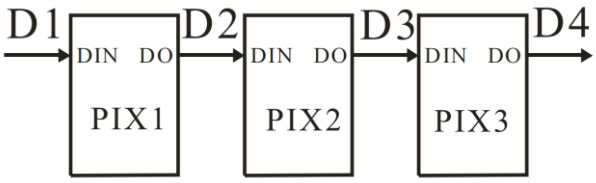

Fig. 4. LEDs cascading mechanism for WS2812 LED strip

In WS2812 each LED has 3 individually controlled diodes: red, green and blue. Each diode has 8-bit color range (0 - 255) which in total gives 3 * 8 = 24 bits (3 Bytes) per LED.

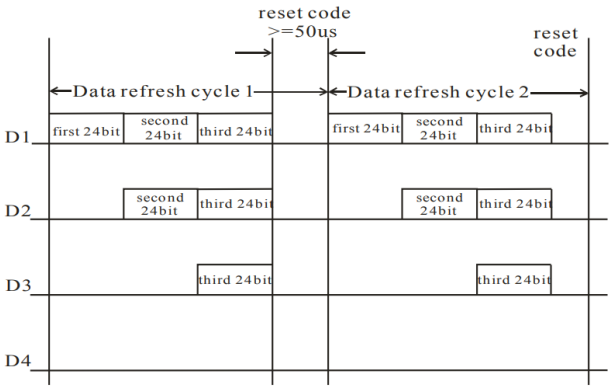

Fig. 5. Data transmission sequence for WS2812 LED strip

Fig. 5 shows that when communication starts, a pixel takes its Red, Green and Blue values from the data stream, then passes on the rest of the bytes to the next pixels.

G7	G6	G5	G4	G3	G2	G1	G0	R7	R6	R5	R4	R3	R2	R1	R0	B7	B6	B5	B4	B3	B2	B1	B0

Fig. 6. Composition of 24 bits of data for WS2812 LED strip

As can be seen from Fig. 6, WS2812 expects color codes in order of Green Reg Blue (GRB) which is different from RGB which everybody is used to.

Color codes sequence isn't the only weird thing in this communication protocol. WS2812 data sheet documents that a “bit 1” is actually encoded as a long pulse of 1 followed by a short pulse of 0. And a “bit 0” is a short pulse of 1 followed by a long pulse of 0. Fig. 7 shows data encoding for WS2812 LED strip and Table 3 shows exact timing requirements for these codes.

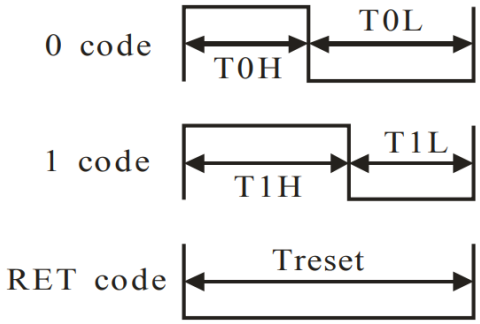

Fig. 7. Data encoding for WS2812 LED strip

Table 3

WS2812 data codes timing requirements

Name	Description	Required time [us]	Deviation [ns]
T0H	0 code, high voltage time	0.4	±150
T1H	1 code, high voltage time	0.8	±150
T0L	0 code, low voltage time	0.85	±150
T1L	1 code, low voltage time	0.45	±150
RES	low voltage time	Above 50	

As can be seen from the Table 3 T1H is double a T0H and a T0L is almost exactly double a T0H. That means that the fundamental unit of time in this system is 0.4uS. Therefore, WS2812 code for "1" can be encoded as 110 and code fore "0" can be encoded as 100.

So, there are 3 colors in one LED, each color is 8-bit value and each bit from that value is encoded as 3 bits in WS2812 protocol therefore 3*8*3 = 72 bits are needed to represent one WS2812 LED.

It is really hard to satisfy timing requirements from Table 3 when using CPU and delays to set the pin value to drive the LEDs. A much better way is to use serial peripheral interphase to transmit the data.

As described previously, bit transmission time for WS2812 is 0.4us, which means that transmission rate is 1/0.4uS = 2.5 Mbps. There are 3 types of peripheral interphases in PSoC: UART, I2C and SPI. Only SPI and I2C are capable of speeds that high. SPI does not require any pull-up resistors and also it can be easily used in custom data transmission rates, so in this project it is used as interphase to transfer the data.

Table 3 also shows that to update the LEDs, the data line must be pulled low for 50 us. SPI data line stays low while transfer is not in progress, so the only necessary thing is to ensure that there is at least 50 us pause between two consecutive transmissions.

Now let's calculate how long does it take to transfer the data for 1 LED. Formula (1) can be used to calculate data transmission time for 1 LED.

$$t_{1\_led} = \frac{bits\_per\_led}{transmission\ rate} = \frac{72}{2.5*10^6} = 28.8\ us \quad (1)$$

where *bits_per_led* is the number of bits per one WS2812 LED, *transmission_rate* is the transmission rate of SPI interphase.

From the Formula (1) it can be seen that it takes 28.8 us to transfer data for one LED, so let's see how much LEDs can be used and which number of Frames Per Second (FPS) can be achieved.

$$t_{n\_leds} = t_{1\_led} *N+50us = 28.8us *N+50us \quad (2)$$

where t_{1_led} is the data transmission time for 1 LED, N is the number of LEDs

Note that extra 50us in Formula (2) are the LEDs update window.

And Formula (3) can be used to calculate FPS.

$$FPS = \frac{1}{t_{n\ leds}} \quad (3)$$

where t_{n_leds} is the data transmission time for N LEDs.

From [12] it is known that human eye tolerates minimum of 24 FPS, but 30 FPS is more preferable target. Using Formulas (2) and (3) it can be calculated that at most 1150 LEDs can be used to produce FPS grater or equal to 30 FPS. WS2812 LEDs density is from 30 LEDs per meter up to 144 LEDs per meter, therefore 1150 is at least 7 meters (with 144 LEDs per meter) up to 38 meters (with 30 LEDs per meter).

This project uses 200 LEDs long WS2812 with 30 LEDs per meter, therefore using Formula (2) transmission time for 200 LEDs is:

$$t_{200\_leds} = 28.8us *200\ 50us = 5810us \quad (4)$$

t_{200_leds} will later be used to choose system parameters.

Let's put all system pieces together and determine which settings can be used to result in a high performance, low latency system.

System parameters should be chosen according to system requirements, so let's go through each system requirement and determine which parameters can be used to satisfy it.

Task 3 (LEDs buffer transfer) execution time is equal to transmission time for 200 LEDs. From Formula (4) it is known that transmission time for 200 LEDs (t_{200_leds}) is 5810us.

Task 1 (FFT + LEDs data computation) execution time is equal to sum of execution times of FFT calculation and visualization function duration. Combining the data from Table 1 and Table 2 gives the following results:

Table 4

Task 1 execution time

FFT size	**32**	**64**	**128**	**256**	**512**	**1024**	**2048**	**4096**
FFT duration [us]	44	89	164	351	731	1414	3061	6462
Visualization time [us]	17	18	21	27	38	61	107	199
Total [us]	61	107	185	378	769	1475	3168	6661

Task 2 (Audio sampling) execution time depends on sampling frequency and FFT buffer size. Formula (5) can be used to calculate Task 2 execution time.

$$t_{task\_2} = \frac{buffer\_size}{sampling\_frequency} \quad (5)$$

where *buffer_size* is the size of the audio buffer, *sampling_frequency* is the frequency used to sample audio.

Requirement 2 states that Task 3 (LEDs buffer transfer) must execute faster than Task 2 (Audio sampling). From Formula (4) it is known that Task 3 execution time is equal to t_{200_leds} (5810us), therefore Task 2 should take longer than 5810us to execute.

Requirement 3 states that system latency should be as low as possible and be less than 45ms. Fig. 2 shows that system latency is the sum of execution times for all three system tasks. Task 3 takes 5810us therefore to ensure that

the system latency is less than 45ms total execution time for Task 1 and Task 2 should be less than 45000-5810=39190us.

Knowing the argumentation for Requirements 2 and 3 heatmap for Task 2 execution time can be created.

		FFT size [bytes]							
		32	64	128	256	512	1024	2048	4096
Sampling frequency [kHz]	44.1	726	1451	2902	5805	11610	23220	46440	92880
	48	667	1333	2667	5333	10667	21333	42667	85333
	88.2	363	726	1451	2902	5805	11610	23220	46440
	96	333	667	1333	2667	5333	10667	21333	42667
	192	167	333	667	1333	2667	5333	10667	21333

Fig. 8. Task 2 execution time heatmap

From Fig. 8 can be seen that configurations that do not satisfy Requirement 2 are marked with red ( ) color and configurations that do not satisfy Requirement 3 are marked with orange ( ) color. Only the configurations that satisfy both Requirement 2 and Requirement 3 are marked with green ( ) color.

Also, from the Fig. 8 and Table 4 it can be seen that that regardless of FFT size and sampling frequency, Task 1 (FFT + LEDs data computation) always executes faster than Task 2 (Audio sampling). Therefore, system Requirement 1 (Task 1 (FFT + LEDs data computation) must execute faster than Task 2 (Audio sampling)) is always satisfied, regardless of settings.

So now it can be seen that all green values from Fig. 8 can be used to result in a system which will satisfy all system requirements. It is better to have audio sampling take a bit longer because then it will capture audio signal over a bigger period of time which will result in FFT which will better represent the signal.

Code and results

Considering everything said above, following system settings were chosen:

- LEDs count: 200
- FFT size 1024
- Audio sampling frequency 44.1 kHz

In this configuration, the system satisfies all the requirements and has a latency of: 5810us+1414us+23220us =30444us.

Overall there is some space for experiments with system parameters because as can be seen from Fig. 8 there are quite a few valid settings that can be used. If smaller latency is needed then smaller FFT buffer size can be chosen and alternatively if higher audio resolution is needed then higher sampling rate can be used.

Code for this project is licensed under GPL-3.0 license [13] and can be found in GitHub repository [14].

It is impossible to show the real time visualization in static paper, so I have prepared a video clip [15] and uploaded it to YouTube.

Conclusions

During this study, methods for light visualization of audio signals were developed. Ways of visualizing audio were investigated and described. System architecture for the project was developed and key system requirements were formulated. Using these requirements investigation on possible system parameters was done, this investigation has shown that there is a wide range of audio sampling frequencies, buffer sizes and LEDs quantities that can be used.

Overall, this article shows practical ways of using fast Fourier transformation on microcontrollers to obtain signal spectrum and methods that can be used for visualization of this spectrum.

There are a number of things to explore/improve for example add verity of visualization functions, add Bluetooth functionality and create a mobile app to allow the user to cycle through visualization modes, explore different hardware, visualize the audio on 2D array matrix of LEDs, etc.

References

1. Fast Fourier transformation: Wikipedia article. URL: https://en.wikipedia.org/wiki/Fast_Fourier_transform. (Accessed May 15, 2022)

2. DiCola T. Fun with Fourier Transforms: online article. URL: https://cdn-learn.adafruit.com/downloads/pdf/fft-fun-with-fourier-transforms.pdf. (Accessed May 15, 2022)

3. Develop FFT apps on low-power MCUs: online article. URL: https://www.embedded.com/develop-fft-apps-on-low-power-mcus/. (Accessed May 15, 2022)

4. Practical FFT on microcontrollers using CMSIS DSP: online article. URL: https://m0agx.eu/2018/05/23/practical-fft-on-microcontrollers-using-cmsis-dsp/. (Accessed May 15, 2022)

5. Audio-to-video synchronization: Wikipedia article. URL: https://en.wikipedia.org/wiki/Audio-to-video_synchronization. (Accessed May 15, 2022)

6. Nyquist-Shannon sampling theorem: Wikipedia article. URL: https://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem. (Accessed May 15, 2022)

7. Hearing range: Wikipedia article. URL: https://en.wikipedia.org/wiki/Hearing_range. (Accessed May 15, 2022)

8. ISO/IEC 13818-3:1998(en) Information technology: Generic coding of moving pictures and associated audio information Part 3: Audio: ISO standard. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:13818:-3:ed-2:v1:en. (Accessed May 15, 2022)

9. Line level: Wikipedia article. URL: https://en.wikipedia.org/wiki/Line_level. (Accessed May 15, 2022)

10. Digital Signal Processing using Arm Cortex-M based Microcontrollers: ARM online documentation. URL: https://www.arm.com/-/media/global/resources/education/textbooks/dsp-sample-chapter.pdf?revision=0a9768b9-0a7a-42fe-aba9-6304f240275b&la=en. (Accessed May 15, 2022)

11. WS2812 Datasheet: online device datasheet. URL: https://pdf1.alldatasheet.com/datasheet-pdf/view/553088/ETC2/WS2812.html. (Accessed May 15, 2022)

12. Bakaus P. The Illusion of Motion: online article. URL: https://paulbakaus.com/tutorials/performance/the-illusion-of-motion/. (Accessed May 15, 2022)

13. GNU General Public License: online copy of license. URL: https://www.gnu.org/licenses/gpl-3.0.en.html (Accessed May 15, 2022)

14. Hunko B. GitHub repository with code for light visualization of audio signals: GitHub repository. URL: https://github.com/hunkob/music_synched_LED_PSoC (Accessed May 15, 2022)

15. Hunko B. Visualization results: YouTube video. URL: https://youtu.be/rnnqywi4_kU. (Accessed May 15, 2022)

https://doi.org/10.31891/csit-2022-2-5
UDC

Vasyl YATSKIV, Volodymyr BODNAROVSKYI
Khmelnytskyi National University

RESEARCH OF METHODS OF ENERGY EFFICIENCY MANAGEMENT IN THE "SMART HOUSE" SYSTEM

The research of methods of energy efficiency management in the "Smart Home" system is carried out in the work. For researchers and practitioners, the problem of monitoring, estimating and reducing energy consumption by homes is an urgent task. According to the results of research, the method of determining user preferences has been improved, which has allowed to achieve more efficient use of energy. The method has also been further developed, which provides the ability to determine the number of people in the room and helps control oxygen levels.

Key words: smart home, energy efficiency, occupant-centric controls, residential buildings, HVAC, software and hardware, smart home management model.

Василь ЯЦКІВ, Володимир БОДНАРОВСЬКИЙ
Хмельницький національний університет

ДОСЛІДЖЕННЯ МЕТОДІВ УПРАВЛІННЯ ЕНЕРГОЕФЕКТИВНІСТЮ В СИСТЕМІ «РОЗУМНИЙ БУДИНОК»

В роботі проведено дослідження методів управління енергоефективністю в системі «Розумний будинок». Для дослідників і практиків проблема моніторингу, оцінки та зниження споживання енергії будинками є актуальною задачею. За результатами проведених досліджень удосконалено метод визначення вподобань користувача, що дозволило досягти більш ефективного використання енергоносіїв. Також набув подальшого розвитку метод, за допомогою якого забезпечується можливість визначення кількості осіб в приміщенні та допомагає контролювати рівень кисню.

Ключові слова: розумний будинок, енергоефективність, елементи керування, орієнтовані на мешканців, житлові будинки, ОВК, програмно-апаратний засіб, модель управління розумним будинком.

Introduction

Analysis of the research topic showed that many researchers consider ordinary apartment buildings (partly private houses) and do not consider commercial and industrial buildings, do not take into account the benefits and requirements for the level of comfort of different users. The considered methods of determining the occupancy of the premises will not allow to effectively predict the transfer of the user to other areas of the premises, which negatively affects the operation of the HVAC system and increase energy efficiency of the "Smart Home" system [1]. They also do not pay attention to the high cost of sensors to detect the presence and number of inhabitants in the controlled space of the room, which is quite relevant against the background of the crisis in the country [2]. The considered works use powerful hardware, which has high energy consumption and high cost, which does not allow to compete with ready-made analogues [3].

Therefore, it is necessary to use appropriate localization methods to distinguish between different occupants of apartments and offices, which are used by many people, and apply their benefits. An additional analysis of employment behavior (data processing) was then conducted, consisting of three stages to determine important employment characteristics, such as the number of residents present, periods of absence and presence, and other random changes in the profile of residents.

Typical meteorological data on the weather in the city are used to model the energy characteristics of the building. In Khmelnytsky warm and humid summers and cold winters. The room has five zones: north, east, center, south and west. In all areas except the central, the wall attached to the outer part has a window-to-wall ratio of 0.4. Ceiling height in all areas is 2.7 meters. Each node represents an area with similar environmental parameters (such as temperature, lighting and air velocity), such as a separate area or one of the layers of the wall. Heat is transferred between nodes by convection, conductivity or radiation.

Basics of the method of energy efficiency management in the "Smart Home" system

In order to achieve the goals of minimizing energy consumption in the building, as well as the hours of discomfort of residents, it is necessary to develop and invest in the optimization algorithm a detailed model of energy simulation of the house.

Since the functioning of building systems is highly dependent on the presence of residents, the integrated model should choose the most optimized parameters of building systems based on information about the dynamic occupation of space.

Using local control of building systems, the simulated space should be divided into several zones to assign appropriate dynamic information about the employment of each zone. Zoning is used to account for the effect:

1. Different activities performed in each zone.
2. Different number of HVAC terminals, as will be discussed in the next section.

3. Different orientation of the facade for the perimeter of the zones.

Knowing the location of a particular resident, the corresponding HVAC terminal and the corresponding light are regulated by local management strategies.

In this study, the proposed methodology includes two main modules: a multi-purpose modeling-based optimization module and a fill-in module, as shown in Figure 1.

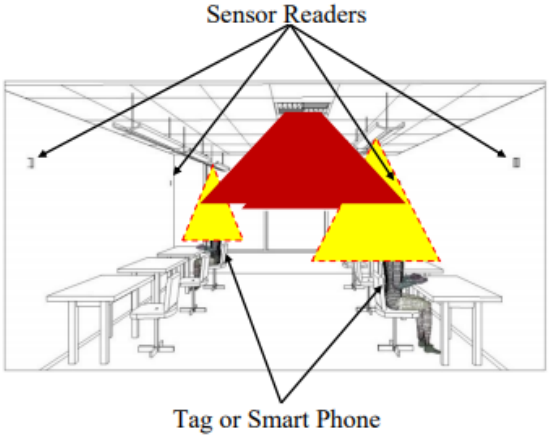

Fig. 1. Local control strategy with population monitoring

The occupancy module is used to determine specific dynamic profiles of residents based on their presence data, as shown in Figures 2 - 4. The main advantages of having dynamic occupancy profiles [4], which reveal the information of residents about his / her location and model of space use, are:

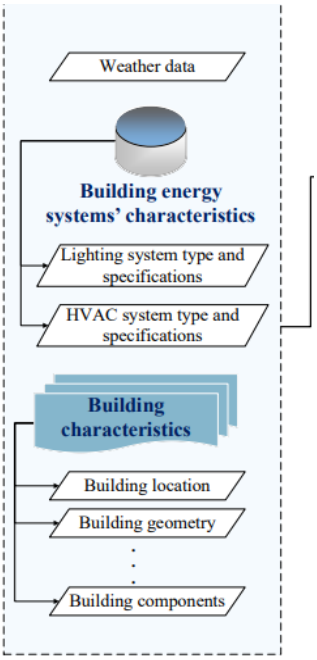

Fig. 2. The main modules of the framework

1. Unlike models that rely on averaging the behavior or schedules of different residents, dynamic occupancy profiles can cover the diversity of behavior of different residents, which is a very important factor in open planning offices.

2. Real-time monitoring and decision-making are the closest ways to imitate the real behavior of residents and their interaction with the energy-consuming systems of the building [5]. Dynamic occupancy profiles allow you to distinguish between schedules and habits of different residents.

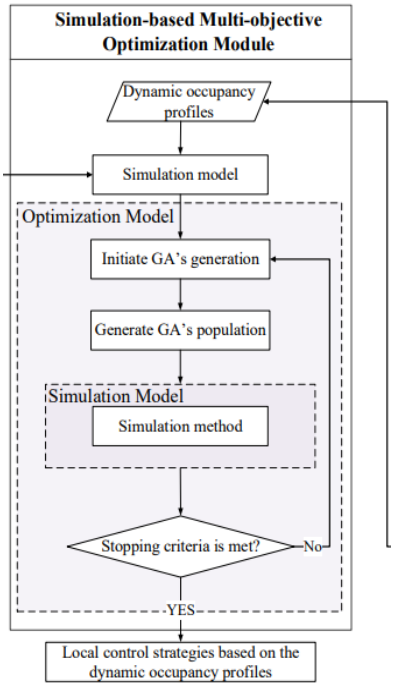

Fig. 3. The main modules of the framework

These profiles can be used to effectively apply the personal preferences of residents.

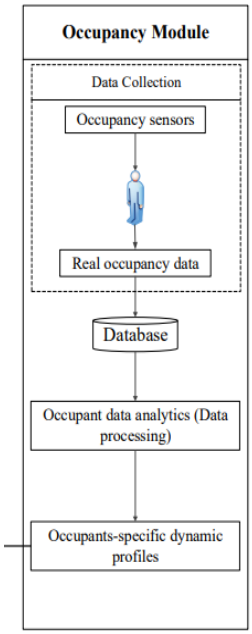

Fig. 4. The main modules of the framework

Experimental studies of the method of energy efficiency management in the "Smart Home" system

In each zone, four environmental parameters of artificial lighting, natural light, room temperature and ventilation rate are automatically monitored every hour [6]. Each zone is equipped with a variable air volume system (VAV), which provides heating, cooling and ventilation. During idle hours, energy management and integrated zone control are based on SOOP energy consumption [7]. The objective function of the SOOP method consists only of the term energy costs. For each zone, the total energy consumption (E_total) per hour is the sum of the energy consumption of artificial lighting, cooler, boiler and fan (1):

$$E_{total} = E_{lighting} + E_{cooler} + E_{boiler} + E_{fan}. \quad (1)$$

The term energy consumption in the target function of the SOOP method is the product of the price of electricity or gas and the associated hourly energy consumption. Fixed tariffs of UAH 1.68 / kWh and UAH 6.99 per

m3 are accepted as prices for electricity and gas in Ukraine [8]. For each hour of simulation, the term energy consumption in the objective function is (2):

$$\text{energy cost} = \left[\text{ElecPrice} \cdot \sum_{z=1}^{5} E_z^{\text{electricity}} + \text{GasPrice} \cdot \sum_{z=1}^{5} E_z^{\text{gas}}\right], \quad (2)$$

E is the consumption of electronic energy in kWh; z- number of the zone (room);
ElecPrice, GasPrice are the prices for electricity and natural gas.

There are many factors that determine the accuracy of the use model, including the identity of the occupants, the length of stay of the occupants, their location in different areas of the building, and their preferences [9].

New RTLSs can provide location and duration of presence, while benefits data can be collected through a simple survey. The fill module is used to define dynamic profiles related to residents [10].

The main advantages are:

1) dynamic employment profiles can cover the diversity of behavior of different residents, which is a very important factor in open planning offices;

2) real-time monitoring and decision-making that arise as a result are the closest ways to imitate the real behavior of residents and their interaction with energy-intensive building systems.

These profiles can be used to effectively apply the personal preferences of residents [11].

The operation of the proposed MOOP (Proposed Case) method and the SOOP (Base Case) method [12] on the thermal comfort of residents is compared.

The temperatures in the room of one-hour simulations in January and July, selected according to the proposed and base case, are compared (Fig. 5).

Optimal Pareto solutions are generated by changing the productivity factor ($ / h) [13].

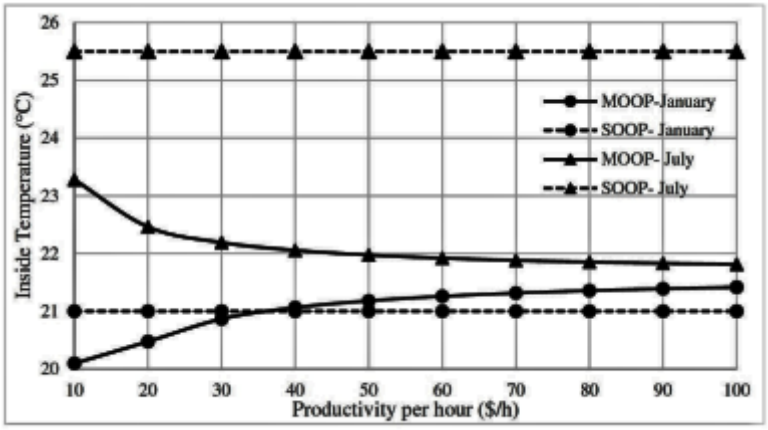

Fig 5. Indoor temperatures (ºC), selected according to the proposed case and the basic option in the eastern zone

In all three zones and for both external weather conditions with an increase in productivity per hour ($ / h) of residents, the average monthly room temperature moves to the maximum comfortable temperature (21.7 ºC). As collective productivity increases, the relative importance of residents 'productivity in terms of electricity costs increases, so the method reduces the relative loss of residents' productivity by approaching the maximum comfortable temperature. Here, the average monthly room temperature (ºC) is chosen to assess thermal comfort. An alternative approach is to demonstrate the frequency distribution of thermal sensory voices and to calculate the number of hours during which thermal sensory voices are in the comfort range. This approach is convenient when thermal comfort indicators such as the PMV index and the PPD index [14] are used to indicate the voices of thermal sensation.

Conclusions

Analysis of the research topic showed that many researchers consider common apartment buildings, do not take into account the preferences and requirements for the level of comfort of different users. The main contributions of this study are:

1) development of a method for obtaining information on population with different time steps from the collected RTLS employment data. This method can capture the different levels of resolution required to apply intelligent local government strategies;

2) development of a new adaptive probabilistic model of employment forecasting on the basis of the received information on employment;

3) development of time-dependent inhomogeneous filling model.

The proposed forecasting model is an adaptive model that evolves and improves over time.

The availability of a occupancy profile forecast for each resident leads to the development of a occupancy level forecast at the zone level.

The forecasting model can accurately estimate the location of residents during most data collection periods during the day. The high accuracy (86% and 68% on average for lighting and air conditioning and ventilation control, respectively) of forecasting employment models also indicates acceptable efficiency of the forecasting model.

References

1. Applied Energy Volume 283, 1 February 2021, 116251 url: https://doi.org/10.1016/j.apenergy.2020.116251

2. Sensors 2020, 20(2), 398; url: https://doi.org/10.3390/s20020398

3. Building and Environment Volume 187, January 2021, 107369 url: https://www.sciencedirect.com/science/article/abs/pii/S0360132320307381

4. Kolokotsa D. Stavrakakis G. Kalaitzakis K. Agoris D., "Genetic algorithms optimized fuzzy controller for the indoor environmental management in buildings implemented using PLC and local operating networks," *Engineering Applications of Artificial Intelligence,* vol. 15, no. 5, p. 417–428, 2020.

5. lcala R. Casillas J. Cordon O. Gonzalez A. Herrera F., "A genetic rule weighting and selection process for fuzzy control of heating, ventilating and air conditioning systems," Engineering Applications of Artificial Intelligence, vol. 18, no. 3, p. 279–296, 2015.

6. Macal C. North M., "Tutorial on agent-based modelling and simulation," *Journal of Simulation*, vol. 4, no. 3, p 151–162, 2018.

7. Joumaa H. Ploix S. Abras S. De Oliveira G., "A MAS integrated into home automation system, for the resolution of power management problem in Розумний будинокs," *Energy Procedia*, vol. 6, p. 786-794, 2015.

8. Hagras H. Callaghan V. Colley M. Clarke G., "A hierarchical fuzzy–genetic multi-agent architecture for intelligent buildings online learning, adaptation and control," *Information Sciences*, vol. 150, no. 1-2, p. 33-57, 2017.

9. Wilde D. Beightler C., Foundations of optimization, United States: Prentice-Hall, 2017.

10. Yang R. Wang L., "Multi-objective optimization for decision-making of energy and comfort management in building automation and control," *Sustainable Cities and Society*, vol. 2, no. 1, p. 1-7, 2018.

11. Wang Z. Yang R. Wang L., "*Multi-agent control system with intelligent optimization for smart and energy-efficient buildings*," IEEE, p. 1144–1149, 2020.

12. Dai C. Lan L., "Method for the determination of optimal work environment in office buildings considering energy consumption and human performance," *Energy and Buildings*, vol. 76, no. June 2014, p. 278–283, 2017.

13. Building SMART. (2018). Technical Vision. Retrieved from buildingSMART International home of openBIM: https://www.buildingsmart.org/siemens-join-buildingsmart-international/

14. Fanger P. O. Langklide G., "Interindividual differences in ambient temperatures preferred by seated persons," ASHRAE Transactions, p. 140-147, 2015.

https://doi.org/10.31891/csit-2022-2-6
UDC 681.5.015:004.681

Volodymyr KHOROSHKO
National Aviation University
Vadym KUDINOV
National Academy of Internal Affairs
Mariia KAPUSTIAN
Khmelnytskyi National University

EVALUATION OF QUALITY INDICATORS OF FUNCTIONING CYBER PROTECTION MANAGEMENT SYSTEMS OF INFORMATION SYSTEMS

Evidence of the complexity of the cybersecurity problem is the rapid increase in the number of information security breaches and losses on cybersecurity threats combined with an increase in the average loss from each of the breaches. Therefore, it is necessary to create requirements for a cybersecurity system that could provide more opportunities in the choice of methods in the management of the protection of automated information systems.

The task of determining the optimal quality indicators of information resource management systems of automated systems is one of the most important problems in designing integrated information security systems. This is due to the complexity of such systems, the presence of many variable parameters, and the complexity of calculating quality indicators. In addition, the determined quality indicators should not only ensure the optimality of the target function, but also the stability of the protection system in a wide range of external adverse effects. The problem is that the existing methods of calculating integrated quadratic estimates (IQE) do not take into account errors in determining quality indicators, as well as the vector nature of these indicators.

The aim of this work is to solve problems (development of algorithms), which are a problem of optimization of stable protection management systems using vector objective functions. Based on the model of information management system protection of information resources in the form of an automatic control system, the method of forming integrated quadratic estimates (IQE) of control error is proposed. This method takes into account the weights of the estimates at the desired installation time and standard transfer functions. Algorithms for calculating IQE according to the modified Katz formula and Ostrom's method for arbitrary order control systems are developed, including vector representation of the objective function of the protection system. The vector penalty function is proposed and the algorithm of its calculation is developed to display the degree of infringement of conditions of stability of parameters of the system of protection by the Rauss-Hurwitz criterion.

Key words: integrated quadratic estimates (IQE) of control error, arbitrary order control systems, vector objective functions, the algorithm for calculating the vector penalty function, and integrated information security systems.

Володимир ХОРОШКО
Національний авіаційний університет, Київ, Україна
Вадим КУДІНОВ
Національна академія внутрішніх справ, Київ, Україна
Марія КАПУСТЯН
Хмельницький національний університет, Хмельницький, Україна

ОЦІНКА ПОКАЗНИКІВ ЯКОСТІ ФУНКЦІУНУВАННЯ СИСТЕМ УПРАВЛІННЯ КІБЕРЗАХИСТОМ ІНФОРМАЦІЙНИХ СИСТЕМ

Підтвердженням складності проблеми кібербезпеки є досить швидке зростання кількості порушень інформаційної безпеки та витрат на загрози кібербезпеки, у поєднанні з ростом середнього збитку від кожного із порушень. Тому необхідно сформувати вимоги до системи кіберзахисту, яка б могла надавати більше можливостей у виборі методів в управлінні захистом автоматизованих інформаційних систем.

Завдання визначення оптимальних показників якості систем управління захистом інформаційних ресурсів автоматизованих систем є однією з найважливіших проблем проектування комплексних систем захисту інформації. Це зумовлено складністю подібних систем, наявністю безлічі варійованих параметрів, складністю обчислення показників якості. Крім того, визначувані показники якості повинні не лише забезпечувати оптимальність цільової функції, але і стійкість функціонування системи захисту в широкому діапазоні зовнішніх несприятливих впливів. Проблема полягає в тому, що існуючі методи обчислення ІКО не враховують помилки визначення показників якості, а також векторний характер цих показників.

Метою даної роботи є рішення завдань (розробка алгоритмів), що становлять проблему оптимізації стійких систем управління захистом при використанні векторних цільових функцій. На основі моделі уявлення системи управління захистом інформаційних ресурсів у вигляді системи автоматичного управління, запропоновано спосіб формування інтегральних квадратичних оцінок (ІКО) помилки управління, що враховує вагові коефіцієнти оцінок за бажаним часом встановлення і стандартні передавальні функції. Крім того, розроблені алгоритми обчислення ІКО за модифікованою формулою Каца і методом Острьома для систем управління довільного порядку, включаючи векторне уявлення цільової функції системи захисту. Запропонована векторна штрафна функція і розроблений алгоритм її обчислення для відображення ступеня порушення умов стійкості параметрів системи захисту за критерієм Рауса-Гурвиця.

Ключові слова: інтегральні квадратичні оцінки (ІКО) помилки управління, системи управління довільного порядку, векторні цільові функції, алгоритм розрахунку векторної штрафної функції, інтегровані системи захисту інформації.

Introduction

Creating and ensuring the functioning of an effective cybersecurity system is a complex and multifaceted process that requires considerable effort. At the same time, cybersecurity provided by the state should not slow down the process of forming a national information space that would correspond to the information and intellectual potential of the state and would not hinder Ukraine's entry into the world information space as a subject of equal international relations. In view of this, the strategic task of state policy on cybersecurity should be to form a system based on scientifically sound criteria and experience in information security management.

Modern information systems are complex and therefore dangerous in themselves, even without the active actions of attackers. New vulnerabilities in software and hardware implementations are constantly being identified. We have to take into account the extremely wide range of hardware and software, the many connections between components and more.

Evidence of the complexity of the cybersecurity problem is the parallel (and fairly rapid) increase in the number of information security breaches and losses on cybersecurity threats combined with an increase in the average loss from each of the breaches. The latter circumstance is another argument in favor of the importance of cybersecurity.

Therefore, it is necessary to create requirements for a cybersecurity system that could provide more opportunities in the choice of methods in the management of the protection of automated information systems.

It should be noted that in Ukraine a single information system is being created, which consists of permanent elements:

- central subsystem;
- functional subsystems;
- transport data network;
- data processing centers, telecommunication networks of system subjects;
- complex information protection systems of subsystems of a single system.

Therefore, this system is distributed. It should be noted that this complicates the task very much. And since in addition to a single system there are other (regional, local, departmental and other) systems, this issue becomes very important and difficult in the system of government, in the banking system and other spheres of society [1, 2].

In addition, in modern conditions the large number of cyberattacks on systems of all classes should be taken into account.

The task of determining the optimal quality indicators of information resource management systems of automated systems is one of the most important problems in designing integrated information security systems. This is due to the complexity of such systems, the presence of many variable parameters, and the complexity of calculating quality indicators. In addition, the determined quality indicators should not only ensure the optimality of the target function, but also the stability of the protection system in a wide range of external adverse effects.

In [3,4] the models of both systems themselves and their control systems and information protection systems are given. But it should be noted that these systems provide very serious mistakes and errors in countering modern cyberattacks.

In the paper [5] it is proposed to use the provisions of the theory of automatic control systems as a model of the protection management system, in which the indicators are given in the form of integrated quadratic estimates (IQE) of control error [6; 7]. The problem is that the existing methods of calculating IQE [8-12] do not take into account errors in determining quality indicators, as well as the vector nature of these indicators.

The most suitable model that was used to solve this problem is the model given in [13]. This model allows to solve all problems which stand at management of cybersecurity of information systems in modern difficult conditions.

The aim of the work is to solve problems (development of algorithms), which are a problem of optimization of stable protection management systems using vector objective functions.

Main part

IQE management error has the form

$$I = \int_0^\infty e^2(t)dt. \tag{1}$$

Control error $e(t)$ can be defined as the weighted sum of the derivatives of the deviation $z(t) = y(\infty) - y(t)$ of the controlled value $y(t)$ from the set value $y(\infty) = 1$ using weights $\omega_k, k = \overline{0,l}; \omega_0 = 1$:

$$e(t) = \sum_{k=0}^{l} \omega_k \cdot z^{(l-k)}(t). \tag{2}$$

Let the control function $y(t)$ correspond to the transfer function

$$W(s) = \frac{\beta(s)}{\alpha(s)}; \alpha(s) = \sum_{i=0}^{n} \alpha_i s^{n-i}; \beta(s) = \sum_{j=0}^{m} \beta_j s^{m-j}; \alpha_n = \beta_m. \tag{3}$$

The Laplace transform $z_k(s)$ for derivatives $z^{(k)}(t)$ in expression (2) is obtained by applying the theorem on the differentiation of the original with the initial conditions $z(0) = 1; z^{(k)}(0) = 0, k = \overline{1,l}$:

$$z_0(s) = \frac{1 - W(s)}{s}; z_k(s) = -s^{k-l} \cdot W(s), k = \overline{1,l}; l \le n - m.$$

Converting Laplace equation (2) and introducing a polynomial with weights

$$\omega(s) = \sum_{k=0}^{l} \omega_k \cdot s^{l-k}, \tag{4}$$

get the error image:

$$E(s) = \frac{1 - W(s) \cdot \omega(s)}{s}. \tag{5}$$

Considering expressions (3) for the transfer function and determining through

$$\delta(s) = \frac{\alpha(s) - \beta(s) \cdot \omega(s)}{s}, \tag{6}$$

the error image is represented as the ratio of two polynomials:

$$E(s) = \frac{\delta(s)}{\alpha(s)}. \tag{7}$$

In the general case, the error can be defined as the difference between the reference function $y_e(t)$ and controlled variable $y(t)$: $e(t) = y_e(t) - y(t).$

The transfer function of the reference function is set as $W_e(s) = 1/\omega(s)$, then the error image will look like

$$E(s) = \frac{W_e(s) - W(s)}{s}. \tag{8}$$

After transformations we come to equality $E(s) = \delta(s)/\alpha(s)/\omega(s).$ To determine the reference function, choose a standard transition function $y_0(t)$ of the desired form with known values of the setting time τ_0 and coefficients of the transfer function $W_0(s) = 1/\gamma(s)$ [4; 5]:

$$\gamma(s) = \sum_{k=o}^{i} \gamma_k \cdot s^{i-k}, \gamma_0 = 1.$$

Setting the value of the time of setting the reference function τ_e, calculate the weights $\omega_k = \mu^{l-k} \cdot \gamma_k, k = \overline{0,l}$, where $\mu = \tau_e / \tau_0 \cdot$ According to these formulas and expressions (6), (7) we will make an algorithm for forming an error image.

Algorithm 1. Input parameters: τ_e – the time of setting the reference function, τ_0 – the standard process setting time, γ – an array of coefficients of the standard transition function, α and β – arrays of coefficients of the denominator and numerator of the transfer function. Output parameters: δ – an array of error image numerator coefficients.

Step 1. Put $l := \dim\gamma - 1; n := \dim\alpha - 1; \mu := \tau_e / \tau_0; \eta := 1; k := l.$

Step 2. Calculate $\omega_k := \gamma_k \cdot \eta; \eta := \eta \cdot \mu.$

Step 3. If $k > 0,$ put $k := k - 1$ and go to step 2.

Step 4. Calculate the convolution of two vectors $c := \beta \otimes \omega.$

Step 5. Calculate $\delta := \alpha - c.$

Step 6. The end.

To calculate the IQE (1) based on error conversion $E(s)$ by Parseval's theorem we write

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} E(s) \cdot E(-s) ds. \tag{9}$$

Let us represent the error image by equality

$$E(s) = \frac{\delta(s)}{\alpha(s)}, \tag{10}$$

where $\alpha(s)$ – polynomial of degree n, all the roots of which lie in the left half-plane, $\beta(s)$ – polynomial of degree $n-1$. Then

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{\delta(s) \cdot \delta(-s)}{\alpha(s) \cdot \alpha(-s)} ds. \tag{11}$$

Denote by $\delta(s^2) = \beta(s) \cdot \beta(-s)$ polynomial of even degrees and rewrite (11):

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{\delta(s^2)}{\alpha(s) \cdot \alpha(-s)} ds. \tag{12}$$

Let in the image (10)

$$\alpha(s) = \sum_{i=0}^{n} \alpha_i s^i; \beta(s) = \sum_{i=0}^{n-1} \beta_i s^i. \tag{13}$$

Considering $m = n - 1$, we present the polynomial of even degrees (12) in the form

$$\delta(s^2) = \sum_{k=0}^{m} \delta_k s^{2k}.$$

Integral (12) is calculated by the modified formula of A.M. Katz [5]:

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{\delta(s^2)}{\alpha(s) \cdot \alpha(-s)} ds = \frac{(-1)^{n-1}}{2\alpha_n} \cdot \frac{G}{H}, \tag{14}$$

where H – Hurwitz determinant for a polynomial $\alpha(s)$, and G – determinant obtained by H replacing the last line with a line of polynomial coefficients $\delta(s^2)$:

$$H = \begin{vmatrix} \alpha_0 & \alpha_2 & \alpha_4 & \alpha_6 & \dots & 0 \\ 0 & \alpha_1 & \alpha_3 & \alpha_5 & \dots & 0 \\ 0 & \alpha_0 & \alpha_2 & \alpha_4 & \dots & 0 \\ 0 & 0 & \alpha_1 & \alpha_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \alpha_{n-1} \end{vmatrix}; G = \begin{vmatrix} \alpha_0 & \alpha_2 & \alpha_4 & \alpha_6 & \dots & 0 \\ 0 & \alpha_1 & \alpha_3 & \alpha_5 & \dots & 0 \\ 0 & \alpha_0 & \alpha_2 & \alpha_4 & \dots & 0 \\ 0 & 0 & \alpha_1 & \alpha_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \delta_0 & \delta_1 & \delta_2 & \delta_3 & \dots & \delta_{n-1} \end{vmatrix} \tag{15}$$

Formula (14) differs from Katz's formula by writing polynomials (13) in ascending order of the degree of the variable s, which simplifies the calculation algorithm IQE. Considering

$\alpha_i^{(0)} = \alpha_i, i = \overline{0:2:n}; \alpha_i^{(1)} = \alpha_i, i = \overline{1:2:n}$ (here is the record of the species $i = \overline{k:2:n}$ means a sequence of integers beginning with k, increases with step 2 and does not exceed n), convert the determinants (15) to a triangular shape:

$$\lambda^{(1)} = \frac{\delta_0}{\alpha_0}; \delta_j^{(1)} = \delta_j - \lambda^{(1)} \cdot \alpha_i, (i, j) = (2,1),(4,2),(6,3),\ldots \tag{16}$$

$$\gamma^{(k)} = \frac{\alpha_{k-2}^{(k-2)}}{\alpha_{k-1}^{(k-1)}}; \lambda^k = \frac{\delta_{k-1}^{(k-1)}}{\alpha_{k-1}^{(k-1)}}, k = \overline{2, n-1}; \tag{17}$$

$$\alpha_i^{(k)} = \alpha_i^{(k-2)} - \gamma^{(k)} \cdot \alpha_{i+1}^{(k-1)}; \delta_j^{(k)} = \delta_j^{(k-1)} - \lambda^{(k)} \cdot \alpha_{i+1}^{(k-1)},$$

$$(i, j) = (k, k),(k+2, k+1),(k+4, k+2),\ldots \tag{18}$$

In this case IQE (15) is calculated by the formula

$$I = \frac{(-1)^{n-1} \cdot \delta_{n-1}^{(n-1)}}{2 \cdot \alpha_n \cdot \alpha_{n-1}^{(n-1)}}. \tag{19}$$

The algorithm for calculating the IQE according to Katz's formula looks like this.

Algorithm 2. Input parameters: α and β – arrays of denominator and numerator coefficients of error conversion. Output parameter: I – IQE.

Step 1. Put $n := \dim \alpha - 1; m := n - 1; \delta := PolPow2(\beta)$ ($PolPow2$ – a function that calculates an array of coefficients of a polynomial of even degrees $\delta(s^2)$).

Step 2. Calculate $\lambda := \delta_0 / \alpha_0$ and put $j := 1; i := 2.$

Step 3. Calculate $\delta_j := \delta_j - \lambda \cdot \alpha_i$ and put $j := j + 1.$

Step 4. If $i < m$, then put $i := i + 2$ and go to step 3.

Step 5. Put $s := 1; k := 2.$

Step 6. Calculate $\gamma := \alpha_{k-2} / \alpha_{k-1}; \lambda := \delta_{k-1} / \alpha_{k-1}$, and put $s := -s; j := k; i := k.$

Step 7. Calculate $\alpha_i := \alpha_i - \gamma \cdot \alpha_{i+1}; \delta_j := \delta_j - \lambda \cdot \alpha_{i+1}$ and put $j := j + 1.$

Step 8. If $i < m$, put $i := i + 2$ and go to step 7.

Step 9. If $k < m$, put $k := k + 1$ and go to step 6.

Step 10. Calculate $I := \delta_{n-1} / (2 \cdot \alpha_{n-1} \cdot \alpha_n).$

Step 11. If $s < 0$, then set $I := I.$

Step 12. The end.

Integral (11) can be calculated by the method of Ostrom using an iterative procedure [11]. Let the polynomials in the image of the error (10) have the form

$$\alpha(s) = \sum_{i=0}^{n} \alpha_i s^{n-1}; \beta(s) = \sum_{i=0}^{n-1} \beta_i s^{n-i}. \tag{20}$$

Considering

$\alpha_i^{(0)} = \alpha_i, i = \overline{0:2:n}; \alpha_i^{(1)} = \alpha_i, i = \overline{1:2:n}; \beta_i^0 = \beta_i, i = \overline{0:2:n-1};$

$\beta_i^{(1)} = \beta_i, i = \overline{1:2:n-1},$

calculate $\gamma^{(k)}$ and $\lambda^{(k)}$ for $k = \overline{1, n}$:

$$\gamma^{(k-1)} = \frac{\alpha_{k-2}^{(k-2)}}{\alpha_{k-1}^{(k-1)}}; \lambda^{(k-1)} = \frac{\beta_{k-2}^{(k-2)}}{\alpha_{k-1}^{(k-1)}}, k = \overline{2, n-1}; \tag{21}$$

$$\alpha_i^{(k)} = \alpha_i^{(k-2)} - \gamma^{(k-1)} \cdot \alpha_{i+1}^{(k-1)};\ \beta_i^{(k)} = \beta_i^{(k-2)} - \lambda^{(k-1)} \cdot \alpha_{i+1}^{(k-1)},\ i = \overline{k:2:n-1}; \tag{22}$$

$$\gamma^{(k)} = \frac{\alpha_{k-1}^{(k-1)}}{\alpha_k^{(k)}};\ \lambda^{(k)} = \frac{\beta_{k-1}^{(k-1)}}{\alpha_k^{(k)}},\ k = \overline{n-1,n}. \tag{23}$$

The value of IQE (11) is calculated by the formula

$$I = 0,5 \cdot \sum_{k=1}^{n} \frac{(\lambda^{(k)})^2}{\gamma^{(k)}}. \tag{24}$$

Algorithm 3. Input and output parameters are similar to algorithm 2.

Step 1. Put $n := \dim\alpha - 1;\ m := n-1;\ I := 0;\ k := 2.$

Step 2. Calculate $\gamma := \alpha_{k-2}/\alpha_{k-1};\ \lambda := \beta_{k-2}/\alpha_{k-1};\ I := I + \lambda^2/\gamma$ and put $i := k.$

Step 3. Calculate $\alpha_i := \alpha_i - \gamma \cdot \alpha_{i+1};\ \beta_i := \beta_i - \lambda \cdot \alpha_{i+1}.$

Step 4. If $i < m$, put $i := i + 2$ and go to step 3.

Step 5. If $k < m$, then put $k := k + 1$ and go to step 2.

Step 6. Put $k := m.$

Step 7. Calculate $\gamma := \alpha_{k-1}/\alpha_k;\ \lambda := \beta_{k-1}/\alpha_k;\ I := I + \lambda^2/\gamma.$

Step 8. If $k < n$, put $k := k + 1$ and go to step 7.

Step 9. Calculate $I = 0,5 \cdot I.$

Step 10. The end.

The input parameters of algorithms 2 and 3 are calculated by algorithm 1.

A necessary condition for the stability of the protection management system is to require the positivity of all coefficients $\alpha_i > 0, i = \overline{0,n}$ [2; 3].

Considering $\alpha_i^{(0)} = \alpha_i, i = \overline{0:2:n};\ \alpha_i^{(1)} = \alpha_i, i = \overline{1:2:n}$, the Rauss-Hurwitz stability criterion is reduced to calculations by formulas

$$\gamma^{(k)} = \frac{\alpha_{k-2}^{(k-2)}}{\alpha_{k-1}^{(k-1)}};\ \alpha_i^k = \alpha_i^{k-2} - \gamma^{(k)} \cdot \alpha_{i+1}^{(k-1)},\ i = \overline{k:2:n-1};\ k = \overline{2,n-1}. \tag{25}$$

We introduce the notation of the elements of the Raus-Hurwitz determinant:

$$\rho_0 = \alpha_0,\ \rho_1 = \alpha_1,\ \rho_k = \alpha_k^k,\ k = \overline{2,n-1},\ \rho_n = \alpha_n. \tag{26}$$

Algorithm 4. Input parameter: α – an array of coefficients of the characteristic polynomial. Output parameter: ρ – an array of Raus-Hurwitz coefficients.

Step 1. Put $n := \dim\alpha - 1;\ m := n-1;\ \rho := \alpha;\ k := 2.$

Step 2. Calculate $\gamma := \rho_{k-2}/\rho_{k-1}$ and put $i := k.$

Step 3. Calculate $\rho_i := \rho_i - \gamma \cdot \rho_{i+1}.$

Step 4. If $i < m$, put $i := i + 2$ and go to step 3.

Step 5. If $k < m$, put $k := k + 1$ and go to step 2.

Step 6. The end.

Necessary and sufficient conditions for the stability of the protection management system are the conditions $\alpha_i > 0, i = \overline{0,n}$, and $\rho_k > 0, k = \overline{2,n-1}$. Taking into account these conditions and algorithm 4, we will develop an algorithm for determining the stability of the control system according to the Rauss-Hurwitz criterion.

Algorithm 5. Input parameter: α – an array of coefficients of the characteristic polynomial. Output parameter: B – a sign of system stability.

Step 1. Put $n := \dim\alpha - 1;\ B := 1;\ i := 0.$

Step 2. If $\alpha_i \le 0$, put $B := 0$ and go to step 10.

Step 3. If $i < n$, put $i := i + 1$ and go to step 2.

Step 4. Put $m := n - 1;\ \rho := \alpha;\ k := 2.$

Step 5. Calculate $\gamma := \rho_{k-2} / \rho_{k-1}$ and put $i := k.$

Step 6. Calculate $\rho_i := \rho_i - \gamma \cdot \rho_{i+1}.$

Step 7. If $i < m$, put $i := i + 2$ and go to step 6.

Step 8. If $\rho_k \le 0$, put $B := 0$ and go to step 10.

Step 9. If $k < m$, put $k := k + 1$ and go to step 5.

Step 10. The end.

Let the coefficients of the characteristic polynomial-function of the vector of the varied parameters of the system $\alpha_i = \alpha_i(x), i = \overline{0,n};\ x \in R^p$. In order for the system to be stable, it is necessary and sufficient to:

$$\alpha_i(x) > 0, i = \overline{0,n};\ \rho_k(x) > 0, k = \overline{2,n-1}. \tag{27}$$

The degree of violation of the first group of inequalities is represented by one penalty function

$$P(x) = \sum_{i=0}^{n} \max\left\{-\alpha_i(x), 0\right\}. \tag{28}$$

Inequalities (27) correspond to the areas of constraints:

$$\Omega_1 = \left\{x \middle| \alpha_i(x) > 0, i = \overline{0,n}\right\};\ \Omega_k = \left\{x \middle| \rho_k(x) > 0\right\},\ k = \overline{2,n-1}. \tag{29}$$

Let's make a system of areas from them: $D_1 = \Omega_1,\ D_k = D_{k-1} \bigcap \Omega_k,\ k = \overline{2,n-1}$, from which by means of differences of sets we will adjust areas of levels of restrictions:

$$H_0 = R^p \setminus D_1;\ H_k = D_k \setminus D_{k+1}, k = \overline{1,n-1};\ H_{n-1} = D_{n-1}. \tag{30}$$

A two-dimensional vector penalty function is proposed for the transition to the stability region:

$$F(x) = \begin{cases} (0; P(x)), & x \in H_0; \\ (k; -\rho_{k+1}(x)), & x \in H_k, k = \overline{1,n-2}; \\ (n-1; 0), & x \in H_{n-1}. \end{cases} \tag{31}$$

The first component reflects the affiliation of the argument to a particular area and is called *a level function*. The second is the penalty for violating the restriction and is called *a penalty function*. When all the constraints are satisfied (the third component), the level function acquires the maximum value $n-1$ and the penalty function is reset.

Algorithm 6. Input parameter: α – an array of coefficients of the characteristic polynomial. Output parameter: B – a sign of system stability; F – the value of the vector penalty function.

Step 1. Put $n := \dim\alpha - 1;\ B := 1;\ i := 0;\ h := 0;\ P := 0.$

Step 2. If $\alpha_i \le 0$, calculate $P := P - \alpha_i$ and put $B := 0.$

Step 3. If $i < n$, put $i := i + 1$ and go to step 2.

Step 4. If $B = 0$, put $F := (h, P)$ and go to step 12.

Step 5. Put $m := n - 1;\ \rho := \alpha;\ k := 2.$

Step 6. Calculate $\gamma := \rho_{k-2} / \rho_{k-1}$ and put $h := h + 1,\ i := k.$

Step 7. Calculate $\rho_i := \rho_i - \gamma \cdot \rho_{i+1}.$

Step 8. If $i < m$, put $i := i + 2$ and go to step 7.

Step 9. If $\rho_k \le 0$, put $F := (h, -\rho_k);\ B := 0$ and go to step 12.

Step 10. If $k < m$, put $k := k + 1$ and go to step 6.

Step 11. Calculate $F := (h : 1, 0)$.

Step 12. The end.

In the case of estimating the vector objective function, we assume that the transfer function (3) depends on the vector of variable parameters x:

$$W(x,s) = \frac{\beta(x,s)}{\alpha(x,s)}, \alpha(x,s) = \sum_{i=0}^{n} \alpha_i(x) s^{n-1}; \tag{32}$$

$$\beta(x,s) = \sum_{j=0}^{m} \beta_j(x) s^{m-j}; \alpha_n(x) = \beta_m(x).$$

Defining the weight polynomial (4) and determining the polynomial (6)

$$\delta(x,s) = \frac{\alpha(x,s) - \beta(x,s) \cdot \omega(s)}{s}, \tag{33}$$

IQE is presented as a function of variable parameters:

$$I(x) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{\delta(x,s) \cdot \delta(x,-s)}{\alpha(x,s) \cdot \alpha(x,-s)} ds. \tag{34}$$

At each value of the vector x from the stability of the system values of this function can be calculated by algorithms 2 and 3. Redefine the vector penalty function (31) into a vector objective function, supplementing it with the value of IQE (34) in the field of stability:

$$F(x) = \begin{cases} (0; P(x)), & x \in H_0; \\ (k; -\rho_{k+1}(x)), & x \in H_k, k = \overline{1, n-2}; \\ (n-1; I(x)), & x \in H_{n-1}. \end{cases} \tag{35}$$

Modify algorithms 2 and 3 according to algorithm 6 to calculate the vector objective function (35) [10].

Algorithm 7. Input parameter: α and β – arrays of denominator and numerator coefficients of error conversion. Output parameter: F – the value of the vector objective function; B – a sign of system stability.

Step 1. Put $n := \dim \alpha - 1; B := 1; i := 0; h := 0; P := 0.$

Step 2. If $\alpha_i \le 0$, calculate $P := P - \alpha_i$ and put $B := 0.$

Step 3. If $i < n$, put $i := i + 1$ and go to step 2.

Step 4. If $B = 0$, put $F := (h, P)$ and go to step 17.

Step 5. Calculate $\delta := PolPow2(\beta); \lambda := \delta_0 / \alpha_0$ and put $m := n - 1; j := 1; i := 2.$

Step 6. Calculate $\delta_j := \delta_j - \lambda \cdot \alpha_i$ and put $j := j + 1.$

Step 7. If $i < m$, put $i := i + 2$ and go to step 6.

Step 8. Put $s := 1; k := 2.$

Step 9. Calculate $\gamma := \alpha_{k-2} / \alpha_{k-1}; \lambda := \delta_{k-1} / \alpha_{k-1}$ and put $h := h + 1; s := -k; j := k; i := k.$

Step 10. Calculate $\alpha_i := \alpha_i - \gamma \cdot \alpha_{i+1}; \delta_j := \delta_j - \lambda \cdot \alpha_{i+1}$ and put $j := j + 1.$

Step 11. If $i < m$, put $i := i + 2$ and go to step 10.

Step 12. If $\alpha_k \le 0$, put $F := (h, -\alpha_k); B := 0$ and go to step 17.

Step 13. If $k < m$, put $k := k + 1$ and go to step 9.

Step 14. Calculate $I := \delta_{n-1} / (2 \cdot \alpha_{n-1} \cdot \alpha_n).$

Step 15. If $s < 0$, put $I := -I.$

Step 16. Calculate $F := (h + 1, I).$

Step 17. The end.

Algorithm for calculating the vector objective function according to Ostrom.

Algorithm 8. Input and output parameters are similar to the algorithm 7.

Step 1. Put $n := \dim\alpha - 1;\ B := 1;\ i := 0;\ h := 0;\ P := 0.$

Step 2. If $\alpha_i \le 0$, calculate $P := P - \alpha_i$ and put $B := 0.$

Step 3. If $i < n$, put $i := i + 1$ and go to step 2.

Step 4. If $B = 0$, put $F := (h, P)$ and go to step 15.

Step 5. Put $m := n - 1;\ I = 0;\ k := 2.$

Step 6. Calculate $\gamma := \alpha_{k-2}/\alpha_{k-1};\ \lambda := \beta_{k-2}/\alpha_{k-1};\ I := I + \lambda^2/\gamma$ and put $h := h + 1;\ i := k.$

Step 7. Calculate $\alpha_i := \alpha_i - \gamma \cdot \alpha_{i+1};\ \beta_i := \beta_i - \lambda \cdot \alpha_{i+1}.$

Step 8. If $i < m$, put $i := i + 2$ and go to step 7.

Step 9. If $\alpha_k \le 0$, put $F := (h, -\alpha_k);\ B := 0$ and go to step 15.

Step 10. If $k < m$, put $k := k + 1$ and go to step 6.

Step 11. Put $k := m.$

Step 12. Calculate $\gamma := \alpha_{k-1}/\alpha_k;\ \lambda := \beta_{k-1}/\alpha_k;\ I := I + \lambda^2/\gamma.$

Step 13. If $k < n$, put $k := k + 1$ and go to step 12.

Step 14. Calculate $F := (h + 1, 0,5 \cdot I).$

Step 15. The end.

Studies have shown the high efficiency of the proposed algorithms. And on the night of February 23-24, 2022, they effectively worked against DDoS attacks carried out by Russian hackers on Ukrainian information systems.

Conclusions

Based on the model of information management system protection of information resources in the form of automatic control system:

1. The method of forming integrated quadratic estimates (IQE) of control error is proposed. This method takes into account the weights of the estimates at the desired installation time and standard transfer functions.

2. Algorithms for calculating IQE according to the modified Katz formula and Ostrom's method for arbitrary order control systems are developed, including vector representation of the objective function of the protection system.

3. The vector penalty function is proposed and the algorithm of its calculation is developed for display of degree of infringement of conditions of stability of parameters of system of protection by Rauss-Hurwitz criterion.

4. Studies have shown the high efficiency of the proposed algorithms. And on the night of February 23-24, 2022, they effectively worked against DDoS attacks carried out by Russian hackers on Ukrainian information systems.

References

1. Brailovskyi M.M., Zybin S.V., Kobozeva A.A., Khoroshko V.O., Khokhlachova Yu.Ie. Analiz kiberzakhyshchenosti informatsiinykh system: monohrafiia - K: FOP Yanchevskyi O.V., 2021.-360s.

2. Luntovskyi A.O., Klymash M.M. Informatsiina bezpeka rozpodilenykh system: monohrafiia - Lviv: NU "Lvivska politekhnika", 2014. -480 s.

3. Hryshchuk R.V., Danyk Yu.H. Osnovy kibernetychnoi bezpeky: monohrafiia - Zhytomyr: ZhNAEU, 2015. -636s.

4. Hubanov D.A., Novykov D.A., Chkhartyshvyly A.H. Sotsyalnыe sety: modely ynformatsyonnoho vlyianyia, upravlenyia y protyvoborstva - M: Yzd. fyzmatlyter., 2010. -228s.

5. Tyskyna E. O., Khoroshko V. A. Modelyrovanye systemы upravlenyia zashchytoi obъekta. Informatsiina bezpeka liudyny, suspilstva, derzhavy. 2010. № 2(4). S. 66-70.

6. Lapko A.P., Krokhov S.V., Chentsov S.Y., Feldman D.A. Obuchaiushchye systemы obrabotky ynformatsyy y pryniatyia reshenyi. – Novosybyrsk: Nauka, 1996. -284 s.

7. Sovremenna teoryia system upravlenyia. Yzd. 2-e / Pod red. K.T. Leondesa – M: Myr, 2000.-521 s.

8. Fuller A. The replacement of saturation constrains by energy constrains in control optimization theory. International Journal of Control. 1997. № 3. P. 201-277.

9. Holovkyn B.A. Raschët kharakterystyk y planyrovanyia protsessov. Yzd. 2-e. – M: Radyo y sviaz, 2003.-274 s.

10. Kudynov V.A., Parkhuts L.T., Khoroshko V. A. Metodyka systemnoho proektyrovanyia korporatyvnыkh setei. Visnyk DUIKT, T3, №3-4, 2005. – S. 184-186.

11. Ostrem K. Yu. Vvedenye v stokhastycheskuiu teoryiu upravlenyia / per. s anhl. pod red. N. S. Raibmana. M., 1973. 322 s.

12. Khymmelblau D. Prykladnoe nelyneinoe prohrammyrovanye / per. s anhl. pod red. M. S. Bыkhovskoho. M., 1975. 536 s.

13. Behma T.V., Kapustian M.V., Khoroshko V.O. Matematychni modeli funktsionuvannia skladnykh system // Visnyk Skhidnoukrainskoho natsionalnoho universytetu im. V.Dalia, №7(161), 2011, Ch.1. -s.252-263.

14. Doktryna informatsiinoi bezpeky Ukrainy (zatverdzhena ukazom Prezydenta Ukrainy №47/2017 vid 25 liutoho 2017 roku): [Elektronnyi resurs] / Ofitsiine predstavnytstvo Prezydenta Ukrainy. – rezhym dostupu: http://www.president.gov.ua/documents/472017-21374 (data zvernennia 25.07.2018). - Nazva z ekranu.

15. Rada natsionalnoi bezpeky i oborony Ukrainy. [Elektronnyi resurs] – Rezhym dostupu: https://www.rnbo.gov.ua/.

16. Kiberataky na ukrainski derzhavni saity (2022). [Elektronnyi resurs] – Rezhym dostupu: https://uk.wikipedia.org/wiki/ (2022).

https://doi.org/10.31891/csit-2022-2-7
UDC 004.942

Lubomyr SIKORA, Nataliia LYSA, Olga FEDEVYCH
Lviv Polytechnic National University, Lviv, Ukraine
Rostyslav TKACHUK
Lviv State University of Life Safety, Lviv, Ukraine

LASER AND INFORMATION TECHNOLOGIES FOR CONTROLLING DYNAMIC DISPLACEMENTS SPATIAL STRUCTURES OF OBJECTS UNDER THE INFLUENCE OF ACTIVE MAN-MADE AND NATURAL RISK FACTORS FOR ACCIDENTS

At the present stage of science development, for technological and technogenic energy-intensive systems, systematic methods of identification of structure, dynamics, and risk assessment are developed, while for spatial objects this problem is not fully solved. This applies to the construction and operation of such objects with a spatially distributed structure such as bridges, large pavilions, high-rise buildings, aggregate lines on a common foundation for color printing, which are subject to a large dynamic, non-uniform load-capacity, operating over a long period of operation. Their destruction with the combined action of dynamic and static heterogeneous flow factors in time of high power, leads to the accidents and human losses. The main factor that leads to cognitive errors in the design of spatial structures is that experts in the design process do not fully take into account the concepts of physical force, power and physical energy factors with stream random structure. In this aspect, the problem of dynamic structural stability under the influence of factors with a stochastic structure drew attention to Y. P. Dragan, introducing the notion of "stochastic process of finite energy" and "finite power of flows (sequences) of active physical force actions". Under certain conditions, the complex action of force factors leads to the emergence of solitons, that is, the formation of the peak of energy and power at a certain time in the weakest node of the structure that destroys it. If the designer, by virtue of his cognitive abilities and level of knowledge, does not take into account the energetic nature of the factors as destructive forces, then this leads to the destruction of infrastructure objects (cities in Genoa, Italy 2015, built in 1967) devastating floods, fires, transport disasters, tsunami. As for the steel construction bridges in the USA (New York), built on the basis of the methods of vibration calculations by S. Tymoshenko, they are operated for more than 100 years, with appropriate technical service.

The assessment of the vibrational stability of spatial structures, both existing and new projects, remains a complex control problem that is not resolved to the fullest, and therefore the development of integrated intellectual methods for designing and controlling their state is relevant

The intensive development of infrastructure, both social and technogenic, results from the impact of transport flows, power plants, harmful emissions, to the growth of force environmental load on spatial structures, corrosion of metal components, and the growth of vibrational effects on elements of objects. Further development of such negative processes leads to a decrease in the strength of structures, their stability, operational reliability and destruction. Reducing the quality of bearing structures, due to neglected negative influences, makes it impossible to forecast the moment of emergency situations. Accordingly, the development of methods for remote control of vibrations of spatial elements of bearing structures is a main problem for various industries.

Key words: construction, vibration, laser, signal, dynamic processes, active factors, data, system, information, project, risks, accident.

ЛЮБОМИР СІКОРА, НАТАЛЯ ЛИСА, ОЛЬГА ФЕДЕВИЧ
Національний університет «Львівська політехніка»
РОСТИСЛАВ ТКАЧУК
Львівський державний університет безпеки життєдіяльності

ЛАЗЕРНІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ КОНТРОЛЮ ДИНАМІЧНИХ ЗМІЩЕНЬ ПРОСТОРОВИХ СТРУКТУР ОБ'ЄКТІВ ПІД ВПЛИВОМ АКТИВНИХ ТЕХНОГЕННИХ ТА ПРИРОДНИХ ФАКТОРІВ

На сучасному етапі розвитку науки, для технологічних і техногенних енергоактивних систем, вироблені системні методи ідентифікації структури, динаміки, оцінки ризику, тоді як для просторових об'єктів ця проблема в повній мірі не розв'язана. Це стосується будівництва і експлуатації таких об'єктів з просторово розподіленою структурою як мости, великі павільйони, висотні будинки, агрегатні лінії на спільному фундаменті для кольорового друку які піддаються великим динамічним неоднорідним по потужності навантаженням, що діють протягом тривалого часу експлуатації. Їх руйнація при сукупній дії динамічних і статичних неоднорідних потокових у часі факторів великої енергетичної потужності, приводить до аварій і людських втрат. Основний фактор, який приводить до когнітивних помилок при проектуванні просторових конструкцій, є те що фахівці у процесі розробки проекту не до кінця враховують поняття фізичної сили, енергії потужності та фізичної енергії факторів з потоковою випадковою структурою. На цей аспект проблеми динамічної стійкості конструкції при дії факторів з стохастичною структурою звернув увагу Я. П. Драган, ввівши поняття «стохастичного процесу скінченої енергії» і «скінченої потужності потоків (послідовностей) активних фізичних силових дій». При певних умовах комплексна дія силових факторів приводить до виникнення солітонів тобто формування піку енергії та потужності у певний момент часу у найслабшому вузлі конструкції, що її руйнує.

Якщо проектант, в силу своїх когнітивних здібностей і рівня знань, не враховує енергетичну сутність факторів, як руйнівних сил, тоді це приводить до руйнування інфраструктурних об'єктів (міст в Генуї, Італія 2018р., збудований у 1967 році, Китай 2019р.) руйнівні повені, пожари, транспортні катастрофи, цунамі. Щодо мостів з металоконструкцій в США (Нью-Йорк), побудованих з врахування методів вібраційних розрахунків С. Тимошенко, то вони експлуатуються більше ніж 100 років, при відповідному технічному обслуговуванні.

Оцінка вібраційної стійкості просторових конструкцій, як існуючих так і нових проектів залишається складною проблемою створення систем контролю і діагностики, нерозв'язаною у повній мірі, і тому розроблення інтегрованих

інтелектуальних методів проектування систем контролю методом дистанційного лазерного зондування є актуальною.

Ключові слова: конструкція, вібрація, лазер, сигнал, динамічні процеси, активні фактори, дані, система, інформація, проект, ризики, аварія.

Introduction

Intensive development of infrastructure, both social and man-made, leads, due to the action of transport flows, power plants, industries with harmful emissions, to an increase in environmental environmental load on spatial structures, corrosion of metal components, increased vibration effects on elements. Further development of such negative processes leads to a decrease in the strength of structures, their stability, operational reliability and destruction. The decline in the quality of load-bearing structures, due to unaccounted for negative impacts, makes it impossible to predict the moment of occurrence of emergencies. Accordingly, the development of methods for remote control of vibrations of the spatial elements of load-bearing structures is an urgent problem for various industries.

Analysis of literature sources

The problem of monitoring and assessing the stability of spatial structures is relevant in recent centuries (1700-2019). It includes problems of building destruction, vehicles, communications, large bridges, high-rise buildings [1] under the influence of various factors [2-3], vibration [4-7], soil landslides and earthquakes [8], aging of components and materials[9] dynamic transport loads and flows [10]. One of the least studied are cognitive factors and deficiencies as well as errors occured in the design process [11-14] and subsequent operation with the participation of operational personnel and designers [15-16].

Therefore, the methods and tools development – both cognitive control of projects and vibration control systems of complex spatial structures is still relevant [17-18], as it requires an integrated approach using signal theory [19-20], the theory of data processing [21-22], interpretation of data and situations, decision making [4, 23-25].

Accordingly, an important task is the construction of objects models and simulation modeling on analogues [24, 26], which ensures the detection of new physical effects [27]. Without taking into account the peculiarities of cognitive thinking of designers, it is impossible to reliably design buildings and their implementation with the appropriate service life and resistance to destruction [28-32].

Introduction. Analysis of dynamic control problems

At the present stage of information and measuring systems development for vibration control of complex structures under the action of a set of energy-active factors by non-contact remote method is not fully considered.

Therefore, the development of laser methods for remote control of dynamic modes of large spatial structures, under the influence of active dynamic factors in time and space is an important problem.

The purpose of the study. To control the vibration of spatial structures and aggregate production systems to create and justify the use of laser probing and develop a block diagram of a laser vibrometer.

Basic tasks to be solved

To solve the problem of remote control of vibration of large spatial structures and aggregate printing systems it is necessary:

– to substantiate structural models of objects and behavior models in time under the influence of active factors;

– to justify the choice of laser remote sensing method for controlling the vibration of structures in critical places of the object of study and the spatial structure of the object structures and (foundation) platform of aggregate production lines of high quality products;

– to develop information technology of laser signals processing and their estimation for determination of vibration parameters.

Research methods

To solve a complex scientific and applied problem of creating systems for remote control of vibration load of spatial structures and boundary modes on the basis of laser sounding, methods and theories were used:

Models of dynamic factors influencing spatial structure.

Since dynamic factors [6, 9, 33] have an energy-active structure, then ignoring their essence leads to the collapse of the mechanical spatial structure due to oscillations and soliton effects [5].

Accordingly, the study of their dynamics requires the use and creation of new methods and control systems based on laser remote sensing, which provided the detection of oscillations of the spatial structure of structures [34]. To solve the above problems, needed:

1. Model of n – dimensional spatial fluctuations for a long section of the bridge (100 m) and a common foundation for aggregate printing production.

2. Energy soliton model for oncoming traffic flows as perturbing factors. $\left(\xrightarrow{\vec{n}^2} \uparrow \vec{n}_n\right)$ vertically and along relative to the supports (Fig. 1).

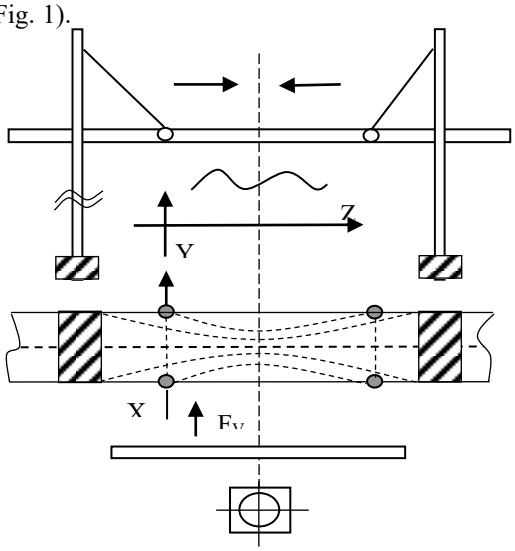

Fig. 1. Method of laser sounding of spatial structure

3. The model of wind load with variable speed as a factor of perturbation of transverse oscillations.

4. Transport flows as excitatory factors with a continuous and discrete structure (homogeneous, heterogeneous, group (unilateral, counter)).

$$TS_1^d = \left\{\left(m_i, V_i\right)\right\}_{Tn},$$

$$TS_2^d = \left\{\sum_{i=1}^{K}\left(m_j, V_j\right)_{ti}\left(t_i \in T_n\right)\right\}, \text{ at the same time } \sum_{i=1}^{n} m_i \le M_d$$

where TS_i – traffic stream, m_i – mass of transport unit of movement, V_i – movement speed, T_n – group time, M_d – maximum mass loading.

5. Model of action of transport flows on the bases and platforms with the established aggregate lines of production of high-quality polygraphic multicolor production (Fig. 2).

6. Laser probing method of the spatial structure study object which is the basis for the development of information and measurement systems.

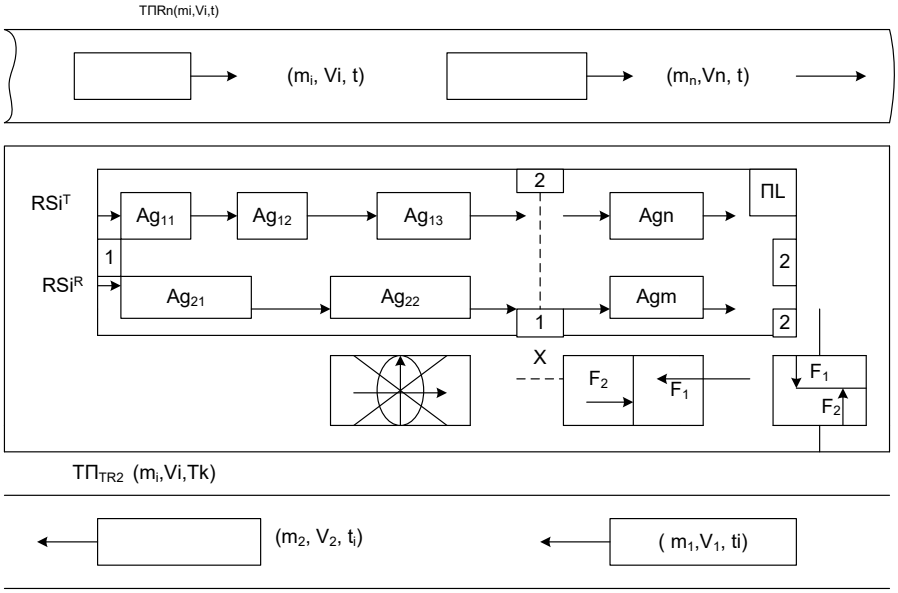

Fig. 2. Model of structural action of factors of influence on the aggregated structure with a common platform: RSi^T – input flow of resources; RSi^R – output flow of resources for the time interval T; $\{Ag_i\}$ – aggregate structure on a common platform $(ПЛ)$; [1], [2] – coordinates of the installation of the laser system for different sensing options.

Active factors influencing the dynamic and structural stability of spatially distributed objects.

According to the conducted researches, such system active factors of action can be allocated.

Absence of effective control structure and object reliability constructions systems, dynamics of destruction due to deformation shifts under the active factors influence – automobile and transport streams and natural dynamic factors:

1. Conflicts and incomplete knowledge that lead to errors at the stage of designing a spatial structure:

– incomplete data about the object, structures, materials, dynamics, factors, loads, destructive factors, loads, destructive forces;

– gaps in the knowledge system of designers lead to systemic and structural errors.

2. Conflicts that arise during operation in the absence of data and knowledge of staff:

– structure of dynamic loadings and their changes on long and short time intervals;

– seasonal, natural factors, cataclysms that lead to structural damage.

3. Transport flows as stimuli of oscillations of spatial structures of bridges, platforms:

– change in the level of reliability and aging of metal and concrete supports, platforms due to vibrational oscillations;

– dynamic destruction of materials, inadequacy of the project and facility structure to the requirements and trends of traffic and its mass parameters and reliability.

4. Exclusion of active factors on the oscillations of the soils of platforms and supports, deformation from dynamic perturbation and gravitational deformations.

Selection and processing of heterogeneous data on the state and dynamics of spatial objects with vibration.

The modern period of development of science and technology needs to pay more attention to basic research, theoretical generalization of known facts and the discovery of new ones, which become the basis for the formation of the knowledge base. One of the ways to obtain new information in stochastic perturbations is the synthesis of robust algorithms for data processing and creation on their basis of information-measuring systems for vibration control using laser remote sensing of areas of greatest stress and displacement of spatial structures [15].

Depending on the type of construction, the level of vibration (data sampling) is estimated by the method of sounding on the reflected beam (mirror surface on the structure) or direct projection sounding (photomatix installed in the control place of the structure) (Fig. 3).

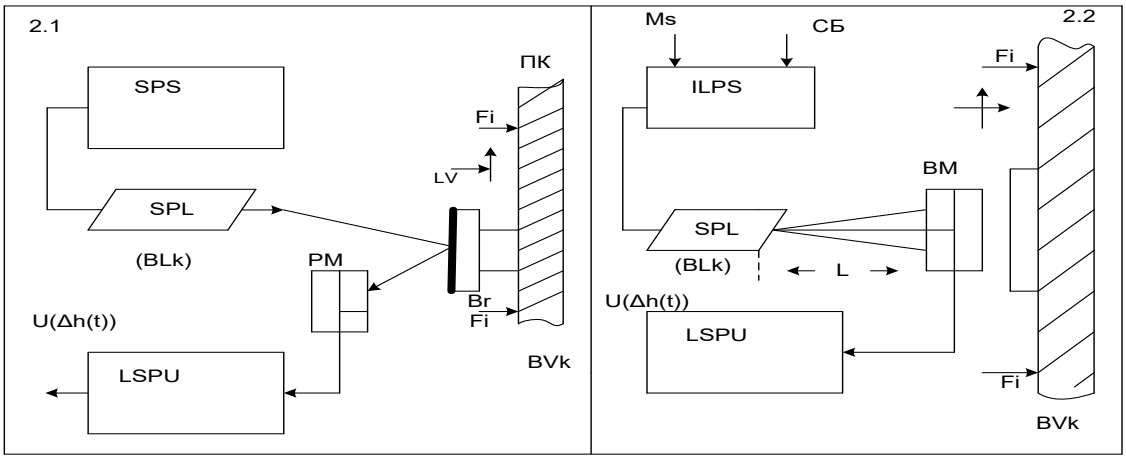

Fig. 3. Schemes of remote sensing of the vibration area of the structure (2.1. – on the reflected beam, 2.2. – projection direct probing)

Symbols in the diagram: NPL – semiconductor projection laser, SPS – stationary power supply, ILPS – integrated laser power supply (network or solar battery), PM – photomatrix with VM basis, LV – laser reflector with Br basis, BOLS – laser signal processing unit, $U(\Delta h(t)$ – vibration signal of the structure. Consider the scheme of vibration transformations of the laser signal when probing a spatial structure (Fig. 4).

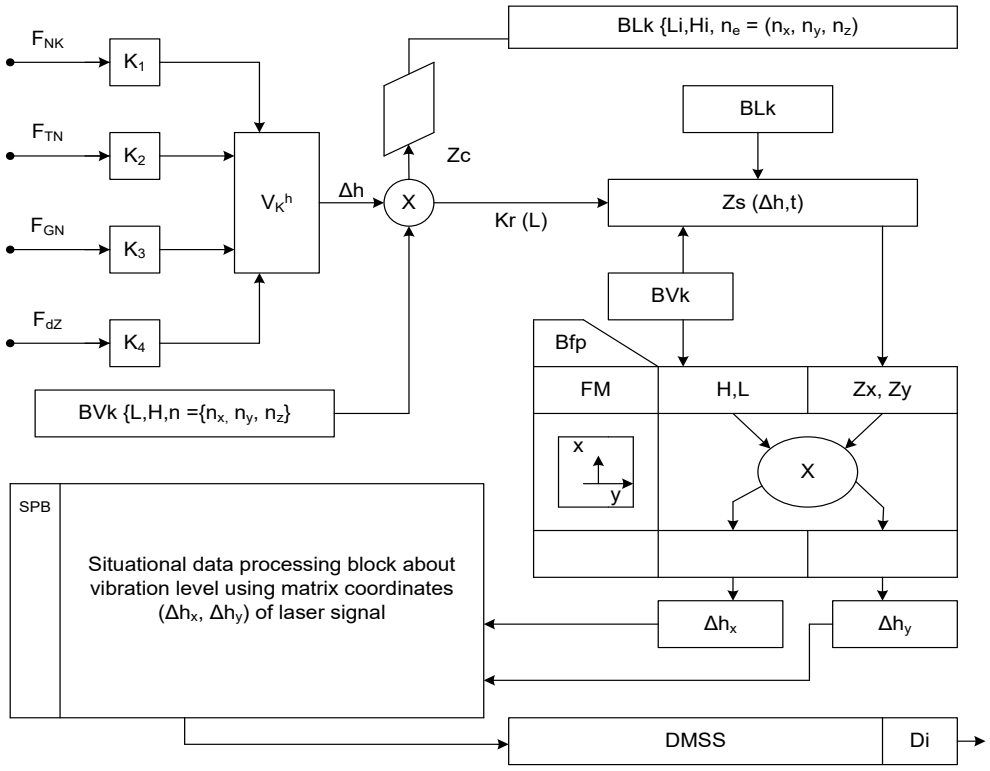

Fig. 4. Block diagram of laser signal transformations at dynamic displacements of the sounding area of the spatial structure: F_{NK} – structures dynamic stress factor; F_{TN} – dynamic transport stress factor; F_{GN} – gravitational stress factor of spatial structure; F_{dZ} – factor of dynamic shifts at seasonal changes temperature, humidity, wind pressure; V_K^h – he control displacement area of structural element; *BVk* – geometric domain of control basis; *BLk* – laser installation basis; *Zc* – probing signal; *Zs* – beaten signal; *Kr(L)* – laser beam scattering coefficient

Information transformations of laser signals in the process of measuring the dynamic displacements of structural elements can be represented as:

1. $\left\{\sum_{i=1}^{u} K_i F_i\right\} \xrightarrow{(t,\tau)} \left\{V_k^n\right\} \xrightarrow{Ar(t,\tau)} \Delta h(t,\tau)\big|_{BVk}$ – model of the process of forming the displacement of the structure control area due to the operator of the influence of factors $A_i = (t,\tau)$ at time moment t on interval τ;

2. $Z_C\left(P,d,F_s(\tau),t\right)\big|_{BLk} \xrightarrow{Az,\Delta} Z_S\left(P_S,d_S,F_S\,|\Delta h,t\right)\big|_{B_{\Phi\Pi}} \to A\left(\Delta h \to \Delta \mathrm{U}\right) \otimes Z_S\left(\Delta h\right) \to \left(\Delta h_x, \Delta h_y\right)$ – model of conversion of laser signal parameters in the region $\left(V_к^{\mathrm{n}}\right)$ probing the displacement of structural elements under the influence of dynamic factors (P_S – laser signal power, d – beam diameter, F_S – signal form).

$$Z_S\left(\Delta h, P_S, t, F_S(\tau)\right) \xrightarrow{V_к^n} \downarrow \otimes \xleftarrow{A_{\Phi M}} K_M\left(\Delta h \to \Delta \mathrm{U}\right)\big|B_{\phi M}$$

$$\Pi D\left(\Delta h_X, \Delta h_y \,|t, T_K\right) \leftarrow A \lg\left(\Delta U(h) \to \Delta \hat{h}\right)$$

The model of measuring data obtained in the process of laser probing of the area V_K formed at the output of the photomatrix in the form of a data stream $\Pi D(\ )$ based on the algorithm.

Based on the proposed models of measurement transformations, the information and control system of vibration measurements is synthesized.

Stochastic methods of ICS synthesis mainly use Gaussian models of changing the parameters of the object and the probing signal, with little attention paid to solving the problem of ensuring the robustness of systems and the algorithms stability and processing observation results procedures.

The main studies are conducted in the following areas of statistics and systems theory [1], which are based on [19, 5, 6]:

– probable models of random processes and fields to describe object vibration and influencing factors;

– procedures for detecting, recognizing, estimating parameters and filtering signals based on selected dynamics models that reflect the state of the technical system or spatial object at the current time;

– algorithms of spatio-temporal signal processing taking into account the stochastic structure of propagation channels and perturbation models for estimating real-time trajectories and trends of change of dynamic parameters;

– procedures for multi-criteria optimization of the decision-making process for management in conditions of data inaccuracy, which changes the load mode;

– procedures for dynamic assessment of the situation in energy-intensive facilities;

– algorithms for pattern recognition (spatio-temporal, situational) formed from data streams in different modes of operation of the object;

– procedures for analysis and synthesis of IBC to assess the state of spatial structures of technological objects with varying degrees of control;

– selection of indicators of signs of limit and emergency modes of the current dynamic situation in constructions concerning the target area of admissible parameters.

The classic approach to the structural synthesis of ICS consists of structure development based on the technical task within the existing analysis and synthesis methods, based on a given measuring system model without taking into account the target orientation. However, information on the study object structure, the conditions of its operation with limited resources, observability and reliability are not always fully inasmuch. First of all, when implementing the ICS synthesis procedure, one should keep in mind the goals of the technological object, which allows building a meaningful model and forming quantitative optimization criteria in the form of a system of quality functionalities.

With the stochastic nature of the functioning of the object control, it is often faced a situation of insufficient priori information.This complex problem arises especially when monitoring the state of technological spatial structures, with unidentified structure and functions, unstable in time and blurred priorities, local goals that have no strategic directions, and decision-making procedures that do not have systematic and effective technological support. In these cases, the principle of dual control of the operation process is used for decision-making, which involves the simultaneous use of signals as a means of studying the technological object, the trajectory of behavior under the influence of perturbing factors. But there are conditions under which optimal surveillance and management becomes impossible. This situation occurs when resource constraints or dynamic disturbances significantly exceed the level of informative useful signal. This leads to the disorientation of LIKS and making incorrect decisions, and in extreme conditions to an emergency situation. Under these conditions, the robustness and efficiency of the ICS, built on the classical theory of filtration basis and the automatic systems theory with feedback using the hierarchical structures of Masarovich, are lost. Problems of synthesis of LIX systems, as well as information aspects of ICS functioning as a dynamic situation image shaper in the control channel of the control system, are practically not considered in the literature, which makes the problem of robust systems synthesis. This requires finding fundamentally new approaches to the synthesis of LICS, taking into account the achievements of program-targeted and situational analysis, which allows to adequately reflect the situation in the target space of the DSS system, and analysis of information about based on the interpretation of the behavior of the trajectory of structures under the influence of active dynamic factors in space and time.

Laser sensing of oscillations of the spatial structure of the bridge with a long span

For detection and identifying transport infrastructure spatial fluctuations and large building structures, a projection laser probing changing method to the trajectory of elements at certain crisis points of structures has been developed. According to [6], the spatial displacements of the coordinates of the supporting structures can be represented as trajectories (Fig. 5).

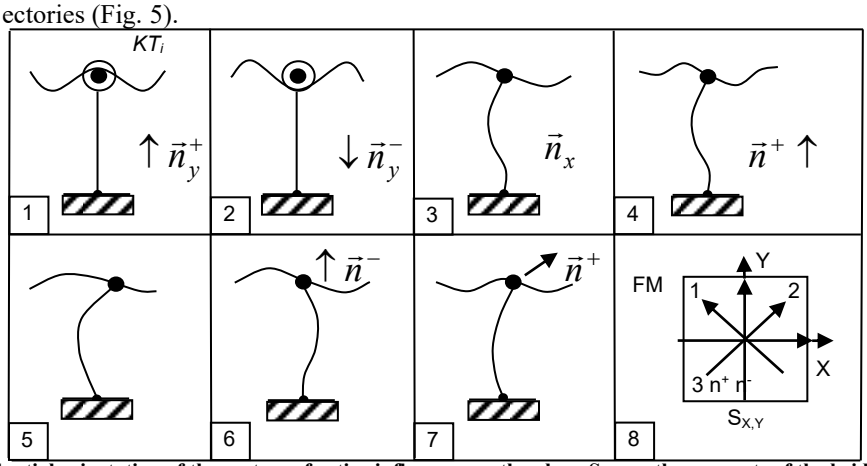

Fig. 5. Spatial orientation of the vectors of active influences on the plane $S_{X,Y}$ on the supports of the bridge with the length of the span from 20 m to 100 and more meters

To receive the spatial flux of laser signals, the photodetector matrix (FM) of the photodetector must have a 4-square structure to estimate the dynamics of the displacements of control points on the vectors $\left(\vec{n}_X, \vec{n}_Y, \vec{n}^+, \vec{n}^-\right)$ according to the difference equation [33]:

$$trak\ \Delta U^t_{\text{var}}\left(\vec{n}_X \Delta x\right) = K_M\left(U^+_{xt} - U^-_{xt}\right) = K_M K_{YS}\left(P^+_{Si} - P^-_{Si}\right)^t,$$

де P^+_{Si}, P^-_{Si} – the power of the received laser beam $\Delta U_{\text{var}}\left(\vec{n}_X \Delta x\right)$; K_M, K_{YS} – the coefficients of transformation of the matrix and scattering of the beam, P^+_{Si}, P^-_{Si} – variation of the voltage at the output of the channel for measuring the oscillations of the control point on the vector $\left(\vec{n}_{Xi}\right)$.

In Fig. 6. the scheme of possible oscillations of the bridge platform and data selection by laser probing of displacement control areas are given.

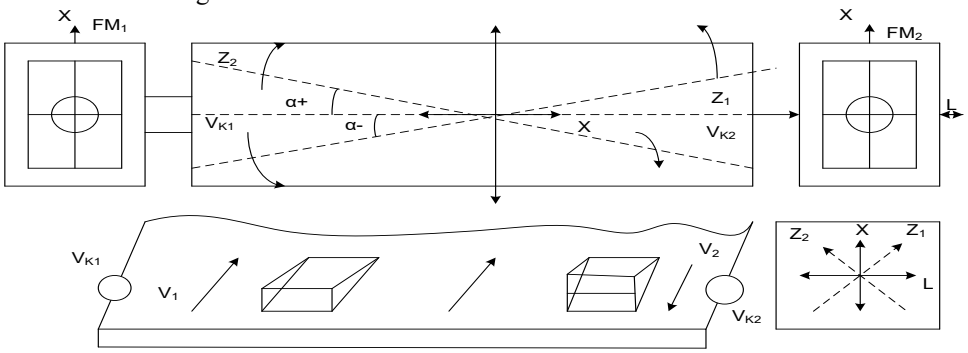

Fig. 6. Oscillations of the bridge deck at oncoming traffic flows with speed $V(t)$ and ground displacements of foundations and platforms

According to Figs. 4, 5 and models of information transformations the structural-information scheme of the method of projection laser sounding of oscillations of the control point of the structure is formed (Fig. 7) [9, 33].

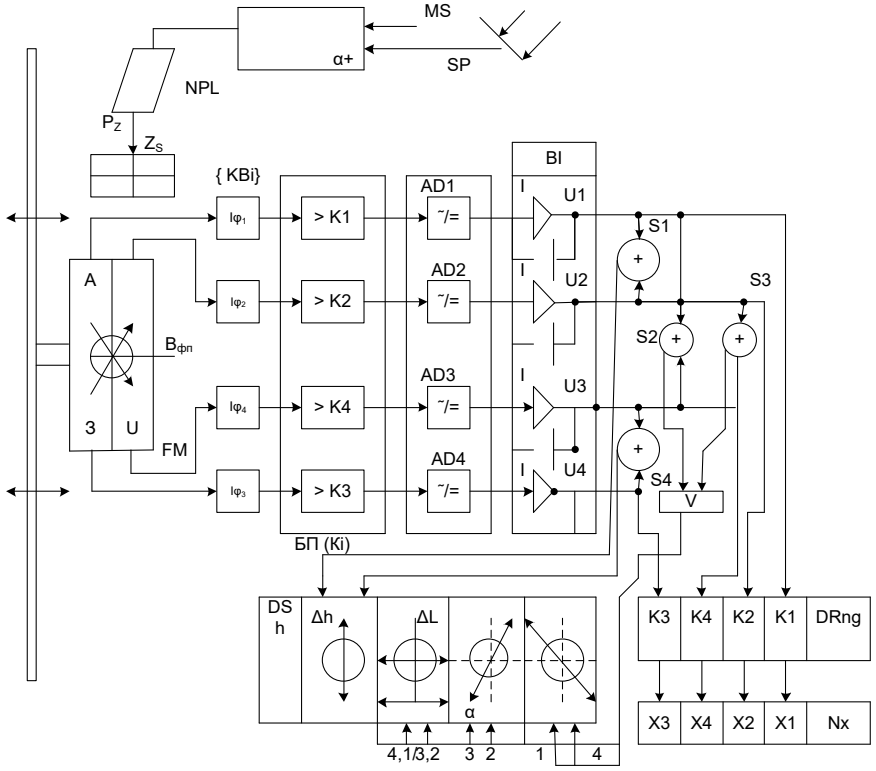

Fig. 7. Laser spatial vibrometer with emission detection (liquid events): *FM* – 4 square matrix photodetector; P_Z – power of the probing laser; P_S – received signal; $\{kB_i\}$ – channels of control of dynamics of oscillations in the base $\{x, y, n\}$ with input filters of signals $\{B_X\Phi_i(f_M, \Delta f)\}$ with frequency f_M and service of transmission Δf; $\{\text{БП}(\text{K}_i)\}$ – block of signal amplifiers with coefficient K_i; AD_i – analog signal detectors; $BI(U_S)$ – signal integrators unit; $\{S_i\}$ – operational signal adders; *DRng* – discrete rank load classifier;

N_X – digital indicator; DS_n – display complex indicator of dynamic displacements along axes (X, Y, Z^+, Z^-);

***MS* – mains supply system, *SP* – solar panels.**

Experimental research

Since special permits and equipment are required to study the vibration of bridges, foundations, platforms. Therefore, analogues were used to estimate spatial waves under the influence of perturbations, using high-temperature hydrodynamic flows of viscous fluid $\left(T^0C=[900-1100\right)$ based on the model of dynamic balance (loading – flow) of resources in different control modes (glass furnace – with lateral selection of molten glass) and laser sounding offset level.

Experiments to assess the perturbation of the glass mass surface by laser sounding were performed on the furnace (Fig. 8) of Rokytne Glass Plant for the period (2010-2011) and other furnaces for the period (1990-2015).

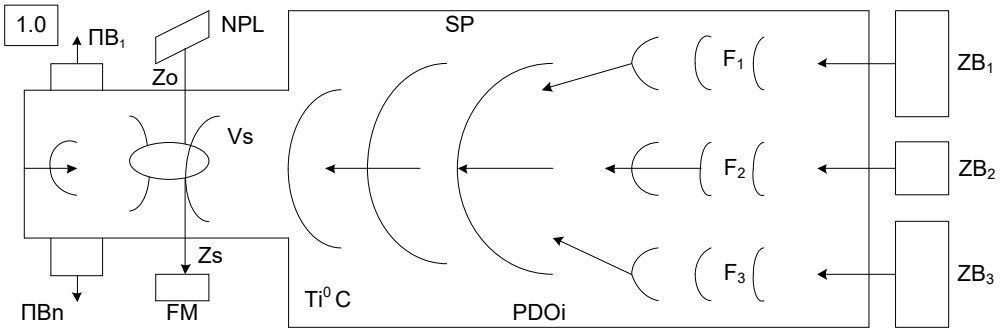

Fig. 8. Hydrodynamic model of surface solitons formationwhere: $\{ZB_1\}$ – loaders (continuous or pulse) charge; $\{ПВ_i\}$ – glass mass extraction devices; V_S – laser control area; PDO_i – spatial dynamic perturbations of the volume and surface of molten glass mass caused by perturbation factors $\{F_i\}$.

In fig. 8 shows the scheme of the glass furnace (SP), Soliton effects occur in the control zone when the direct wave of surface perturbation in the mass extraction channel meets the reflected one.

Graphs of the surface displacement trajectory in the sounding field and the dynamic soliton formation mode in the channel of mass selection. The graphs show changes in the object operation modes under different load modes (discrete pushing mechanism). During executing control commands on the loading mechanism inhomogeneous streams of the charge fall on the glass mass surface and excite longitudinal waves. Longitudinal waves propagate along the molten viscous surface, pass along the furnace into the mass extraction channel where under certain conditions solitons of the surface are formed (Fig. 9).

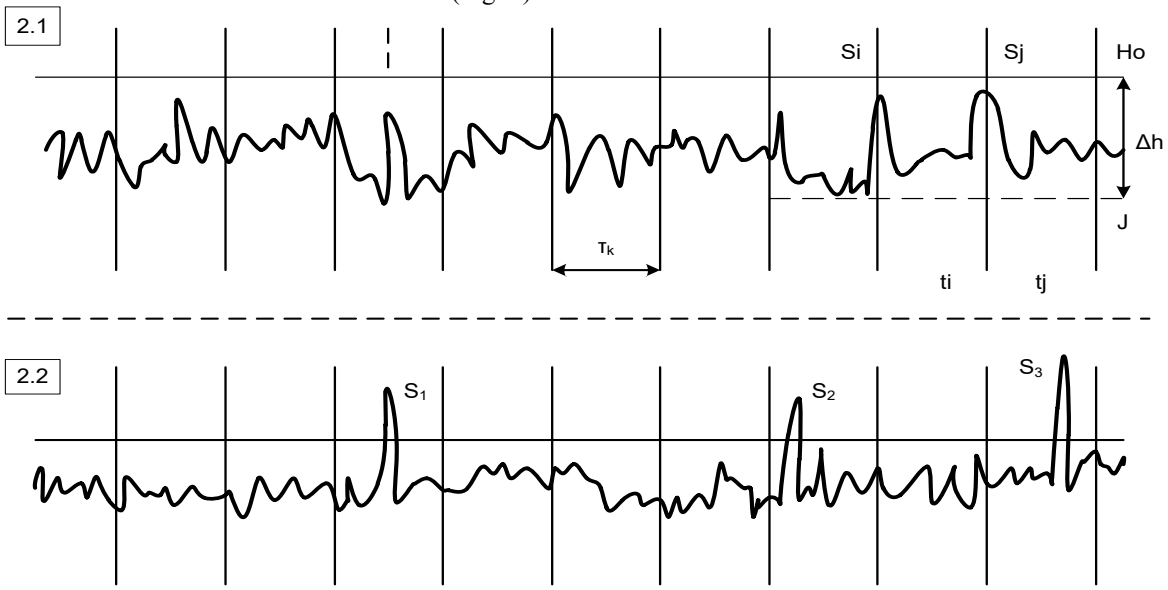

Fig. 9. Hydrodynamic model of surface solitons formation: (top) – the structure of the glass furnace, as a model; where SP – glass furnace in the contour image; $\{3B_1\}$ – charge loaders, as perturbing factors; V_S – speed and loading time $\left(V_{зі},\tau_{зі}\right)$; $\{ПВ_i\}$ – flow of product selection; $\{T_i^0C\}$ – surface thermal field $\left(T_i^0C-980^0C\right)$; *NPL* – semiconductor projection laser for surface sensing *FM* – photomatrix of laser signal Z_S reflected from the control area V_S; $\{F_i\}$ – factors that disturb surface waves at dynamic loading of the resource $F_i=\{m_i,\tau_{iз}\}$ – mass m_i, during the time interval $\tau_{iз}$.

2.1. The graph of the trajectory of oscillations of the viscous mass surface in the probing region V_S, is characterized by the following parameters: $\left(\tau_K = 60\ сек\right)$ control time interval, $\Delta\tau_{ij}$ – distance between solitons, Δh_i – emission amplitude, Δ_n – rank of the pulse amplitude scale at $\left\{N_S = 0.5\ мм,\ A_n = 0{,}005\ мм\right\}$.

For: 2.1. – soliton amplitude: $S_i = 0.3\ мм,\ S_j = 0.2\ мм$;

2.2. – soliton amplitude $S_1 = 0.2\ мм,\ S_2 = 0.25\ мм,\ S_3 = 0.3\ мм.$

In fig. 10. The scheme of soliton formation is given.

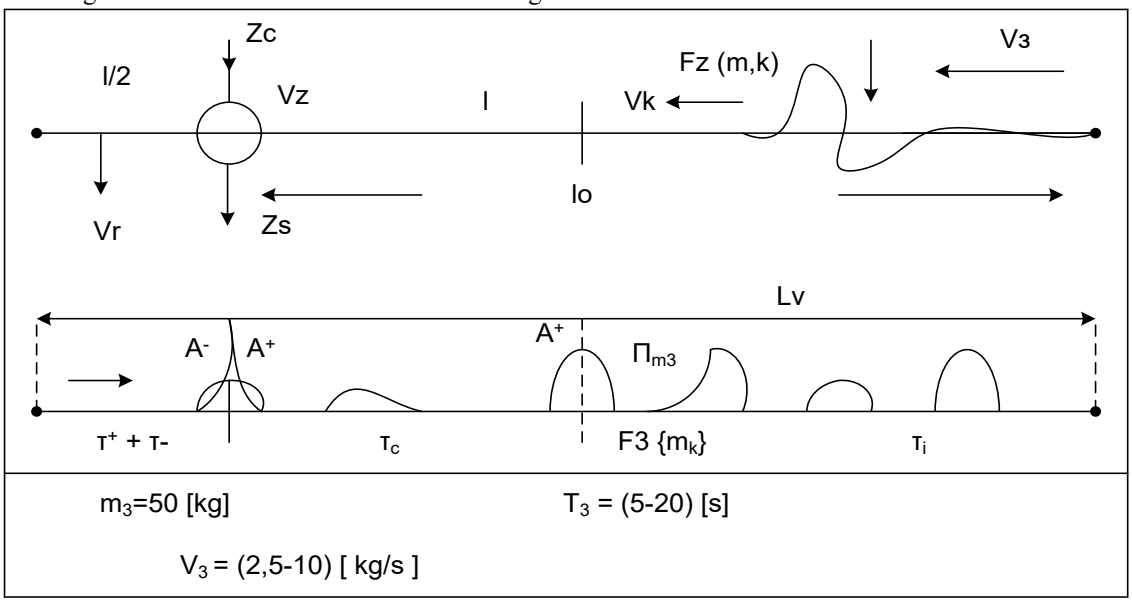

Fig. 10. Model of formation of soliton impulse at loading of a resource on a bath entrance

Designations: V_Z – sounding area, Z_C – sounding signal, Z_S – reflected signal, L_V – bath length, L_O – distance to control area, V_K – loading flow rate, V_r – mass sampling rate at object output, $F_3\left(m_K\right)$ – perturbation factor at pulse loading of resource, V_3 – flow rate loading, τ_e – wavelengths at perturbed surfaces: 1. $L_0 = 16\ m$,

2. $L_V = 20\ m$, 3. $U = l = 4\ m$.

Condition of soliton perturbation on the surface in the control area: If at a distance $\left(l/2\right)$ to $\left(V_Z\right)$ the perturbed surface of the impulse with an amplitude A_1^t at the time t_1, passing to the point l_k at the time t_1 the impulse A_2^+ appeared, then at the meeting at the moment $t_2, \left(\tau = t_2 - t_1\right)$ formed a soliton with amplitude $A_{sl} = \left(A_1^+ + A_2^+\right)$

For finding the necessary TS control strategy, it is required to conduct a simulation game on the model $\left\langle CUS \leftrightarrow TS\right\rangle$ for different classes of perturbations, determine stable Lyapunov regions I_{ZTS} in phase space based on the interval partition, and then the parameters of the strategy for controlling the reliability of spatial structures (Fig. 11).

Typical design errors of ASI, IAS

Expert long-term experience from one of the authors (FMI_AN of Ukraine, Soyuzavtomatika, Center for Strategic Studies – Scientific, Production, Design) indicates that the shortcomings of the project are identified in the first 72 hours after launch of automated production systems of both structural and functional type.

Accordingly, the following mistakes, made by designers can be identified in the development of ACS (automated control system) for both organization and production.

1. Structural errors through the ACS project developing

1.1. The structure of the ACS does not meet the objectives and does not ensure the solution of problems in full.

1.2. The project of the structural organization does not provide for the separation of automatic and operational control and interfaces for its coordination.

1.3. The block diagram of the connection of units does not ensure the safety of energy-intensive units.

2. Mistakes in the orientation of ACS management systems regarding the way to solve a problem situation in a complex system with a hierarchical structure.

2.1. The selected strategies for managing the system mode do not provide a solution to the problem of keeping units within the specified limits.

2.2. The procedures for the transition to the limit area of operation () at maximum loads due to incorrect estimation of the allowable values of the parameter intervals and the intervals of the terminal time of exit from the pre-emergency situation are unclear.

2.3. Incorrectly substantiated logical procedures for the formation of management teams (automatic and operational) to ensure the guaranteed functioning of the ACS and the technological process.

2.4. Professional, cognitive, psychological characteristics and skills and experience of operational staff do not meet the level of requirements for functional management actions.

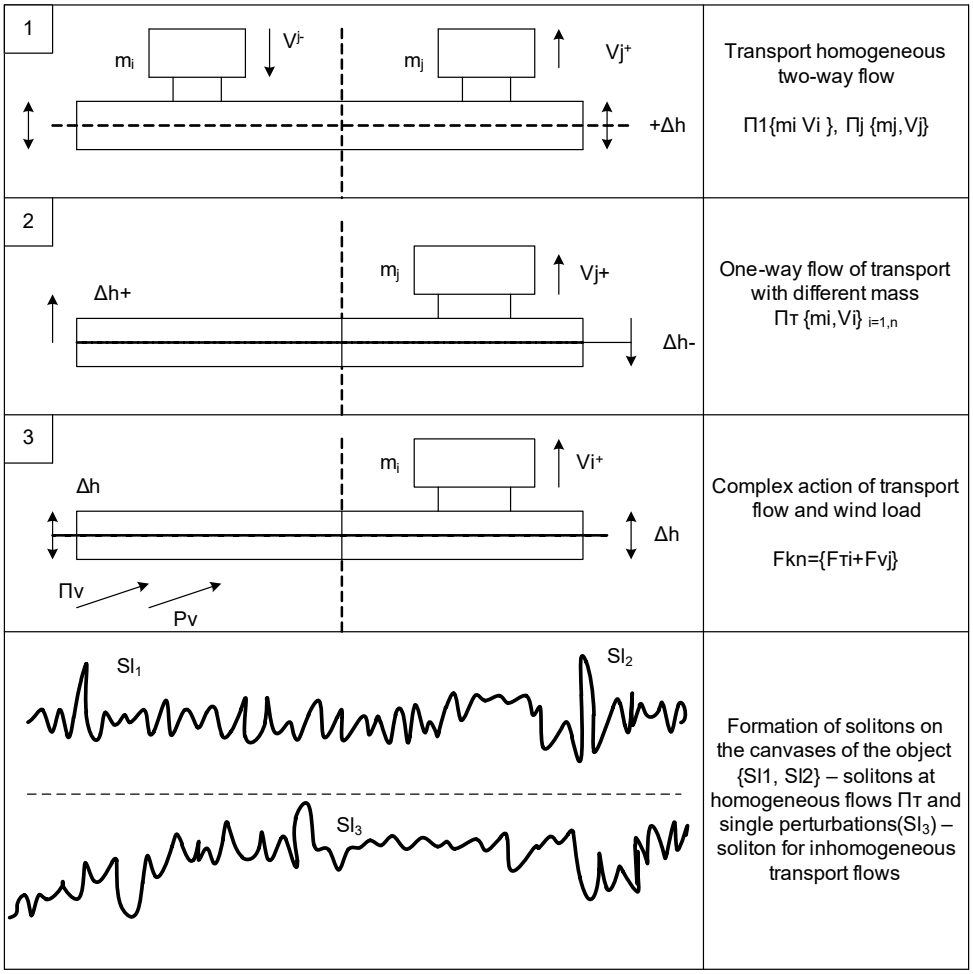

Fig. 11. Models of dynamic loads on the canvas of the bridge

3. Non-coordination of dynamic characteristics of resource flows and dynamic characteristics with the mode of technological process and control actions, as necessary for the sustainable operation of a man-made system or organization.

3.1. Uncoordinated design and regulatory parameters, which is necessary to assess the dynamics of the process of exit from the pre-emergency mode at maximum loads of energy-active units.

3.2. The designers have not fully agreed on the requirements for the dynamics of technological process management, under the influence of obstacles to the control structure and information channels.

3.3. Failure to reconcile the powers of operational personnel to make decisions in extreme man-made situations with the upper levels of the hierarchy of systems.

3.4. Insufficient protection of data collection and transmission channels from information attacks, which leads to disorientation of personnel at the operational and strategic levels to assess the content of the situation.

According to a set of test tables for assessing the level of competence of personnel, candidates for project and operational activities, it is possible to determine the level of cognitive – professional coefficients that determine the projected level of efficiency during project work and operational management.

$Sh\left(KF_i\right) \in \left[0,00-1,0\right] = I$ and the risk scale.

$Sh\left(\alpha_{risk}\left(KF_i\right)\right) \subset [0,00-1,0] = I_{\alpha_{risk}}$ build a table of balance of the allowable level of risks of selection of project and operational staff, which provide a guaranteed implementation of the project (Table 1).

Table 1

Balance of acceptable level of risks of selection of design and operational personnel

№	$\sum KF_i$	$\sum \alpha_{risk}$	Type of intellectual, project and management operations
1.	(0,9-1,0)	$\alpha_{risk} < 0,1$	Project work – development of ideas and strategies
2.	(0,8-1,0)	$\alpha_{risk} < 0,15$	Development of the structure and management strategies of the system according to the project objectives
3.	(0,75-1,0)	$\alpha_{risk} < 0,2$	Choice of architecture of control systems and TP units
4.	(0,6-0,8)	$\alpha_{risk} < 0,2$	Document flow, ancillary work
5.	(0,75-1,0)	$\alpha_{risk} < 0,3$	Installation and adjustment works and operational management
6.	(0,5-1,0)	$\alpha_{risk} < 0,4$	Installation work and maintenance
7.	(0,2-0,4)	$\alpha_{risk} > 0,7$	Installation of work inadmissibility

Conclusion

Based on systems analysis and cognitive concepts, crisis situations that arise in man-made systems with energy-intensive factors are considered. It is argued that only if these factors are taken into account at all stages: from design to construction and operation can ensure a high level of trouble-free operation of man-made regional and global structures.

In order to ensure a high level of reliability of man-made systems, it is necessary to take into account in the design process active, informational and cognitive factors influencing the design and implementation of the project, accounting the development of real dynamic situations.

References

1. Sikora L., Lysa N., Martsyshyn R., Miyushkovych Y. Information Technology for Assessing the Situation in Energy-Active Facilities by the Operator of an Automated Control System During Data Sampling. Lecture Notes on Data Engineering and Communications Technologies. 2022. Vol. 77. Pp. 177–187.

2. MarshalI V. The main dangers of chemical production. M.: Mir, 1989. 671 p.

3. Dryzdel D. Introduction to the dynamics of fires. M.: Stroyizdat, 1990. 424 p.

4. Haley E. J., Kumamoto H. Reliability engineering and risk assessment (Reliability engineering and risk assessment) M.: Mashinostroenie, 1984. 528 p.

5. Timoshenko S. P. Fluctuations in engineering. M.: Nauka, 1967. 444 p. / Vibration Problems in Engineering, bu Timoschenko. New York - 1955.

6. Dynamics of high-speed transport / ed. Tivilov T. A. - Transport, 1988. 215 p. / Dynamics of High-speed Vehicles, Springer Verlag, V. No. 274.

7. Timochenko S.P. Fluctuations in engineering. M. Science. 1967 - 444s.

8. G. Augusti, A. Baratta, F. Casciata. Probabilistic Methods in Structural Engineering. – London. 1984.- 584p.

9. G. P. Karzov, B. Z. Margolin, and V. A. Shvetsova, Physico-mechanical modeling of fracture processes. St. Petersburg: Politekhnika, 1993. 391 p.

10. Poturaev V.N., Belobrov V.I. Analysis of the dynamics of mechanical systems. K. Vishcha school. 1989 - 151p.

11. Chikriy AA Conflict-controlled processes. K .: Naukova dumka, 1992. 384s.

12. Pavlov VV Conflicts in technical systems. K .: Higher school, 1982. 184p.

13. Pospelov D. A. Situational management: theory and practice. M.: Nauka, 1986. 288 p.

14. Roberts F. S. Discrete Mathematical Models with Applications to Social Biological and Ecological Problems. / Ed. Geiman A. I. M.: Nauka, 1986. 496 p.

15. Perhach OL, Podolchak N. Yu. Corporate conflicts and methods of overcoming them. Lviv: Ed. NU "LP", 2014. 192 p.

16. Kunchenko-Kharchenko VT Information and management documentation in hierarchical systems: Concepts of information security. Lviv: Ukrainian Academy of Printing, 2015. 376 p.

17. Gladun V. P. Decision planning. K.: Nauk. Dumka, 1987. 168 p.

18. Ackoff R., Emery F. On Purposeful Systems. M.: Sov. radio, 1974. 272p.

19. Dragan Ya. P., Sikora L. S., Yavorsky B. I. System analysis will become the basis of the modern theory of stochastic signals: energy concept; mathematical substrate; physical clouding. Lviv: NVF "Ukrainian Technologies", 2014. 240 p.

20. Dragan Ya. P. Energy theory of linear models of stochastic signals. Lviv: TsSD.1997. 361 p.

21. Durnyak B. V., Sikora L. S., Grunik A. I., Timchenko O. V. Laser control of the vibration of aggregates in the transport infrastructure. Lviv: Ukrainian Academy of Health, 2011. 157 p.

22. Hayes D. Causal analysis in statistical studies. M. Finance and statistics, 1981. 255 p.

23. O'Conor D., McDermat I. Systemic thoughts and search for extraordinary creative solutions. K.: Our format, 2018. 240 p.

24. Skurikhin V. I., Kvachev V. G., Valkman Yu. R., Yakovenko L. P. Information technologies in testing complex objects: methods and means. Kiev: Naukova Dumka, 1990. 320 p.

25. Mesarovich M., Mako D., Takahara I. Theory of hierarchical multilevel systems. M.: Mir, 1973. 344 p.

26. Ishimaru A. Propagation and scattering of waves in randomly inhomogeneous media. Moscow: Mir, vol. 1,

27. Filippov A. T. Multimeasure soliton. Science, 1990.- 288s.

28. Duryak B. V., Sikora L. S., Atonik M. S., Tkachuk R. L. Automated human-machine systems for managing the integration of hierarchical organizational and viral structures in the minds of risks and conflicts. Lviv: Ukrainian Academy of Health, 2013. 514 p.

29. Duryak B. V., Sikora L. S., Lisa N. K., Tkachuk R L., Yavorsky B. I. Information and laser technologies for the selection of data flows and their cognitive interpretation in automated control systems. Lviv: Ukrainian Academy of Health, 2017. 644 p.

30. Bychenok N. N., Gaiduk O. V., Mostovoy V. V., Tereshchenko V. S., Senchenko A. D. Predictive-analytical decision support system for regional security. K .: Control systems and machines, No. 4. 2000. - S. 88-95.

31. Vasilenko V. A. Geneza, zmist and ways of implementing the concept of international environmental safety. K.: Bulletin of the National Academy of Sciences of Ukraine, No. 7. 2017. - S. 89-96.

32. Shapar A. G., Mikheev O. V. Conceptual approaches to understanding the processes of anthropogenic destabilization of ecological systems. K.: Bulletin of the National Academy of Sciences of Ukraine, No. 3. 2018. - S. 56-66.

33. Sikora L. S. Laser information and simulation systems for control of TSS technological processes. T2. Lviv: Kamenyar, TsSD "EBTES", 1988. 445 p.

34. Grigoruk V.I., Korotkov P.A., Khizhiyak A.I. Laser physics. - K. MP "Lesya" 1994. - 478s.

https://doi.org/10.31891/csit-2022-2-8
UDC 004.9

Vira SHENDRYK, Yuliia PARFENENKO,
Valentyn MAIKOVSKYI, Denys YURCHENKO
Sumy State University, Sumy, Ukraine
Sergii SHENDRYK
Sumy National Agrarian University, Sumy, Ukraine

SUBSYSTEM OF COLLECTION, STORAGE AND VISUALIZATION OF OPERATING DATA OF THE DECISION SUPPORT SYSTEM FOR MICROGRID MANAGEMENT

The paper states that the introduction of various types of renewable energy sources into energy systems turns them into complex cyber technical systems and come to growing complexity of interaction processes in such technical energy systems increases the complexity of the management process. The complexity of management is especially increased when such an energy system is operated as a microgrid because in this case, and effective management is possible only if the decision-maker will have quality information support and be able to fully monitor all information flows. To ensure the efficient operation of the microgrid, it is necessary to collect and use in the analysis process data on the current state of the power system, on the parameters of the environment. The monitoring process must be performed in real-time. The difficulty of data collection is that the data comes from different sources and at different time intervals. To ensure the data collection process, a functional and mathematical model is proposed, which describes the process of pre-processing a large amount of collected data, their verification, and storage. Data exchange and storage for the monitoring process are provided by the operational database, the logical scheme of which is also proposed in this paper. For the successful formation of the solution, a large amount of controlled data and the results of forecasting the production and consumption of electricity are proposed to be implemented as a subsystem of visualization. It is the combination of operational monitoring of the current state of the power system with the decisions of intelligent decision support services through the operational database that provides the decision-maker with a convenient visualized form of information about the current state of the system and the optimal configuration of resource planning and use services plan.

Key words: renewable energy sources, microgrid, information support, monitoring, database, visualization

Віра ШЕНДРИК, Юлія ПАРФЕНЕНКО,
Валентин МАЙКОВСЬКИЙ, Денис ЮРЧЕНКО
Сумський державний університет, Суми, Україна
Сергій ШЕНДРИК
Сумський національний аграрний університет, Суми, Україна

ПІДСИСТЕМА ЗБОРУ, ЗБЕРІГАННЯ ТА ВІЗУАЛІЗАЦІЇ ОПЕРАТИВНИХ ДАНИХ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ УПРАВЛІННЯ МІКРОГІД

У статті визначено, що впровадження до енергосистем різного роду відновлюваних джерел енергії перетворюють їх на складні кібертехнічні системами, які потребують інтегрованого інтелектуального управління. Зростання складності процесів взаємодії у таких технічних енергетичних системах підвищує складність процесу управління ними. Особливо збільшується складність управління, коли така енергетична система експлуатується як мікрогрід, оскільки вона у цьому випадку має певну автономію у функціонуванні, і досягти ефективного управління нею можливо лише, якщо особа, яка приймає рішення буде мати якісну інформаційну підтримку, та мати змогу повністю відслідковувати всі інформаційні потоки. Для забезпечення ефективного функціонування мікрогрід необхідно збирати та використовувати у процесі аналізу дані про поточний стан енергосистеми, про параметри оточуючого середовища. Процес моніторингу необхідно виконувати у реальному часі. Це дозволяє відслідковувати різного роду зміни, та вчасно на них реагувати, змінюючи параметри мікрогрід. Складність збору даних полягає у тому, що дані надходять з різних джерел та у різні часові інтервали. Для забезпечення процесу збору даних запропоновано функціональну та математичну модель, які описують процес попередньої обробки великого обсягу зібраних даних, їх верифікацію та зберігання. Обмін даними та їх зберігання для процесу моніторингу забезпечується оперативною базою даних, логічна схема, якої також запропонована у цій статті. Для успішного формування рішення великий обсяг контрольованих даних та результати прогнозування виробництва та споживання електричної енергії запропоновано реалізувати як підсистему візуалізації. Саме поєднання оперативного моніторингу поточного стану енергосистеми з рішеннями інтелектуальних служб системи підтримки прийняття рішень через оперативну базу даних надають особі, що приймає рішення, у зручній візуалізованій формі інформацію про поточний стан системи та оптимальну конфігурацію планування ресурсів і план використання, який орієнтується на підвищення рівня електричних послуг.

Ключові слова: відновлювальні джерела енергії, мікрогрід, інформаційна підтримка, моніторинг, база даних, візуалізація.

Introduction

The global energy crisis, caused by russian invasion of Ukraine, requires the final abandonment of fossil fuels and the rapid transition to renewable energy sources (RES) usage for electricity generation in the housing and utility sector and industry for all sectors of the economy. Therefore, energy production, distribution, and consumption require the development and creation of more innovative and energy efficient approaches. Replacing traditional power generation with generation from RES also poses new challenges to electricity consumption and distribution, and requires the implementation of Smart Grid technology, as well as the integrated management of next-generation power grids. In turn, the modern development of Smart Grid technology turns it into a global

concept of energy interconnection, allowing it to replace traditional approaches to the management of production, distribution, and consumption of electricity with more efficient ones.

Thus, the new generation of Smart Grid is transformed into cyber technical systems with integrated intelligent control. This transformation, in addition, allows to save electricity, avoid transportation losses, and develop energy savings. Environmental friendliness and energy efficiency are often achieved by changing the paradigm of "centralized» to the "decentralized" electricity production through the various RES usage, reversing the logic of production and distribution of electricity, and the use of new software to perform functional energy optimization. Along with this, there are many questions related to digitalization the processes that accompany the electricity life cycle, from the moment of its production to storage, distribution, and consumption. The use of integrated intelligent management throughout the energy life cycle can increase profits, and such a transformation is aimed at more efficient use of available energy resources, especially renewable energy resources. In turn, the ability to intelligently manage complex cyber technical systems with RES is the ability of decision support systems (DSS) to collect data, convert it into information, store information, use the results of information processing for decision-making and change the power system behavior. It allows saving expert knowledge gained on the basis of previous experience. Conventionally, DSS tasks for Smart Grid energy management can be divided into those involving decision-makers and those that are not controlled by such a person. The quality of the decision very often depends on how fully the decision-maker receives information and can perceive it. In other words, the quality of management decisions depends on the quality of information support of the decision-maker. Such information support is formed by data collection, storage of data that provides monitoring, as well as the process of providing information in a user-friendly form for decision-making.

Related works

The development of the Internet of Things (IoT) [1] and smart home technologies [2] is designed to monitor and manage energy consumption in buildings. The combination of these technologies allows to monitor energy consumption in different rooms or locations or give information on the current state of functioning of energy sources. An example of successful energy management is monitoring and managing energy consumption via the Internet. The solution proposed in [3] allows for a review of how people use, store, and manage energy. Once connected to this system, any device becomes an intelligent space with a rich set of functions and capabilities for energy consumption monitoring and control.

It should also be noted the advantages of monitoring systems that perform monitoring at the moment. Because the observed objects are constantly changing or evolving, dynamic monitoring systems have the ability to predict the failure of the object to be observed or to determine the probability of danger at the time of its operation [4 - 5].

IoT technologies also facilitate the collection of information about the operation of solar panels or charging stations and monitor the movement of the sun by using special trackers to monitor the sun [6], which increases the amount of electricity generated by about 35-50% [7], as powered by solar panels behind the sun.

Analysis of the results of the previous study revealed that to increase the effectiveness of the decision, information from the decision support system to the decision maker should come in the form of charts, graphs, tables, etc., which visualize data for energy management. Consider a few software products that visualize real-time monitoring data.

The first application was the Microgrid Digital Control Solution from Schneider Electric [7]. It is a desktop software that allows viewing real-time data with a simulation function, where you can see where and where the data comes from, as well as view weather conditions. Additionally, you can see the maximum voltage and energy consumption percentage. In the interface of this software, it is possible to view in the form of graphs how much energy is produced. Its feature is the ability to view individual network components and the percentage of their energy.

The application "Grafana" was also considered [8]. The Grafana Dashboard collects information and displays it in a convenient form for the user via the web interface. The Grafana software consists of a graph of time series so that the operator can observe the evolution of the measured temperatures. The user can zoom in on the graph, as well as select the time interval for visualization. In addition, the sensor illustrates the instantaneous value of one temperature to get a first look at the value on the sensor. Another element that was considered relevant is the table, which presents the indicators, namely the current, minimum, and maximum temperature for each sensor.

The third studied application is a modification of the MathLab software, more precisely its library. Simscape Power Systems is a standard SIMULINK library. It consists of a database of electronic components, as well as modeling and analysis tools for power systems. The library is also designed to develop and test control systems in integration with power systems; this makes it suitable for microgrids modeling [9]. This library allows performing a simulation, as a result of which the program displays visual indicators in the form of graphs depending on the specified parameters.

Consider a full-featured application for real-time microgrid simulation ETAP Microgrid Management [10]. After enabling the simulation, you can view detailed indicators of individual devices in real-time, as if physical devices are now connected. The application displays devices connected to the system and provide the ability to track

their performance in real-time. The program can display information in the form of graphs. You can view the performance of a separately connected device. It is possible to view a specific energy capacity in the form of a graph, as well as individual indicators such as speed, temperature, voltage, humidity, etc. in the form of a bar chart and a regular table. It is possible to view individual indicators in the form of a pie chart.

The functions of the information system, which is currently being developed but is attracting attention, are also considered - it is Monash Microgrid [11]. This system allows you to visualize the microsystem as a 3D model. More precisely, it visualizes real buildings on a Google map and depicts the system of connecting power and its connection to buildings.

It would also like to consider the oldest but still popular visualization system - LabVIEW Visualization [12]. The graphical interface is designed using LabVIEW software to visualize system parameters in real-time. As an instance, you can visualize the values of direct current inverters, such as direct current voltage, solar panel power, etc. Using this software, it is possible to make several modes of data visualization in real-time.

After analyzing the existing information technologies and software tools, it can be concluded that in order to collect, store and visualize data in the DSS and to organize effective interaction of the decision-maker, the DSS must provide real-time data visualization and support for several types of data visualization.

Purpose

Increasingly, engineers create smart microgrids with renewable energy sources - systems that can operate separately from the centralized power grid. But such microgrids often have problems related to their internal technological limitations of power supplies [13]. Because of this, it is usually impossible to determine how efficiently energy is produced and how efficiently it is used without monitoring. In this regard, there is the problem of how to perceive the collected data and control the energy of the end consumer. To solve this current problem the tools and methods of data collection, storage and visualization are used, which in turn determines the relevance of this study. Through operational monitoring of the current state of the microgrid system, intelligent DSS services can provide users with a convenient visualized form of the optimal energy resource planning and improve the level of electrical services [14].

The object of the study – to provide information support for the decision-making process in the energy management of microgrids.

The subject of the study – information technology of data collection, storage, and visualization in the energy management of microgrids.

The aim of the study is to develop information technology for data collection, storage, and visualization for the DSS system in the energy management of microgrids. This information technology should be implemented in the subsystem of collection, storage, and visualization of operating data of the DSS for microgrid management.

Proposed technique

In the study, we consider a typical small microgrid, which consists of a set of photovoltaic panels, wind turbines, and a common energy storage bank, and also has a connection to the external power grid.

The process of decision support in the management of such a microgrid pre-includes the collection and processing of data on weather data, the current state of the power system, determining the forecasted value of electricity generation and consumption, and quality indicators, and consists of recommendations for the optimal operating model. This process is divided into sub-processes that are implemented by the respective DSS subsystems. Some of these subsystems have auxiliary functions of data visualization. Appropriate technical and software tools for data collection, processing, and storage are required for the implementation of information collection and pre-processing.

Data that enters the DSS can be input once by the authorized user when it's registering in the system, as well as entered from external information sources online. In turn, operational data are divided into data coming from the automated control system and characterizing the current technical condition of the object of observation - the microgrid, and data collected from external information sources - data on forecast meteorological conditions. At the preliminary stage, it is necessary to check the accuracy and integrity of data transmitted from the automated control system, as well as from weather forecast websites. After the validation collected data are stored in the database. The data obtained online are used in evaluating the list of alternatives to the modes of the microgrid operation, as well as in calculating the criteria for choosing a solution regarding the microgrid operation mode. The decomposition of the process of collecting and pre-processing information in the form of a diagram in the IDEF0 notation is shown in Figure 1. The components of this process are "Data Collection", "Data Validation", "Storage to Database" and "Data Conversion".

All monitored data can be described as a set $M_p = M_{po} \cup M_{pi}$. It consists of two subsets:

- subset M_{po}, that consists of parameters that are collected from external information sources (these data are variable in time) and entered by the user;
- subset M_{pi}, that contains the calculated parameters.

Set of the parameters collected and entered by the user M_{po}, can be presented in the form:

$$M_{po} = \{Mwo, Mres, Mg, MPl, M_{tech}\},$$

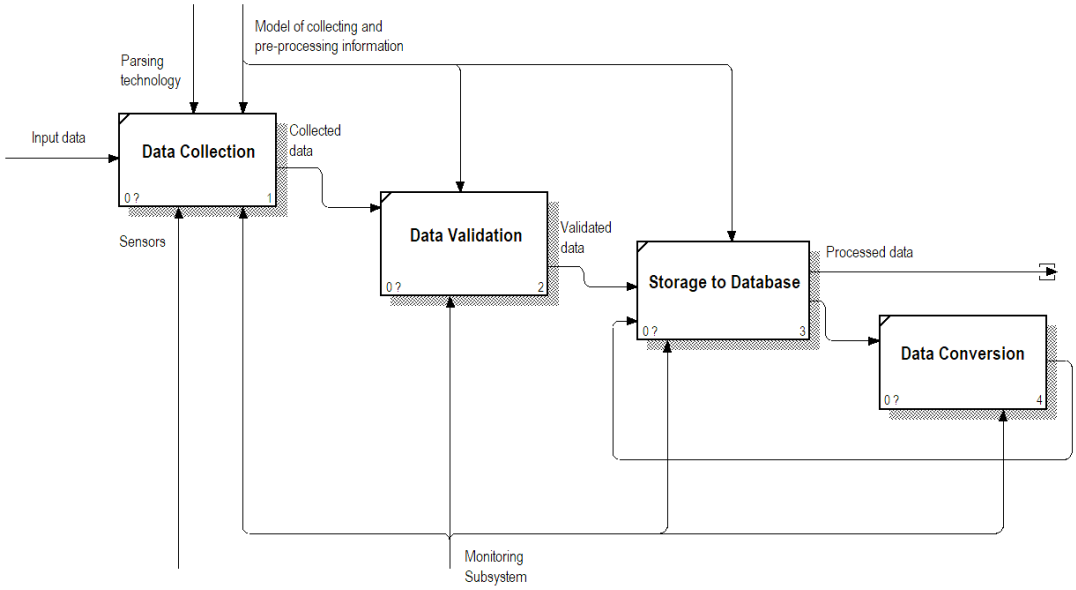

Fig. 1. Decomposition of the process of collecting and pre-processing information in IDEF0 notation

where M_{WO} is a set of forecast weather conditions; M_{res} – a set of data on the available microgrid with RES configuration; M_{Pl} – a set of data on geographical location of microgrid with RES; M_g – a set of data on the local distribution networks to which it is planned to connect the microgrid; M_{tech} – a set of data that characterizes the current technical condition of the microgrid (for example, the battery charge level).

Data M_{res}, M_g, M_{Pl} are entered once at user registration, data M_{WO}, M_{tech} are collected from external sources real-time, в they are the main factors that affect the microgrid current state and are crucial for the formation of energy management decisions.

Model data collected from external sources (M_{wo}), can be presented as a set of weather forecast data that are real-time collected. M_{wo} is described by an ordered set of elements:

$$M_{wo} = \{(t, E, T, V)\},$$

where t – time interval for which meteorological indicators are provided on the weather site (hours);

E – the level of insolation and precipitation in qualitative characteristics (clear; variable cloudiness; cloudy; cloudy and precipitation);

T – temperature, (°C);

V – wind speed can also be represented by the range of values from initial to final (м/с).

The data flow diagram in the process of collecting, storing, and displaying information to the user is shown in Fig.2.

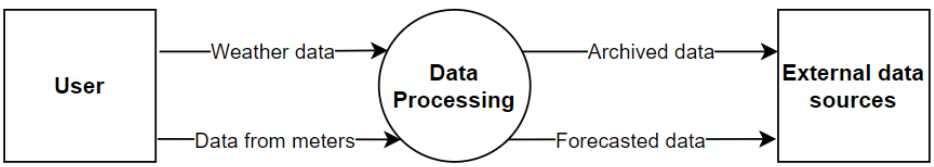

Fig. 2. Data flow diagram

Data on the current state of the microgrid enters the operating base from various sources, such as sensors or meters. At three-hour intervals, forecast weather data is recorded to provide forecasts of energy generation by devices for the day, three days, or a week ahead. Data from microgrid devices are received every hour. These data are visualized on the web interface and used for data warehouse filling. The data in the data warehouse should be updated at midnight Greenwich Mean Time in order to minimize delays in the operation of the entire system.

Figure 3 shows a logical model of the operational database that stores current information.

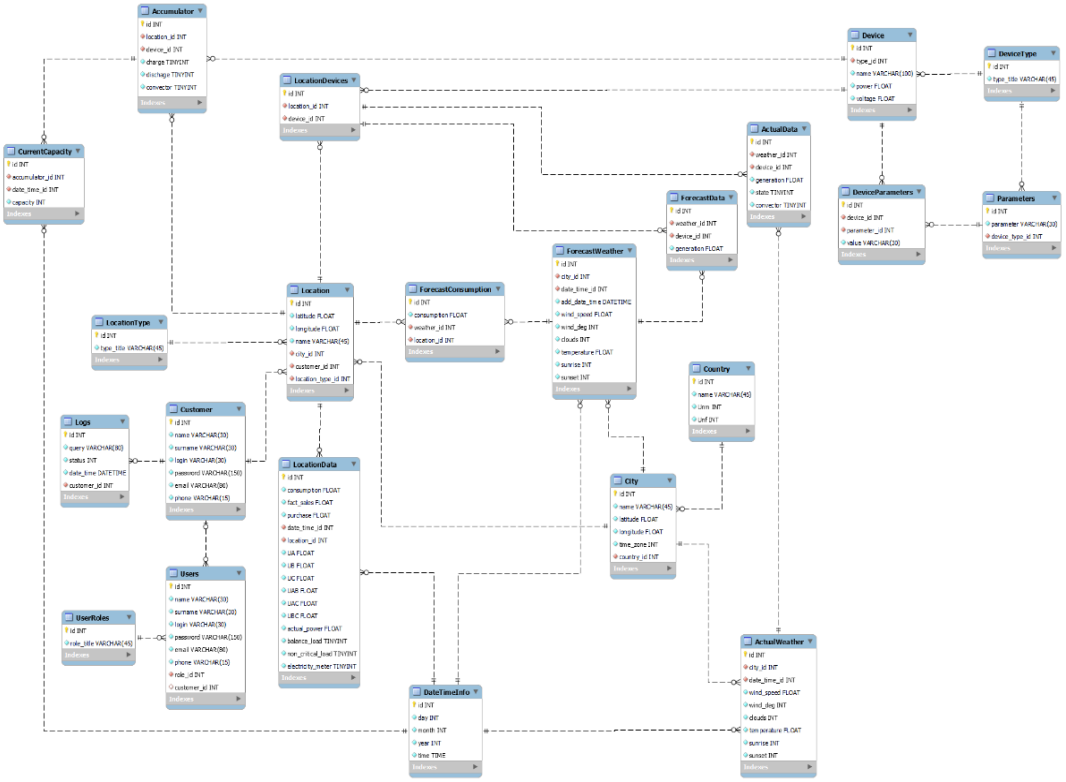

Fig. 3. Logical model of database

Information support of the decision-maker is provided through the subsystem of visualization of monitoring results and analysis. The collected data, processed data and data from analysis process are displayed in an easy-to-understand format in the form of diagrams, graphs and tables.

Thanks to the graphics module, which allows to build different types of graphics and configure them, provides visual output, namely:

– construction of graphs with data updates every 5 minutes with data display: on the X axis time (day, month, or week), on the B axis indicator of electricity use.

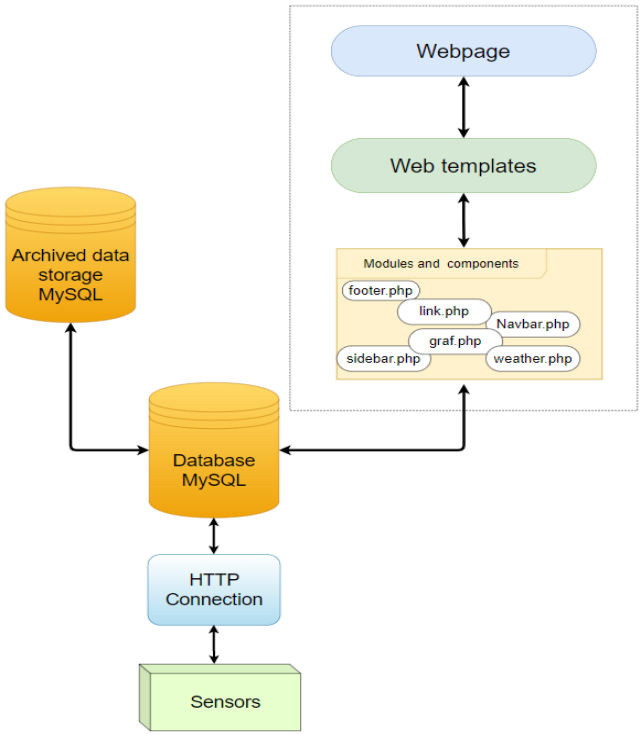

Fig. 4. Architecture of data visualization subsystem

- display tables with the ability to add and edit data with information about users, energy data, etc.;
- display of weather data as a separate widget.

The subsystem of data collection, storage and visualization allows to differentiate access rights of separate groups of users:

- user – does not have full access to the system, namely it can view data and add them to the database;
- administrator – has full access to the system and its settings, such as the user, as well as the ability to add and remove users, delete information from the database. The use case diagram is shown in Figure 5.

The visualization subsystem performs the following functions:

- displays data coming from the database in the form of graphs, tables and charts;
- implements saving data in .png format;
- implements fast work of updating data in the system.

The visualization subsystem allows:

- personal data editing – allows to change or add personal data about the client or administrator;
- authorization – allows to authorize in the system for further work;
- registration – gives the client the opportunity to register in the system;
- remote control for microgrid – allows to turn on or off a separate device in the microgrid;
- view microgrid energy data – the user can view all data in the microgrid;
- editing microgrid data – depending on the access rules, the user can edit the data in the microgrid.

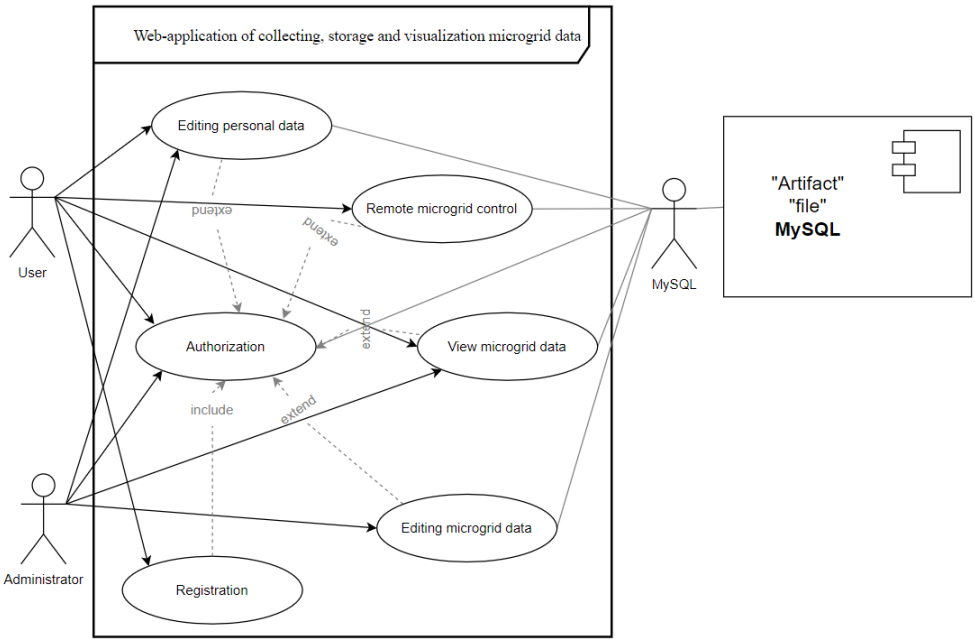

Fig. 5. Use Case Diagram

The data visualization subsystem consists of components and modules, which are listed in table.

Table 1

Modules and components

№	Modules and components	Description
1	link.php	All external links used in the web application
2	grafscript.php	The module contains settings for graphs
3	navbar.php	Navigation menu module
4	weather.php	Configure the weather API used in the web application
5	footer.php	Footer with contact information
6	include_database.php	Database connection module
7	config.php	An application configuration component

The data enters the visualization subsystem from the database. Information is obtained from the OpenWeatherMap API and devices, using the HTTP protocol. The visualization subsystem is created with the help of modules that are interconnected and form full-fledged page templates, and are connected into a full-fledged data visualization subsystem, which provides informational support to the decision-maker in DSS management of microgrids.

Results

Here is an example of working with the data visualization subsystem. After confirming all the data, the user gets to the main administrative page (Fig. 6), where we can view the current weather indicators (Fig. 7).

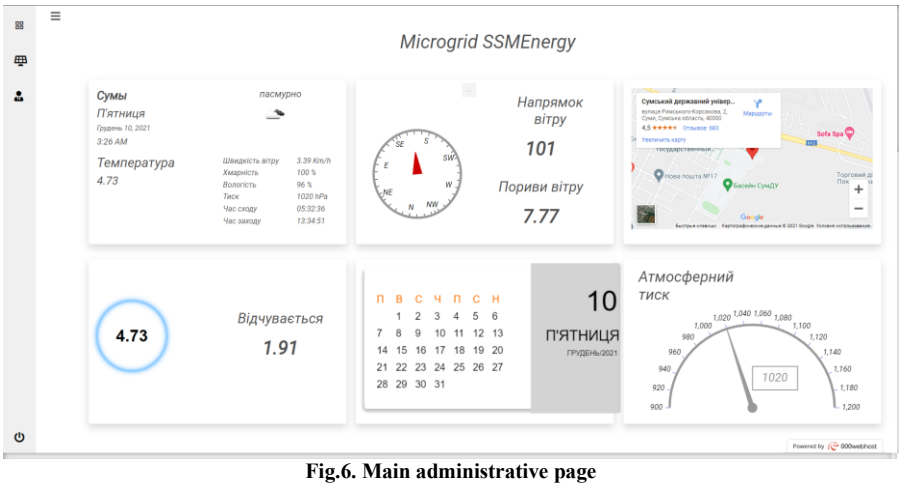

Fig.6. Main administrative page

Fig 7. Weather indicators

Below the weather indicators on the same page, the user can view the archived information on the performance of solar panels (fig. 8).

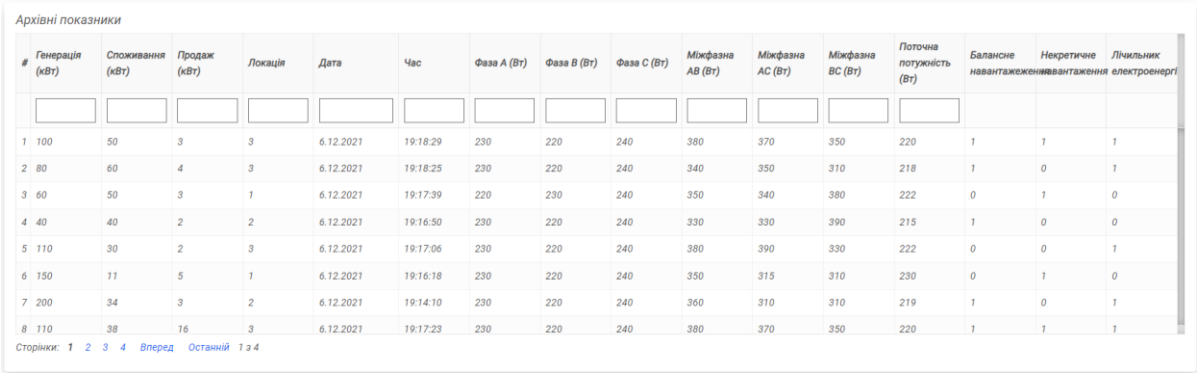

Fig 8. Table with archived data

Below from the presented table it is possible to view the voltage of the phases of solar panels in the form of a bar plot, which is presented in Figure 9.

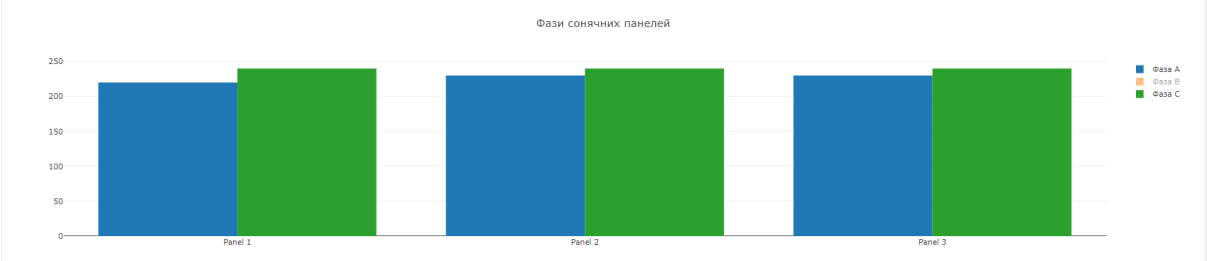

Fig 9. Bar plot of solar panel voltages

Also, you can view information about the voltage of solar panels relative to their power, in the form of a scatter plot (Fig.10).

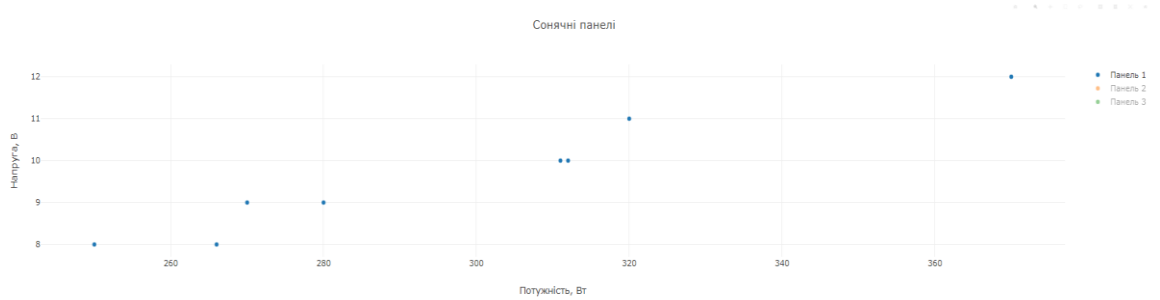

Fig 10. Scatter plot of solar panel voltage relative to their power

Conclusions

The paper presents information technology for data collection, processing, storage, and visualization. Information technology is implemented as a subsystem of the decision support system in the management of microgrids with renewable energy sources.

Previous analysis of technologies and methods that support the collection, storage, and visualization of monitoring data, confirmed the relevance of the study because in the decision-making process in microgrid management to improve the quality of the decision is possible only by improving the information support of the decision-maker.

Developed functional model of data collection and storage, the mathematical model of data monitoring, allow to formalize these processes and provide effective monitoring of a large amount of heterogeneous data at different time intervals.

An operational database has been developed to store the information received in the process of data collection, forecasting of consumption, and production of electricity, in connection with the subsystem of data visualization. The operational database stores information on the current state of the power system and short-term forecast data.

The visualization subsystem outputs data in a user-friendly form. It is designed for different categories of users and provides each user with information following his authority. This subsystem will be used as the component of the decision support system in the management of microgrids with renewable energy sources.

The developed subsystem of the decision support system for microgrid management has an advantage over existing data visualization systems for energy microgrids operation. It is that the use of the subsystem will provide comprehensive information support for decision-making in the intelligent energy microgrids management with renewable energy sources.

Acknowledgements

The studies were carried out within the frame of the project №0121U109558 «Intelligent Information-analytical Technologies and Means of Presentation, Assessment, and Management of the Country's Energy Infrastructure» at Sumy State University.

References

1. Vira Shendryk, Olha Boiko, Yuliia Parfenenko, Sergii Shendryk, Sergii Tymchuk. Decision Making for Energy Management in Smart Grid. *Research Anthology on Smart Grid and Microgrid Development* : monorgaph. IGI Global, 2022, P.1742-1776. DOI: 10.4018/978-1-6684-3666-0.ch057.
2. Kholoud Maswadinorjihan Binti, Abdul Ghani, Suraya Binti Hamid. Systematic Literature Review of Smart Home Monitoring Technologies Based on IoT for the Elderly. *IEEE Access*, vol.8., 2020. DOI:10.1109/ACCESS.2020.2992727.
3. S. Karnouskos, P. G. da Silva, D. Ilić, Developing a web application for monitoring and management of Smart Grid neighborhoods: Proceedings of the 11th IEEE International Conference on Industrial Informatics (INDIN), 2013. P. 408-413. DOI: 10.1109/INDIN.2013.6622919.
4. IoT Power Consumption Control and Monitoring, 2016. *Digiteum*. URL: https://www.digiteum.com/portfolio/electricity-consumption-monitoring-remote-control/ (accessed May 27, 2022).
5. Monitoring of Infrastructure and Services, 2018. *Solution Catalog of ProNet company*. URL: https://pronet.ua/en/monitoring-of-infrastructure-and-services/ (accessed May 27, 2022).
6. K. Dinesh, Lakshmi Priya, Preethi. T, Sandhya. M., Sangeetha. P. IoT Based Solar Panel Tracking System with Weather Monitoring System. Recent Trends in Intensive Computing. 2021. P. 795-799. DOI:10.3233/APC210282
7. Microgrid Case study: Schneider Electric helps make an ambitious Microgrid campus project. *Schneider Electric*. 2021. URL: https://www.se.com/ww/en/download/document/998-21080285/ (accessed May 27, 2022).
8. I. G. Pérez, A. J. Calderón Godoy. Monitoring interfaces for photovoltaic systems and dc microgrids: brief survey and application case. *XLII JORNADAS DE AUTOMÁTICA : LIBRO DE ACTAS*, 2021. P. 183–189. DOI:10.17979/spudc.9788497498043.183

9. Energy Management Systems for Optimal Operation of Electrical Micro/Nanogrids / ed. by Maria Carmela Di Piazza. MDPI, 2021, 137 p.
10. ETAP Microgrid Management, 2021. *The official webpage of Operation Technology, Inc.* URL: https://etap.com/solutions/microgrid (accessed May 27, 2022).
11. Monash Microgrid: data visualisation and platform, load and generation management, operating systems. *Monash Energy Institute*. 2020. URL: https://www.monash.edu/energy-institute/news-events/past-events/events/2020/monash-microgrid-data-visualisation-and-platform,-load-and-generation-management,-operating-systems (accessed May 27, 2022).
12. A. Maashri, N. Tarhuni, A. Elhaffar. A Real-Time Monitoring Platform for Distributed Energy Resources in a Microgrid. *Electronics*, vol. 10(15), 2021, P. 1-21. DOI:10.3390/electronics10151803.
13. Manson S. Microgrid Systems: Design, Control Functions, Modeling, and Field Experience. *Schweitzer Engineering Laboratories, Inc.* 2018. URL: https://cms-cdn.selinc.com/assets/Literature/Publications/Technical%20Papers/6701_MicrogridSystems_SM_20180726_Web.pdf?v=20190325-145907(accessed May 27, 2022).
14. Li Y. A Data Warehouse Architecture supporting Energy Management of Intelligent Electricity System : Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), P.1-4. DOI:10.2991/iccsee.2013.177.

https://doi.org/10.31891/csit-2022-2-9
UDC 519.7:007.52; 519.711.3

Iryna VECHIRSKA, Oleg KOBYLIN, Stepan PROKOPIEV, Anna VECHIRSKA
Kharkiv National University of Radio Electronics
Maksym KUCHERENKO
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

BUILDING A LOGICAL NETWORK FOR SOLVING THE PROBLEM OF CAR RENTAL BY MEANS ALGEBRA OF FINITE PREDICATES

The article is devoted to the research of the tools of algebra of finite predicates for the system analysis and formalization of the task of automating car rental according to the selected parameters. In particular, the process of optimal car selection depends on the following parameters: car class, brand, availability of driver, type of trip and its duration, fuel type, tariff type, as well as season and weather conditions. Each of these criteria has its own area of definition, where you need to take into account all the relationships and influences between the values of the entered variables (criteria). The aim of the work is to increase the speed of data processing in the problem of car rental by dividing the input multi-place ratio into a binary composition. The technique is based on the means and methods of algebra of finite predicates. Introduction of the predicate of object recognition in the specified subject area allowed to formally describe data of any type, and the applied method of construction of logical networks provides increase in speed of information processing due to parallelization of processing processes. Thus, a complex multi-place relation was divided into a composition of binary relations described in the language of predicate algebra, taking into account the detailed system analysis of the subject area. A scientific novelty is the constructed mathematical model of the car rental problem, which is represented by a predicate that depends on thirteen variables. This predicate is characterized by a system of twelve binary relations, which are represented in the article by dual graphs and formulas of the corresponding predicates. The model predicate is a composition of all constructed binary predicates. The practical significance is due to the logical network built on the basis of a mathematical model, which allows from the relationship "many to many" to move to the relationship "to each other" and parallelize the process of information processing. The result is a logical network of car rental problems, which works iteratively until it receives stable results in two consecutive steps and allows you to solve problems of analysis and synthesis for car rental according to selected parameters.

Keywords: algebra of finite predicates, predicate of object recognition, logical network, mathematical model, relation, car rental, subject area, criterion

Ірина ВЕЧІРСЬКА, Олег КОБИЛІН, Степан ПРОКОП'ЄВ, Анна ВЕЧІРСЬКА
Харківський національний університет радіоелектроніки, Харків, Україна
Максим КУЧЕРЕНКО
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

ПОБУДОВА ЛОГІЧНОЇ МЕРЕЖІ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧІ ОРЕНДИ АВТОМОБІЛІВ ЗАСОБАМИ АЛГЕБРИ СКІНЧЕННИХ ПРЕДИКАТІВ

Статтю присвячено дослідженню інструментарія алгебри скінченних предикатів для проведення системного аналізу та формалізації задачі автоматизації оренди автомобілів за обраними параметрами. Зокрема, процес оптимального вибору автомобіля залежить від наступних параметрів: класу автомобіля, його марки, наявності водія, типу подорожі та її тривалості, типу палива, типу тарифу, а також сезону та погодних умов. Кожен із зазначених критеріїв має свою область визначення, де потрібно враховувати усі взаємозв'язки та впливи саме між значеннями введених змінних (критеріїв). Метою роботи є підвищення швидкодії обробки данних в задачі оренди автомобілів за рахунок розбиття вхідного багатомісцевого відношення на композицію бінарних. Методика грунтується на засобах та методах алгебри скінченних предикатів. Введення предикату впізнавання предметів на вказаній предметній області дозволило формально описати дані будь-якого типу, а застосований метод побудови логічних мереж забезпечує підвищення швидкості обробки інформації за рахунок розпаралелювання процесів обробки. Таким чином, складне багатомісцеве відношення було розбито на композицію бінарних відношень, що описуються мовою алгебри предикатів з урахуванням деталізованого системного аналізу предметної області. Науковою новизною являється побудована математична модель задачі оренди автомобілів, яку подано предикатом, що залежить від тринадцяти змінних. Цей предикат характеризується системою дванадцяти бінарних відношень, які в статті представлено дводольними графами та формулами відповідних предикатів. Предикат моделі є композицією усіх побудованих бінарних предикатів. Практична значимість обумовлюється побудованою на основі математичної моделі логічною мережею, що дозволяє від відношення «багато до багатьох» перейти до відношень «один до одного» та розпаралелити процес обробки інформації. Результатом роботи є побудована логічна мережа задачі оренди автомобілів, що працює ітераційно до тих пір, поки не отримає сталі результати на двох кроках підряд та дозволяє розв'язувати задачі аналізу та синтезу для оренди автомобілів за вибраними параметрами.

Ключові слова: алгебра скінченних предикатів, предикат упізнавання предметів, логічна мережа, математична модель, відношення, оренда автомобілів, предметна область, критерій.

Introduction

The information processing and transmission, especially in large volumes, accumulates and intersects, which in the wrong organization leads to the loss of some data or to finding the wrong result, the system can also simply be overloaded. Relationships within the system are quite complex [1]. If we consider all the elements of the system to be equivalent and consider their interactions, the structure of the object under study will be too complex for automation and further improvement, as well as for the end user. In addition, the information in the system is

often presented in different forms. Therefore, it is important to have the tools to formally describe such data, to describe complex relationships and to build an appropriate mathematical model.

Related works

Nowadays, there are many programs for rent and lease accounting [2-4], designed to automate paperwork, staff work, direct accounting and optimization of various equipment, vehicles, real estate and more. As a rule, they allow you to set the price, fill in the nomenclature, attach a photo, barcode, specify the duration of the lease, even different markups for certain categories of customers. However, little attention is paid (or not taken into account at all) to the fact that the data themselves can be complex, and no direct interdependence between different data is studied, which leads to duplication of information and reduce the speed of its processing.

It is very important to analyze the subject area in detail, and only then build a mathematical model. And to analyze the system, it is advisable to formalize it, identify variables, values of variables and identify the essential relationships.

The language of algebra of finite predicates allows not only to formally describe the process of data processing taking into account the detailed system analysis of the subject area, but also allows to build an economical model in the form of a logical network by decomposing the input many-placed relationship [5,6]. Relationships are described by predicates of algebra of finite predicates, and the introduced predicate of object recognition allows to describe data of any nature [7-9].

Purpose

The problem of automating car rental is solved by building an appropriate logical network by describing the subject area by means of algebra of finite predicates, decomposition of the original multi-place predicate into a binary composition and introduction of additional internal variables characterizing the subject area. The aim of the work is to increase the speed of data processing in the problem of car rental in the form of a logical network by dividing the input multi-place relation into a composition of binary means of algebra of finite predicates.

Proposed technique and Results

The subject area "Car Rental Company" was taken to perform this task. The task consisted of building a logical network using predicate algebra and building a mathematical model.

Regarding the formal description and selection of criteria, three criteria were selected, namely: x_1– car class, x_2 – car name, x_3 – presence of a driver. It should be noted that the attribute of the presence of the driver actually determines the type of service provided (taxi or car sharing).

Regarding the allowable values of the first attribute, it was decided to use six main classes of cars. It should be noted that this classification is generally accepted, but may differ due to certain conditions, for example, if all cars of a certain class are not in sufficient demand. This phenomenon may be caused by insufficient earnings of citizens of the settlement.

x_1^1 – Economy, x_1^2 – Comfort, x_1^3 – Comfort+, x_1^4 – Business, x_1^5 – Premium, x_1^6 – Elite

The company's fleet consists of the following cars:

x_2^1 – KIA Rio, x_2^2 – Chevrolet Lacetti, x_2^3 – Volkswagen Polo, x_2^4 – Hyundai Solaris, x_2^5 – Renault Logan, x_2^6 – Skoda Octavia, x_2^7 – Hyundai Elantra, x_2^8 – Toyota Camri, x_2^9 – Kia Optima, x_2^{10} – Hyundai Sonata, x_2^{11} – Mercedes-Benz E-class, x_2^{12} – BMW 5, x_2^{13} – Audi A6, x_2^{14} – Mercedes-Benz S-class, x_2^{15} – BMW 7, x_2^{16} – Audi A8, x_2^{17} – Mercedes-Maybach S-class

The last attribute can take only two values, and it was decided to order the order of any car listed in the list of available, both with the driver and without him.

The predicate x_3^1 means a presence of a driver, the predicate x_3^2 means a driver is not present

The general view of the car rental company will be as follows (table 1), where the intersection of all attributes instead of a certain value will contain an expression obtained from the algebra of predicates and a pre-specified list of attributes and their valid values.

Table 1

Relationship between class, car name and driver availability

Class	Car Name	With driver	No driver
1	2	3	4
Economy	KIA Rio	$x_1^1x_2^1x_3^1=q_1$	$x_1^1x_2^1x_3^2=q_{18}$
	Chevrolet Lacetti	$x_1^1x_2^2x_3^1=q_2$	$x_1^1x_2^2x_3^2=q_{19}$
	Volkswagen Polo	$x_1^1x_2^3x_3^1=q_3$	$x_1^1x_2^3x_3^2=q_{20}$
	Hyundai Solaris	$x_1^1x_2^4x_3^1=q_4$	$x_1^1x_2^4x_3^2=q_{21}$
	Renault Logan	$x_1^1x_2^5x_3^1=q_5$	$x_1^1x_2^5x_3^2=q_{22}$
Comfort	Skoda Octavia	$x_1^2x_2^6x_3^1=q_6$	$x_1^2x_2^6x_3^2=q_{23}$
	Hyundai Elantra	$x_1^2x_2^7x_3^1=q_7$	$x_1^2x_2^7x_3^2=q_{24}$
Comfort+	Toyota Camri	$x_1^3x_2^8x_3^1=q_8$	$x_1^3x_2^8x_3^2=q_{25}$
	Kia Optima	$x_1^3x_2^9x_3^1=q_9$	$x_1^3x_2^9x_3^2=q_{26}$
	Hyundai Sonata	$x_1^3x_2^{10}x_3^1=q_{10}$	$x_1^3x_2^{10}x_3^2=q_{27}$

1	2	3	4
Business	Mercedes-Benz E-class	$x_1^4x_2^{11}x_3^1=q_{11}$	$x_1^4x_2^{11}x_3^2=q_{28}$
	BMW 5	$x_1^4x_2^{12}x_3^1=q_{12}$	$x_1^4x_2^{12}x_3^2=q_{29}$
	Audi A6	$x_1^4x_2^{13}x_3^1=q_{13}$	$x_1^4x_2^{13}x_3^2=q_{30}$
Premium	Mercedes-Benz S-class	$x_1^5x_2^{14}x_3^1=q_{14}$	$x_1^5x_2^{14}x_3^2=q_{31}$
	BMW 7	$x_1^5x_2^{15}x_3^1=q_{15}$	$x_1^5x_2^{15}x_3^2=q_{32}$
	Audi A8	$x_1^5x_2^{16}x_3^1=q_{16}$	$x_1^5x_2^{16}x_3^2=q_{33}$
Elite	Mercedes-Maybach S-class	$x_1^6x_2^{17}x_3^1=q_{17}$	$x_1^6x_2^{17}x_3^2=q_{34}$

Next you need to use the variable m – determination of the name of the car according to its class and the presence of the driver. To define m you must consider the values of all possible combinations of attributes x_1,x_2 and x_3. In this case the predicate of the car name will look like this:

$$m(x_1,x_2,x_3)= x_1^1x_2^1x_3^1 \vee x_1^1x_2^1x_3^2 \vee x_1^1x_2^2x_3^1 \vee x_1^1x_2^2x_3^2 \vee x_1^1x_2^3x_3^1 \vee x_1^1x_2^3x_3^2 \vee x_1^1x_2^4x_3^1 \vee x_1^1x_2^4x_3^2 \vee x_1^1x_2^5x_3^1 \vee x_1^1x_2^5x_3^2 \vee x_1^2x_2^6x_3^1 \vee x_1^2x_2^6x_3^2 \vee x_1^2x_2^7x_3^1 \vee x_1^2x_2^7x_3^2 \vee x_1^3x_2^8x_3^1 \vee x_1^3x_2^8x_3^2 \vee x_1^3x_2^9x_3^1 \vee x_1^3x_2^9x_3^2 \vee x_1^3x_2^{10}x_3^1 \vee x_1^3x_2^{10}x_3^2 \vee x_1^4x_2^{11}x_3^1 \vee x_1^4x_2^{11}x_3^2 \vee x_1^4x_2^{12}x_3^1 \vee x_1^4x_2^{12}x_3^2 \vee x_1^4x_2^{13}x_3^1 \vee x_1^4x_2^{13}x_3^2 \vee x_1^5x_2^{14}x_3^1 \vee x_1^5x_2^{14}x_3^2 \vee x_1^5x_2^{15}x_3^1 \vee x_1^5x_2^{15}x_3^2 \vee x_1^5x_2^{16}x_3^1 \vee x_1^5x_2^{16}x_3^2 \vee x_1^6x_2^{17}x_3^1 \vee x_1^6x_2^{17}x_3^2 .$$

Performing a disjunction operation on a variable:

$$\begin{gathered}
x_1^1x_2^1x_3^1 \vee x_1^1x_2^1x_3^2 = q_1 \vee q_{18} = m_1,\\
x_1^1x_2^2x_3^1 \vee x_1^1x_2^2x_3^2 = q_2 \vee q_{19} = m_2,\\
x_1^1x_2^3x_3^1 \vee x_1^1x_2^3x_3^2 = q_3 \vee q_{20} = m_3,\\
x_1^1x_2^4x_3^1 \vee x_1^1x_2^4x_3^2 = q_4 \vee q_{21} = m_4,\\
x_1^1x_2^5x_3^1 \vee x_1^1x_2^5x_3^2 = q_5 \vee q_{22} = m_5,\\
x_1^2x_2^6x_3^1 \vee x_1^2x_2^6x_3^2 = q_6 \vee q_{23} = m_6,\\
x_1^2x_2^7x_3^1 \vee x_1^2x_2^7x_3^2 = q_7 \vee q_{24} = m_7,\\
x_1^3x_2^8x_3^1 \vee x_1^3x_2^8x_3^2 = q_8 \vee q_{25} = m_8,\\
x_1^3x_2^9x_3^1 \vee x_1^3x_2^9x_3^2 = q_9 \vee q_{26} = m_9,\\
x_1^3x_2^{10}x_3^1 \vee x_1^3x_2^{10}x_3^2 = q_{10} \vee q_{27} = m_{10},\\
x_1^4x_2^{11}x_3^1 \vee x_1^4x_2^{11}x_3^2 = q_{11} \vee q_{28} = m_{11},\\
x_1^4x_2^{12}x_3^1 \vee x_1^4x_2^{12}x_3^2 = q_{12} \vee q_{29} = m_{12},\\
x_1^4x_2^{13}x_3^1 \vee x_1^4x_2^{13}x_3^2 = q_{13} \vee q_{30} = m_{13},\\
x_1^5x_2^{14}x_3^1 \vee x_1^5x_2^{14}x_3^2 = q_{14} \vee q_{31} = m_{14},\\
x_1^5x_2^{15}x_3^1 \vee x_1^5x_2^{15}x_3^2 = q_{15} \vee q_{32} = m_{15},\\
x_1^5x_2^{16}x_3^1 \vee x_1^5x_2^{16}x_3^2 = q_{16} \vee q_{33} = m_{16},\\
x_1^6x_2^{17}x_3^1 \vee x_1^6x_2^{17}x_3^2 = q_{17} \vee q_{34} = m_{17}.
\end{gathered}$$

The motive that prompted the operation of a disjunction operation on a variable is the desire to obtain an economical system of influences of definitions, in which each influence of the definition of the name of the car would correspond to one and only one name.

Binarization of a predicate that combines m with variables x_1, x_2 and x_3

$$\begin{gathered}
P_1(x_1, m) = x_1^1(m_1 \vee m_2 \vee m_3 \vee m_4 \vee m_5) \vee x_1^2(m_6 \vee m_7) \vee x_1^3(m_8 \vee m_9 \vee m_{10}) \vee x_1^4(m_{11} \vee m_{12} \vee m_{13}) \vee x_1^5(m_{14} \vee m_{15} \vee m_{16}) \vee x_1^6 m_{17},\\
P_2(x_2, m) = x_2^1m_1 \vee x_2^2m_2 \vee x_2^3m_3 \vee x_2^4m_4 \vee x_2^5m_5 \vee x_2^6m_6 \vee x_2^7m_7 \vee x_2^8m_8 \vee x_2^9m_9 \vee x_2^{10}m_{10} \vee x_2^{11}m_{11} \vee x_2^{12}m_{12} \vee x_2^{13}m_{13} \vee x_2^{14}m_{14} \vee x_2^{15}m_{15} \vee x_2^{16}m_{16} \vee x_2^{17}m_{17},\\
P_3(x_3, m) = (x_3^1 \vee x_3^2)(m_1 \vee m_2 \vee m_3 \vee m_4 \vee m_5 \vee m_6 \vee m_7 \vee m_8 \vee m_9 \vee m_{10} \vee m_{11} \vee m_{12} \vee m_{13} \vee m_{14} \vee m_{15} \vee m_{16} \vee m_{17}).
\end{gathered}$$

Let's represent the relations obtained by binarizing the predicate using graphs.

Graph of the relationship between m and x_1 variables shown on fig.1.

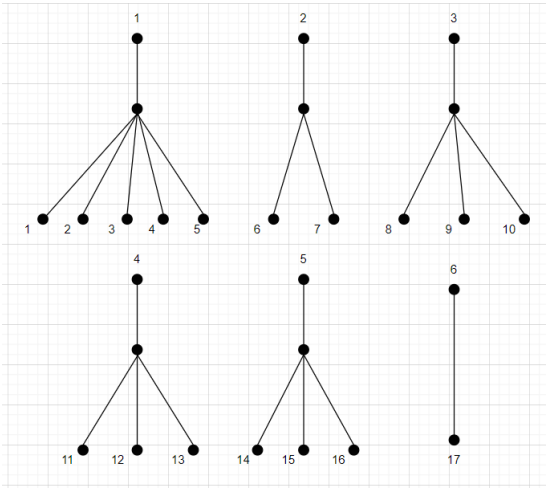

Fig.1. Graph of the relationship between m and x_1 variables

Graph of the relationship between m and x_2 variables shown on fig.2, where N belongs to the range from 1 to 17 inclusive.

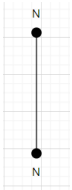

Fig.2. Graph of the relationship between m and x_2 variables

Graph of the relationship between m and x_3 variables shown on fig.3.

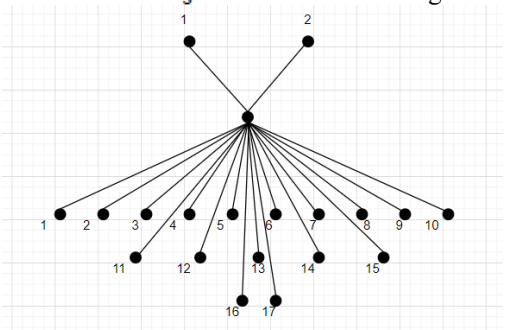

Fig.3. Graph of the relationship between m and x_3 variables

Let's classify further the cars available for rent according to the following criteria: y_1– rate type, y_2 – trip type.

According to the first attribute, rates are divided into: y_1^1 – morning, y_1^2 – day, y_1^3 – evening, y_1^4 – night, y_1^5 – mixed, for the second attribute, it can take the following values: y_2^1 – trip through the city, y_2^2 – trip between towns.

Table 2

Relationship between rate type and trip type

Rate type	Trip through the city	Trip between towns
morning	$y_1^1 y_2^1 = k_1$	$y_1^1 y_2^2 = k_6$
day	$y_1^2 y_2^1 = k_2$	$y_1^2 y_2^2 = k_7$
Evening	$y_1^3 y_2^1 = k_3$	$y_1^3 y_2^2 = k_8$
Night	$y_1^4 y_2^1 = k_4$	$y_1^4 y_2^2 = k_9$
Mixed	$y_1^5 y_2^1 = k_5$	$y_1^5 y_2^2 = k_{10}$

Next you need to use the variable n – determination the name of the car according to the type of available rate and trip. To determine n you must consider the values of all possible combinations of attributes $y_1$and y_2. In this case the predicate of the car name will look like this:

$$n\,(y_1, y_2) = y_1^1 y_2^1 \vee y_1^1 y_2^2 \vee y_1^2 y_2^1 \vee y_1^2 y_2^2 \vee y_1^3 y_2^1 \vee y_1^3 y_2^2 \vee y_1^4 y_2^1 \vee y_1^4 y_2^2 \vee y_1^5 y_2^1 \vee y_1^5 y_2^2.$$

Performing a disjunction operation on a variable:

$$y_1^1 y_2^1 \vee y_1^1 y_2^2 = k_1 \vee k_6 = n_1,$$
$$y_1^2 y_2^1 \vee y_1^2 y_2^2 = k_2 \vee k_7 = n_2,$$
$$y_1^3 y_2^1 \vee y_1^3 y_2^2 = k_3 \vee k_8 = n_3,$$
$$y_1^4 y_2^1 \vee y_1^4 y_2^2 = k_4 \vee k_9 = n_4,$$
$$y_1^5 y_2^1 \vee y_1^5 y_2^2 = k_5 \vee k_{10} = n_5.$$

Let's binarize the predicate that combines n with variables $y_1$and y_2:

$$P_4(y_1, n) = y_1^1 n_1 \vee y_1^2 n_2 \vee y_1^3 n_3 \vee y_1^4 n_4 \vee y_1^5 n_5,$$
$$P_5(y_2, n) = (y_2^1 \vee y_2^2)(n_1 \vee n_2 \vee n_3 \vee n_4 \vee n_5).$$

Let's represent the relations received at binarization of a predicate using graphs.

Graph of the relationship between n and y_1 variables shown on fig.4, where N belongs to the range from 1 to 5 inclusive.

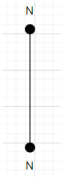

Fig.4. Graph of the relationship between n and y_1 variables

Graph of the relationship between n and y_2 variables shown on fig. 5.

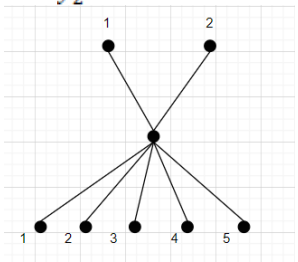

Fig.5. Graph of the relationship between n and $y_2$variables

In the case of long trips or long-term car rental, cars should be classified according to the following criteria: z_1– a fuel type that is accepted by car, z_2 – a duration of trip with a full tank (table 3).

According to the first attribute, its allowable values are the three most popular fuels, namely: z_1^1 – petrol, z_1^2 – electricity, z_1^3 – diesel. As for the second attribute, it was decided to divide it into the following values: z_2^1 – up to 2 hours, z_2^2 –up to 6 hours, z_2^3 – until a day, z_2^4 – more than a day.

Table 3

Relationship between class, car name and driver availability

Trip duration	Fuel type		
	Petrol	Electricity	Diesel
Up to 2 hours	$z_1^1 z_2^1 = l_1$	$z_1^2 z_2^1 = l_5$	–
Up to 6 hours	$z_1^1 z_2^2 = l_2$	$z_1^2 z_2^2 = l_6$	$z_1^3 z_2^2 = l_9$
Until a Day	$z_1^1 z_2^3 = l_3$	$z_1^2 z_2^3 = l_7$	$z_1^3 z_2^3 = l_{10}$
More than a day	$z_1^1 z_2^4 = l_4$	$z_1^2 z_2^4 = l_8$	–

Next, you need to use the variable b – determining the name of the car according to the type of fuel and duration of the trip. To determine b you must consider the values of all possible combinations of attributes z_1 and z_2. In this case the predicate of the car name will look like this:

$$b\,(z_1, z_2) = z_1^1 z_2^1 \vee z_1^1 z_2^2 \vee z_1^1 z_2^3 \vee z_1^1 z_2^4 \vee z_1^2 z_2^1 \vee z_1^2 z_2^2 \vee z_1^2 z_2^3 \vee z_1^2 z_2^4 \vee z_1^3 z_2^2 \vee z_1^3 z_2^3.$$

Performing a disjunction operation on a variable:

$$z_1^1 z_2^1 \vee z_1^2 z_2^1 = l_1 \vee l_5 = b_1,$$
$$z_1^1 z_2^2 \vee z_1^2 z_2^2 \vee z_1^3 z_2^2 = l_2 \vee l_6 \vee l_9 = b_2,$$
$$z_1^1 z_2^3 \vee z_1^2 z_2^3 \vee z_1^3 z_2^3 = l_3 \vee l_7 \vee l_{10} = b_3,$$

$$z_1^1 z_2^4 \vee z_1^2 z_2^4 = l_4 \vee l_8 = b_4.$$

Let's binarize the predicate that combines b with z_1та z_2 variables:

$$P_6(z_1, b) = (z_1^1 \vee z_1^2)(b_1 \vee b_2 \vee b_3 \vee b_4) \vee z_1^3(b_2 \vee b_3),$$
$$P_7(z_2, b) = z_2^1 b_1 \vee z_2^2 b_2 \vee z_2^3 b_3 \vee z_2^4 b_4 .$$

Let's represent the relations received at binarization of a predicate using graphs.
Graph of the relationship between b and z_1 variables shown on fig. 6.

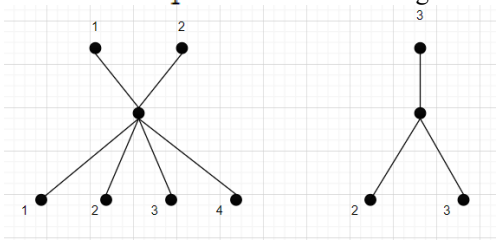

Fig.6. Graph of the relationship between n and y_2 variables

Graph of the relationship between b and z_2 variables shown on fig. 7, where N belongs to the range from 1 to 4 inclusive.

Fig.7. Graph of the relationship between b and z_2 variables

In order to increase the safety of our customers and eliminate unwanted traffic accidents while driving, it was necessary to classify all cars according to the following criteria: r_1– season, r_2 – weather conditions (table 4).

According to the first attribute, its allowable values are the three most popular fuels, namely: r_1^1 – summer, r_1^2 – autumn, r_1^3 – winter, r_1^4 – spring. Regarding the second attribute, it was decided to take into account the following weather conditions: r_2^1 – fog, r_2^2 – rain, r_2^3 – wind, r_2^4 – snow, r_2^5 – sleet.

Table 4

Relationship between season and weather conditions

Weather conditions	Season			
	Summer	Autumn	Winter	Spring
Fog	$r_1^1 r_2^1 = p_1$	$r_1^2 r_2^1 = p_4$	$r_1^3 r_2^1 = p_7$	$r_1^4 r_2^1 = p_{11}$
Rain	$r_1^1 r_2^2 = p_2$	$r_1^2 r_2^2 = p_5$	—	$r_1^4 r_2^2 = p_{12}$
Wind	$r_1^1 r_2^3 = p_3$	$r_1^2 r_2^3 = p_6$	$r_1^3 r_2^3 = p_8$	$r_1^4 r_2^3 = p_{13}$
Snow	—	—	$r_1^3 r_2^4 = p_9$	—
Sleet	—	—	$r_1^3 r_2^5 = p_{10}$	—

Next you need to use the variable c – determination of car name according to season and weather conditions. To determine c you must consider the values of all possible attribute combinations r_1 and r_2. In this case the predicate of car name will look like this:

$$c(r_1, r_2) = r_1^1 r_2^1 \vee r_1^1 r_2^2 \vee r_1^1 r_2^3 \vee r_1^2 r_2^1 \vee r_1^2 r_2^2 \vee r_1^2 r_2^3 \vee r_1^3 r_2^1 \vee r_1^3 r_2^3 \vee r_1^3 r_2^4 \vee r_1^3 r_2^5 \vee$$
$$r_1^4 r_2^1 \vee r_1^4 r_2^2 \vee r_1^4 r_2^3.$$

Performing a disjunction operation on a variable:

$$r_1^1 r_2^1 \vee r_1^2 r_2^1 \vee r_1^3 r_2^1 \vee r_1^4 r_2^1 = p_1 \vee p_4 \vee p_7 \vee p_{11} = c_1,$$
$$r_1^1 r_2^2 \vee r_1^2 r_2^2 \vee r_1^4 r_2^2 = p_2 \vee p_5 \vee p_{12} = c_2,$$
$$r_1^1 r_2^3 \vee r_1^2 r_2^3 \vee r_1^3 r_2^3 \vee r_1^4 r_2^3 = p_3 \vee p_6 \vee p_8 \vee p_{13} = c_3,$$
$$r_1^3 r_2^4 = p_9 = c_4,$$
$$r_1^3 r_2^5 = p_{10} = c_5.$$

Let's binarize the connecting predicate c with $r_1$and r_2 variables:

$$P_8(r_1, c) = (r_1^1 \vee r_1^2 \vee r_1^4)(c_1 \vee c_2 \vee c_3) \vee r_1^3(c_1 \vee c_3 \vee c_4 \vee c_5),$$
$$P_9(r_2, c) = r_2^1 c_1 \vee r_2^2 c_2 \vee r_2^3 c_3 \vee r_2^4 c_4 \vee r_2^5 c_5.$$

Let's represent the relations received at binarization of a predicate using graphs.
Graph of the relationship between c and r_1 variables shown on fig. 8.

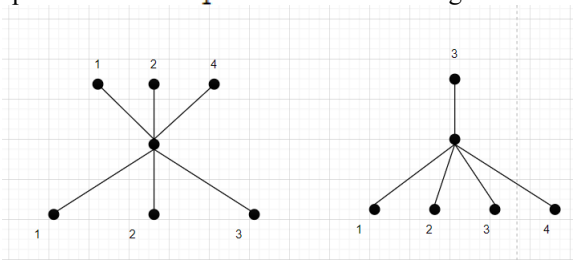

Fig.8. Graph of the relationship between c and r_1 variables

Graph of the relationship between c and r_2 variables shown on fig. 9, where N belongs a range from 1 to 5 exclusive.

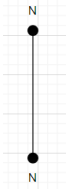

Fig.9. Graph of the relationship between c and r_2 variables

Let's build a mathematical model for determining the car in relation to its characteristics by class and car name, as well as rate.
A paradigmatic relationship table between m and n displayed in Table 5.

Table 5

Relationship between car class, name and rate

Class	Name	Rate				
		Morning	Day	Evening	Night	Mixed
Economy	KIA Rio	$x_1^1x_2^1y_1^1$	$x_1^1x_2^1y_1^2$	$x_1^1x_2^1y_1^3$	$x_1^1x_2^1y_1^4$	$x_1^1x_2^1y_1^5$
	Chevrolet Lacetti	$x_1^1x_2^2y_1^1$	$x_1^1x_2^2y_1^2$	$x_1^1x_2^2y_1^3$	$x_1^1x_2^2y_1^4$	$x_1^1x_2^2y_1^5$
	Volkswagen Polo	$x_1^1x_2^3y_1^1$	$x_1^1x_2^3y_1^2$	$x_1^1x_2^3y_1^3$	$x_1^1x_2^3y_1^4$	$x_1^1x_2^3y_1^5$
	Hyundai Solaris	$x_1^1x_2^4y_1^1$	$x_1^1x_2^4y_1^2$	$x_1^1x_2^4y_1^3$	$x_1^1x_2^4y_1^4$	$x_1^1x_2^4y_1^5$
	Renault Logan	$x_1^1x_2^5y_1^1$	$x_1^1x_2^5y_1^2$	$x_1^1x_2^5y_1^3$	$x_1^1x_2^5y_1^4$	$x_1^1x_2^5y_1^5$
Comfort	Skoda Octavia	$x_1^2x_2^6y_1^1$	$x_1^2x_2^6y_1^2$	$x_1^2x_2^6y_1^3$	$x_1^2x_2^6y_1^4$	$x_1^2x_2^6y_1^5$
	Hyundai Elantra	$x_1^2x_2^7y_1^1$	$x_1^2x_2^7y_1^2$	$x_1^2x_2^7y_1^3$	$x_1^2x_2^7y_1^4$	$x_1^2x_2^7y_1^5$
Comfort+	Toyota Camri	$x_1^3x_2^8y_1^1$	$x_1^3x_2^8y_1^2$	$x_1^3x_2^8y_1^3$	$x_1^3x_2^8y_1^4$	-
	Kia Optima	$x_1^3x_2^9y_1^1$	$x_1^3x_2^9y_1^2$	$x_1^3x_2^9y_1^3$	-	$x_1^3x_2^9y_1^5$
	Hyundai Sonata	-	$x_1^3x_2^{10}y_1^2$	$x_1^3x_2^{10}y_1^3$	$x_1^3x_2^{10}y_1^4$	$x_1^3x_2^{10}y_1^5$
Business	Mercedes-Benz E-class	$x_1^4x_2^{11}y_1^1$	$x_1^4x_2^{11}y_1^2$	$x_1^4x_2^{11}y_1^3$	-	-
	BMW 5	$x_1^4x_2^{12}y_1^1$	$x_1^4x_2^{12}y_1^2$	$x_1^4x_2^{12}y_1^3$	$x_1^4x_2^{12}y_1^4$	$x_1^4x_2^{12}y_1^5$
	Audi A6	$x_1^4x_2^{13}y_1^1$	$x_1^4x_2^{13}y_1^2$	$x_1^4x_2^{13}y$	-	-
Premium	Mercedes-Benz S-class	-	$x_1^5x_2^{14}y_1^2$	$x_1^5x_2^{14}y_1^2$	-	$x_1^5x_2^{14}y_1^2$
	BMW 7	$x_1^5x_2^{15}y_1^1$	$x_1^5x_2^{15}y_1^2$	$x_1^5x_2^{15}y_1^3$	-	-
	Audi A8	$x_1^5x_2^{16}y_1^1$	$x_1^5x_2^{16}y_1^2$	$x_1^5x_2^{16}y_1^3$	$x_1^5x_2^{16}y_1^4$	-
Elite	Mercedes-Maybach S-class	-	$x_1^6x_2^{17}y_1^2$	$x_1^6x_2^{17}y_1^3$	-	$x_1^6x_2^{17}y_1^4$

We binarize:

$$P_{10}(x_1,x_2,y_1) = y_1^1(x_1^1x_2^1 \vee x_1^1x_2^2 \vee x_1^1x_2^3 \vee x_1^1x_2^4 \vee x_1^1x_2^5 \vee x_1^2x_2^6 \vee x_1^2x_2^7 \vee x_1^3x_2^8 \vee x_1^3x_2^9 \vee x_1^4x_2^{11} \vee x_1^4x_2^{12} \vee x_1^4x_2^{13} \vee x_1^5x_2^{15} \vee x_1^5x_2^{16}) \vee y_1^2(x_1^1x_2^1 \vee x_1^1x_2^2 \vee x_1^1x_2^3 \vee x_1^1x_2^4 \vee x_1^1x_2^5 \vee x_1^2x_2^6 \vee x_1^2x_2^7 \vee x_1^3x_2^8 \vee x_1^3x_2^9 \vee x_1^3x_2^{10} \vee x_1^4x_2^{11} \vee x_1^4x_2^{12} \vee x_1^4x_2^{13} \vee x_1^5x_2^{14} \vee x_1^5x_2^{15} \vee x_1^5x_2^{16} \vee x_1^6x_2^{17}) \vee y_1^3(x_1^1x_2^1 \vee x_1^1x_2^2 \vee x_1^1x_2^3 \vee x_1^1x_2^4 \vee x_1^1x_2^5 \vee x_1^2x_2^6 \vee x_1^2x_2^7 \vee x_1^3x_2^8 \vee x_1^3x_2^9 \vee x_1^3x_2^{10} \vee x_1^4x_2^{11} \vee x_1^4x_2^{12} \vee x_1^4x_2^{13} \vee x_1^5x_2^{14} \vee x_1^5x_2^{15} \vee x_1^5x_2^{16} \vee x_1^6x_2^{17}) \vee y_1^4(x_1^1x_2^1 \vee x_1^1x_2^2 \vee x_1^1x_2^3 \vee x_1^1x_2^4 \vee x_1^1x_2^5 \vee x_1^2x_2^6 \vee x_1^2x_2^7 \vee x_1^3x_2^8 \vee x_1^3x_2^{10} \vee x_1^4x_2^{12} \vee x_1^5x_2^{16}) \vee y_1^5(x_1^1x_2^1 \vee x_1^1x_2^2 \vee x_1^1x_2^3 \vee x_1^1x_2^4 \vee x_1^1x_2^5 \vee x_1^2x_2^6 \vee x_1^2x_2^7 \vee x_1^3x_2^9 \vee x_1^3x_2^{10} \vee x_1^4x_2^{12} \vee x_1^5x_2^{14} \vee x_1^6x_2^{17}).$$

Let's build a mathematical model for determining the car in relation to its characteristics by trip type and car name, as well as fuel type.

Paradigmatic table of relationships between entered intermediate m and b variables shown Table 6

Table 6

Relationship between trip type, car name and fuel type

Name	With driver			Without driver		
	Petrol	Electricity	Diesel	Petrol	Electricity	Diesel
KIA Rio	$x_2^1x_3^1z_1^1$	—	—	$x_2^1x_3^2z_1^1$	—	—
Chevrolet Lacetti	$x_2^2x_3^1z_1^1$	—	—	$x_2^2x_3^2z_1^1$	—	—
Volkswagen Polo	—	—	$x_2^3x_3^1z_1^3$	—	—	$x_2^3x_3^2z_1^3$
Hyundai Solaris	—	—	$x_2^4x_3^1z_1^3$	—	—	$x_2^4x_3^2z_1^3$
Renault Logan	$x_2^5x_3^1z_1^1$	—	—	$x_2^5x_3^2z_1^1$	—	—
Skoda Octavia	$x_2^6x_3^1z_1^1$	—	—	$x_2^6x_3^2z_1^1$	—	—
Hyundai Elantra	—	$x_2^7x_3^1z_1^2$	—	—	$x_2^7x_3^2z_1^2$	—
Toyota Camri	—	—	$x_2^8x_3^1z_1^3$	—	—	$x_2^8x_3^2z_1^3$
Kia Optima	$x_2^9x_3^1z_1^1$	—	—	$x_2^9x_3^2z_1^1$	—	—
Hyundai Sonata	—	—	$x_2^{10}x_3^1z_1^3$	—	—	$x_2^{10}x_3^2z_1^3$
Mercedes-Benz E-class	$x_2^{11}x_3^1z_1^1$	—	—	$x_2^{11}x_3^2z_1^1$	—	—
BMW 5	$x_2^{12}x_3^1z_1^1$	—	—	$x_2^{12}x_3^2z_1^1$	—	—
Audi A6	—	$x_2^{13}x_3^1z_1^2$	—	—	$x_2^{13}x_3^2z_1^2$	—
Mercedes-Benz S-class	—	$x_2^{14}x_3^1z_1^2$	—	—	$x_2^{14}x_3^2z_1^2$	—
BMW 7	—	—	$x_2^{15}x_3^1z_1^3$	—	—	$x_2^{15}x_3^2z_1^3$
Audi A8	—	$x_2^{16}x_3^1z_1^2$	—	—	$x_2^{16}x_3^2z_1^2$	—
Mercedes-Maybach S-class	$x_2^{17}x_3^1z_1^1$	—	—	$x_2^{17}x_3^2z_1^1$	—	—

We binarize:

$$P_{11}(x_2,x_3,z_1) =$$
$$z_1^1(x_2^1x_3^1 \vee x_2^2x_3^1 \vee x_2^5x_3^1 \vee x_2^6x_3^1 \vee x_2^9x_3^1 \vee x_2^{11}x_3^1 \vee x_2^{12}x_3^1 \vee x_2^{17}x_3^1 \vee x_2^1x_3^2 \vee x_2^2x_3^2 \vee x_2^5x_3^2 \vee x_2^6x_3^2 \vee x_2^9x_3^2 \vee x_2^{11}x_3^2 \vee x_2^{12}x_3^2 \vee x_2^{17}x_3^2)$$

$$\vee\ z_1^2(x_2^7x_3^1 \vee x_2^{13}x_3^1 \vee x_2^{14}x_3^1 \vee x_2^{16}x_3^1 \vee x_2^7x_3^2 \vee x_2^{13}x_3^2 \vee x_2^{14}x_3^2 \vee x_2^{16}x_3^2) \vee z_1^3(x_2^3x_3^1 \vee x_2^4x_3^1 \vee x_2^8x_3^1 \vee x_2^{10}x_3^1 \vee x_2^{15}x_3^1 \vee x_2^3x_3^2 \vee x_2^4x_3^2 \vee\ x_2^8x_3^2 \vee x_2^{10}x_3^2 \vee x_2^{15}x_3^2)$$

.

Let's build a mathematical model for determining the car in relation to its characteristics by trip type and rate type, as well as weather conditions.

Paradigmatic table of relationships between entered intermediate n and c variables shown on Table 7.

Table 7

Relationship between travel type, rate type and weather conditions

Rate type	Weather conditions	In town	Between cities
1	2	3	4
Morning	Fog	$y_1^1y_2^1r_2^1$	$y_1^1y_2^2r_2^1$
	Rain	$y_1^1y_2^1r_2^2$	$y_1^1y_2^2r_2^2$
	Wind	$y_1^1y_2^1r_2^3$	$y_1^1y_2^2r_2^3$
	Snow	$y_1^1y_2^1r_2^4$	$y_1^1y_2^2r_2^4$
	Sleet	$y_1^1y_2^1r_2^5$	$y_1^1y_2^2r_2^5$
Day	Fog	$y_1^2y_2^1r_2^1$	$y_1^2y_2^2r_2^1$
	Rain	$y_1^2y_2^1r_2^2$	$y_1^2y_2^2r_2^2$
	Wind	$y_1^2y_2^1r_2^3$	$y_1^2y_2^2r_2^3$
	Snow	$y_1^2y_2^1r_2^4$	$y_1^2y_2^2r_2^4$
	Sleet	$y_1^2y_2^1r_2^5$	$y_1^2y_2^2r_2^5$
Evening	Fog	$y_1^3y_2^1r_2^1$	$y_1^3y_2^2r_2^1$
	Rain	$y_1^3y_2^1r_2^2$	$y_1^3y_2^2r_2^2$
	Wind	$y_1^3y_2^1r_2^3$	$y_1^3y_2^2r_2^3$
	Snow	$y_1^3y_2^1r_2^4$	$y_1^3y_2^2r_2^4$
	Sleet	$y_1^3y_2^1r_2^5$	$y_1^3y_2^2r_2^5$

1	2	3	4
Night	Fog	$y_1^4 y_2^1 r_2^1$	$y_1^4 y_2^2 r_2^1$
	Rain	$y_1^4 y_2^1 r_2^2$	$y_1^4 y_2^2 r_2^2$
	Wind	$y_1^4 y_2^1 r_2^3$	$y_1^4 y_2^2 r_2^3$
	Snow	$y_1^4 y_2^1 r_2^4$	$y_1^4 y_2^2 r_2^4$
	Sleet	$y_1^4 y_2^1 r_2^5$	$y_1^4 y_2^2 r_2^5$
Mixed	Fog	$y_1^5 y_2^1 r_2^1$	$y_1^5 y_2^2 r_2^1$
	Rain	$y_1^5 y_2^1 r_2^2$	$y_1^5 y_2^2 r_2^2$
	Wind	$y_1^5 y_2^1 r_2^3$	$y_1^5 y_2^2 r_2^3$
	Snow	$y_1^5 y_2^1 r_2^4$	$y_1^5 y_2^2 r_2^4$
	Sleet	—	—

We binarize:

$$P_{12}(y_1, y_2, r_2) =$$
$$(r_2^1 \vee r_2^2 \vee r_2^3 \vee r_2^4)(y_1^1 y_2^1 \vee y_1^2 y_2^1 \vee y_1^3 y_2^1 \vee y_1^4 y_2^1 \vee y_1^5 y_2^1 \vee y_1^1 y_2^2 \vee y_1^2 y_2^2 \vee y_1^3 y_2^2 \vee y_1^4 y_2^2 \vee y_1^5 y_2^2)$$
$$\vee r_2^5 (y_1^1 y_2^1 \vee y_1^2 y \vee y_1^3 y_2^1 \vee y_1^4 y_2^1 \vee y_1^1 y_2^2 \vee y_1^2 y_2^2 \vee y_1^3 y_2^2 \vee y_1^4 y_2^2).$$

Conclusions

Thus, a mathematical model was built in the course of the work. This model combines certain characteristics and with the help of this model it is possible to accurately determine the attribute – car name. The multi-place predicate, which reflected all the selection criteria (9 variables) by the user for car rental, is presented in the form of a composition of binary:

$$P(x_1, x_2, x_3, m, y_1, y_2, n_1, r_1, r_2, c, z_1, z_2, b) = P_1(x_1, m) \wedge P_2(x_2, m) \wedge P_3(x_3, m) \wedge P_4(y_1, n) \wedge$$
$$\wedge P_5(y_2, n) \wedge P_6(z_1, b) \wedge P_7(z_2, b) \wedge P_8(r_1, c) \wedge P_9(r_2, c) \wedge P_{10}(m, b) \wedge P_{11}(n, c) \wedge P_{12}(m, n).$$

We see that the constructed mathematical model in the form of a predicate P depends on 13 variables: 9 selected parameters and 4 intermediate variables. It is the introduction of these intermediate variables that made it possible to break the multi-place relation into binary ones. The complexity of the method of building a logical network is precisely the analysis of the system for optimal input of intermediate variables. To solve this problem, it was necessary to systematize all parameters in the form of tables, write the values obtained by the formulas of algebra of finite predicates through predicates of object recognition, then write the predicates corresponding to paradigmatic tables, and conduct, where possible, disjunctive gluing operation, binarization.

The results and data obtained during the work can be taken to the next display of the logical network (fig. 10):

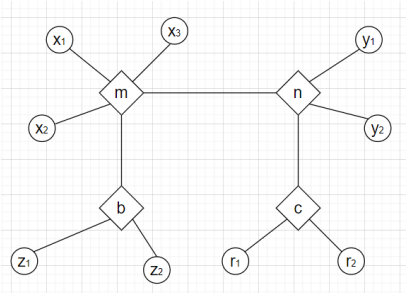

Fig.10. Logical car rental network

It should be noted that in addition to network performance, due to the parallelization of processes, it is important to be able to solve the network not only analysis problems, but also synthesis problems [10-11]. That is, the network can not only issue a car brand (or some other unknown parameter) based on the values entered in the nodes, but, working iteratively, allows you to set the value on all nodes and determine its truth, or, conversely, the known value of one of the parameters, find all relevant possible solutions.

References

1. Hlavcheva, Y.M., Kanishcheva, O.V. & Borysova, N.V. A Survey of Informetric Methods and Technologies. *Cybern Syst Anal 55*. 2019. P. 503–513.

2. CRM dlia avtoprokatu. URL: https://xn----7sbbaqhlkm9ah9aiq.net/ua/services-new/crm/crm-dlya-avtoprokata.html (дата звернення: 28.05.2022).

3. Transport: ekspluatatsiia, perevezennia, remonty. Debet-kredyt. Servisy dlia bukhhaltera.URL: https://services.dtkt.ua/tematic_roz/tematic_roz/103?_ga=2.252505609.146834866.1653898846-1112811398.1653898846 (дата звернення: 26.05.2022).

4. Carshare. URL: https://ekar.app/ae/en/carshare (дата звернення: 30.04.2022).

5. Shepelev G., Khairova N., Kochueva Z. Method "Mean – Risk" for Comparing Poly-Interval Objects in Intelligent Systems [Electronic resource]. *Computational linguistics and intelligent systems (COLINS 2019)* : proc. of the 3d Intern. Conf., April 18-19, 2019. Vol. I: Main Conference. Lviv, 2019. – P. 12-21. – URL: http://ceur-ws.org/Vol-2362/paper2.pdf (дата звернення: 15.12.2020)

6. Chetverykov G., Kobylin O., Lyashenko V., Deineko Z. Wavelets as a Tool for Data Mining Technology. *Information, Communication, Society (ICS-2019).* 16-18 May. Chynadiyovo, Ukraine, 2019. – P. 249-250. – URL: https://ics.skid-lp.info/2019/ics_2019.pdf#page=249 (дата звернення: 12.02.2022).

7. Tymoshuk R., Vilchynska Kh., Shyrokov V., Nadutenko M. Semantic interpretation of phraseological units in Ukrainian-Polish electronic phraseological dictionary. *Cognitive studies | Études cognitives*. Warsaw, 2015. Vol.15. P. 319–325.

8. Kharchenko V., Kondratenk, Y., Kacprzyk J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. *Studies in Systems, Decision and Control*. 2017. Vol. 74. P. 3-19.

9. Gorokhovatskyi, V.A., Vechirska, I.D. & Chetverikov, G.G. (2016). Method for building of logical data transform in the problem of establishing links between the objects in intellectual telecommunication systems. Telecommunications and engineering. 18, Vol.75. P. 1645-1655.

10. Vechirska I.D., Goncharov I.E., Khamitov T.M. Pobudova lohichnoi merezhi dlia diahnostyky ta upravlinnia nadzvychainymy sytuatsiiamy. Bionika intelektu: nauk.-tekhn. zhurnal. 2015. № 2 (85). S.41 – 51.

11. Shostak, I., Kapitan R., Volobuyeva L., Danova M.: Ontological approach to the construction of multi-agent systems for the maintenance supporting processes of production equipment. IEEE International Scientific and Practical Conference *Problems of Infocommunications. Science and Technology* (*PIC S&T-2018*). Kharkiv, 9–12 Oct 2018. IEEE. Pp 209–214.

https://doi.org/10.31891/csit-2022-2-10
UDC 004.023

Vasyl PRYIMAK, Svitlana PRYIMA, Olga HOLUBNYK
Ivan Franko National University of Lviv, Lviv, Ukraine

FUZZY TECHNOLOGIES IN THE MANAGEMENT OF SOCIO-ECONOMIC PROCESSES

The purpose of the article is to analyze the course of socio-economic processes in Ukraine over the past eight years to make recommendations for improving their management policy. The complexity of the task was that the development of the studied processes is characterized by not one but several primary factors, the simultaneous analysis of which is not a trivial task. Among these factors are not only quantitative but also qualitative indicators, which makes our task poorly structured.

In the analysis were used statistical and expert data that characterize socio-economic processes. This made it possible to consider the Russian military aggression against Ukraine and the situation with COVID-19 when making calculations. The use of qualitative information in the analysis makes it impossible to use methods of multidimensional statistics. To perform this task, it is proposed to use the theory of fuzzy logic and fuzzy sets. To describe the linguistic terms of fuzzy sets are used trapezoidal membership functions, the parameters of which are determined by experts. The constructed model of approximation of a nonlinear object with linguistic expressions allows finding the desired result faster and easier in comparison with the classical procedure of similar calculations.

Using the considered approach in the article the tendencies of the development of social and economic processes in Ukraine from 2013 to 2020 are analyzed. The result of the calculations is the value of the generalized indicator, the value of which characterizes the level of development of socio-economic processes in the year. The obtained calculations showed that the value of this indicator was the highest in 2013. Over the next two years, this figure decreased and then began to grow slowly. The increase lasted until 2019, and in 2020 the value decreased again. However, in none of the years, 2014-2020 did the figure reach 2013. In our opinion, the main reason for the decrease in the level of development of socio-economic processes in 2014 and 2015 was Russia's annexation of the Autonomous Republic of Crimea and the war in eastern Ukraine, and in 2020 - the effects of the COVID-19 pandemic.

Keywords: socio-economic processes, fuzzy set, linguistic variable, trapezoidal membership function, qualitative indicator, generalized indicator.

Василь ПРИЙМАК, Світлана ПРИЙМА, Ольга ГОЛУБНИК
Львівський національний університет

НЕЧІТКІ ТЕХНОЛОГІЇ В УПРАВЛІННІ СОЦІАЛЬНО-ЕКОНОМІЧНИМИ ПРОЦЕСАМИ

Метою статті було виконання аналізу перебігу соціально-економічних процесів в Україні протягом останніх восьми років для вироблення рекомендацій щодо удосконалення політики управління ними. Складність поставленого завдання полягала у тому, що розвиток досліджуваних процесів характеризується не одним, а декількома первинними чинниками, одночасний аналіз яких не є тривіальним завданням. Серед цих чинників є не тільки кількісні, а й якісні показники, що робить нашу задачу слабоструктурованою.

У роботі при виконанні аналізу було використано статистичні й експертні дані, які характеризують соціально-економічні процеси. Це дало змогу при проведенні розрахунків урахувати російську військову агресію проти України і ситуацію з COVID-19. Використання якісної інформації при проведенні аналізу унеможливлює застосування методів багатовимірної статистики. Для виконання цього завдання пропонується використовувати теорію нечіткої логіки і нечітких множин. Для опису лінгвістичних термів нечітких множин запропоновано скористатися трапецієподібними функціями належності, параметри яких визначені експертами. Побудована модель апроксимації нелінійного об'єкта з лінгвістичними висловлюваннями дозволяє швидше та легше знайти потрібний результат порівняно з класичною процедурою подібних обчислень.

Використовуючи розглянутий підхід у статті проаналізовано тенденції розвитку соціально-економічних процесів в Україні протягом 2013-2020 років. Результатом виконаних розрахунків є значення узагальненого показника, величина якого характеризує рівень розвитку соціально-економічних процесів у відповідному році. Отримані розрахунки показали, що величина цього показника була найбільшою у 2013 році. В наступних двох роках ця величина зменшувалась, а далі почала повільно зростати. Збільшення відбувалось аж до 2019 року, а в 2020 р. величина показника знову зменшилася. Проте в жодному з 2014-2020 років показник не досяг величини, яка була у 2013 році. На нашу думку, головною причиною зменшення величини рівня розвитку соціально-економічних процесів у 2014 і 2015 роках була анексія Росією Автономної республіки Крим і війна на сході України, а в 2020 р. – наслідки пандемії COVID-19.

Ключові слова: соціально-економічні процеси, нечітка множина, лінгвістична змінна, трапецієподібна функція належності, якісний показник, узагальнений показник.

Introduction

The development of the world economy is constantly accompanied by a contradiction between its capitalization and socialization. Earlier, the economy's capitalization won in this contradiction, the transformation of an arbitrary resource into capital. Currently, the priorities have changed. There is an awareness that the increase of capital is impossible without the economy's socialization and humanization. The goal of economic progress, its basic capital, is man. Instead of saving on people, they are looking for ways to invest in it, to develop it comprehensively. Investments in people pay off quickly and are effective in accelerating economic development. The formation and development of human capital are impossible without its involvement in production. This indicates the

interdependence and intertwining of social and economic processes. Therefore, they need to be studied together, which indicates the relevance of research to assess the development of socio-economic processes.

It is necessary to study the development of socio-economic processes in order to develop an effective socio-economic policy. It is not possible to make the right decision to manage a particular object or process without precise information about it. The adopted management actions will achieve the goal in the case of available information about the reaction of the system to previous management actions. This once again confirms the relevance of the procedure for analyzing the development of socio-economic processes.

Assessing the development of socio-economic processes is associated with certain difficulties. The first thing to emphasize is that the development of socio-economic processes is characterized not by one but by several primary factors, the simultaneous analysis of which is not a trivial task. The second thing that complicates the analysis - among the factors could be not only quantitative but also qualitative indicators. In the case of only quantitative primary factors, the procedure is simplified. Then you can use the methods and algorithms of multidimensional statistics to perform the analysis. Depending on the task, you can organize the studied processes according to the level of their development by time or region, find the best or worst of them or perform their clustering. A sufficient number of statistical methods and software products have been developed to date to implement these procedures.

The considered task is sharply complicated in case of a choice for the analysis, as primary, at least one qualitative indicator. Then the use of multidimensional statistics is impossible. We need to use fuzzy set theory, which solves poorly structured problems present in our case.

In this paper, we have analyzed the development of socio-economic processes in Ukraine in recent years using quantitative and qualitative indicators.

Related works

Several publications are devoted to the coverage of the results of scientific research on socio-economic processes and their management. In particular, the scientific works [1]-[3] consider the mechanisms of public management of these processes at the regional level. Works [4]-[5] investigated the socio-economic development management issues, respectively, of the municipality and the united territorial communities. But the effectiveness of management of socio-economic processes depends on accurate and reliable information about the state of these processes and their response to previous management actions. Therefore, it is necessary periodically to diagnose socio-economic processes and assess their condition. It is impossible to perform such an assessment with the help of any one primary indicator, as several primary partial indicators characterize these processes.

It is possible to estimate the dynamics of socio-economic processes and their intensity by analyzing all such primary partial indicators or to build on their basis one generalized integrated indicator and based on its value to draw certain conclusions. For example, the results of a study of most international rankings and indices, as well as some primary partial indicators of Ukraine's economy allowed the author of a scientific article [6] to develop certain recommendations, the implementation of which will ensure stable socio-economic development in the country. In [7], a comparative analysis of socio-economic processes in the regions of Poland used an algorithm for constructing Hellwig's taxonomic integral index. The extension of this method was used by the authors of scientific work [8] to study education in the countries of the European Union.

However, in the specified scientific works for estimation of social and economic processes only quantitative primary partial indicators and the method of multidimensional statistics are used. If high-quality primary information is taken into account, it isn't possible to use this method. In this case, it is advisable to use the methods of fuzzy set theory and fuzzy logic. This theory operates with so-called «soft» or otherwise «fuzzy» data typical of many economic problems and control systems in general. These tasks are associated with the uncertainty that cannot be accurately and unambiguously disclosed.

The fuzzy sets theory makes it possible to apply a linguistic description of weakly structured processes and formalize linguistic variables in decision making information systems [9].

The fuzzy sets theory was developed and used by foreign and Ukrainian scientists [10]-[12] and many others. The main concepts in this theory are the concept of fuzzy set and linguistic variable.

In more detail, a fuzzy set is defined as a set of pairs of the following type:

$$Y = \{(x, \mu_Y(x)),\ x \in X\}, \tag{1}$$

where Y is a fuzzy (blurred) set; X is the base scale or, in other words, the universal set; $\mu_Y(x)$ is a function of the membership of the set Y to the universal set X. This function can take values from the interval [0, 1] and be discrete or continuous. It determines the subjective measure of the expert's confidence that a given specific value of the base scale corresponds to a fuzzy set. It cannot be identified with probability, because the distribution function is unknown, there is no repetition of experiments [10].

Thus, taking into account the study of socio-economic processes of qualitative primary indicators that characterize these processes requires the use of fuzzy set theory elements. Thus, it is necessary to construct an integral indicator in a multidimensional fuzzy statement. Scientists have developed a method of constructing such

indicators, which generally characterize the studied processes. It is called a method of constructing a classical «model of approximation of a nonlinear object according to linguistic statements». The specified model displays an object with n inputs and one output:

$$y = f_y(x_1, x_2, \dots, x_N), \tag{2}$$

where y і x_i $\left(i = \overline{1, N}\right)$ – respectively, the output and input variables, which can be both quantitative and qualitative. It is assumed that for quantitative variables, the domains of their change are known, and for qualitative variables - sets of all their possible values.

This method of constructing a fuzzy model of the object of economic analysis has been developed in detail. Scientists widely use it to make recommendations for decision-making in poorly structured situations in various human activity areas. In short, its essence can be described as follows. The first stage of the considered technique, called the fuzzification of variables, is to translate the primary data into a fuzzy format. At this stage, determine the linguistic estimates of variables and necessary for their membership function's formalization. At the next stage, using the available expert information, a matrix of knowledge is built, and on its basis - a base of fuzzy knowledge. Here, if necessary, perform sampling of the continuous output y. Next, using this information, a system of logical equations is derived, which is used to perform calculations of fuzzy initial values of y. The last stage of this technique, called defuzzification of variables, is to convert the obtained initial values into a «clear» format. As a result, we get the desired value of the generalized indicator, which characterizes the studied socio-economic processes in a country or its region for a certain period of time, such as a year.

Despite the widespread use of this technique by scientists, its application is associated with significant difficulties. The fact is that the procedure of building a knowledge base with a large enough sample of primary data is quite time-consuming. This requires a significant amount of expert time. This complicates the use of this technique to assess the development of socio-economic processes.

To solve this problem in qualitative primary indicators, you can also use a simplified method of calculating the value of this integrated indicator [12]. This technique also consists of three steps: fuzzification of input data, processing of the received information, and its defuzzification. Simultaneously, it is not necessary to develop a knowledge base, and defuzzification is performed according to the approximate algorithm, which is convenient in calculations. Therefore, to assess the development of socio-economic processes in Ukraine, we will use this approach.

Research method

Suppose that to assess the development of socio-economic processes in Ukraine, we selected n primary indicators $X_1, X_2, \dots, X_n$, which are sufficient for this study. The values of these indicators for a certain period of time, such as a year, are denoted by $x_1, x_2, \dots, x_n$. They can be quantitative, determined based on statistical data, or qualitative, obtained from experts. The task is to determine the type of function.

$$Q = f_Q(X_1, X_2, \dots, X_n), \tag{3}$$

the value of which determines the level of development of the studied socio-economic processes in the country. The larger the value of the function (3), the higher this level. Moreover, for a better structure of our problem, we will look for this function in the normalized form ($Q \in [0,1]$)).

The set of states C of socio-economic processes is divided into several fuzzy subsets, which correspond to certain development levels. To represent the membership functions of these subsets, we use trapezoidal numbers:

$$\gamma = (b_1, b_2, b_3, b_4), \tag{4}$$

where numbers b_1, b_2, b_3, b_4 abscissas of the vertices of the trapezoid $OABC$ with coordinates $O(b_1, 0)$, $A(b_2, 1), B(b_3, 1), C(b_4, 0)$ in the Cartesian coordinate system (X, μ).

The number of these subsets can be arbitrary. In our research, we will use five subsets of C_1, C_2, C_3, C_4, C_5. These subsets intersect. Each of these subsets means, respectively, low, satisfactory, medium, good, and high level of socio-economic processes. Using the formula (4), the membership functions of these subsets are given in the form:

$$\begin{gathered}\theta_1(Q) = \gamma_1 = (0{,}0; 0{,}0; 0{,}15; 0{,}25);\\ \theta_2(Q) = \gamma_2 = (0{,}15; 0{,}25; 0{,}35; 0{,}45);\\ \theta_3(Q) = \gamma_3 = (0{,}35; 0{,}45; 0{,}55; 0{,}65);\\ \theta_4(Q) = \gamma_4 = (0{,}55; 0{,}65; 0{,}75; 0{,}85);\\ \theta_5(Q) = \gamma_5 = (0{,}75; 0{,}85; 1{,}0; 1{,}0).\end{gathered} \tag{5}$$

To perform calculations, you need to choose the primary factors, the number of which should not be large, but their total informativeness should be sufficient to reflect all aspects of the studied processes. At the same time, it is possible to consider their priority. Besides, they should be divided into two groups. The first group is stimulators, and the second - destimulators. The first of them include indicators, the increase in the value of which leads to an increase in the level of development of processes. The second of these groups includes indicators, the increase in the value of which leads to a decrease in the level of development of the studied processes.

Next, similarly to the initial indicator, it is necessary to determine the sets of states of each input indicator $X_i\ (i = \overline{1,n})$ with the involvement of experts. That is, for each of these indicators, it is necessary to specify the linguistic variable "The value of the indicator X_i" and its term set. Besides, you need to specify fuzzy subsets of the domain of this indicator $D(X_i)$, as well as their membership functions.

These term sets may have different, or the same number of terms, and these terms may be different. For simplicity, let's focus on random ones, where all term sets have five elements. Moreover, the terms have names: the value of the indicator $X_i\ (i = \overline{1,n})$"very bad", "bad", "average", "good" and "very good".

Denote by $D_{ij}\ (i = \overline{1,n}, j = \overline{1,5})$ fuzzy set, which corresponds to the j-th term $(j = \overline{1,5})$ of the indicator$X_i\ (i = \overline{1,n})$. We determine the membership functions of these sets with the help of experts in the form of trapezoidal numbers (4) so that they intersect for each $X_i\ (i = \overline{1,n})$ Denote them by $\lambda_j(x_i) = \lambda_{ij}\ (i = \overline{1,n}, j = \overline{1,5})$.

At the first stage of calculations, ie at the stage of fuzzification of indicators $X_i\ (i = \overline{1,n})$ for each of them, you need to specify the trapezoidal numbers of these membership functions, and then the algorithms for determining these functions themselves. It is possible to take into account the affiliation of each of these indicators to the classification group of stimulators or destimulators at this or the next stage. Let's do this at this stage.

Let us illustrate the actions of the first stage of the considered approach on the example of the indicator-destimulator X_1, which we will call "Force majeure". This is a qualitative indicator determined expertly for each period (year) of the study on a scale with a gradation from zero to ten points. The higher the number of points for this indicator, the worse the corresponding socio-economic process. The survey of experts showed that the membership functions $\lambda_j(j = \overline{1,5})$ of the corresponding fuzzy subsets $D_j\ (j = \overline{1,5})$ of this indicator have the form:

$$\lambda_1 = (8{,}0; 8{,}5; 10{,}0; 10{,}0);\ \lambda_2 = (6{,}0; 6{,}5; 8{,}0; 8{,}5);$$
$$\lambda_3 = (4{,}0; 4{,}5; 6{,}0; 6{,}5);\ \lambda_4 = (2{,}0; 2{,}5; 4{,}0; 4{,}5);$$
$$\lambda_5 = (0{,}0; 0{,}0; 2{,}0; 2{,}5); \qquad (6)$$

The algorithm for calculating the corresponding terms and values of the membership function of fuzzy sets of this indicator is given in the table. 1.

Table 1

Classification of values of the indicator-destimulator "Force majeure"

The name of the indicator	Value range	Classification group of the indicator's value	Degree of assessed confidence (membership function)
Force majeure circumstances (designation of the indicator) – X_1, of the value – K_1)	$8{,}5 \le K_1 \le 10$	"Very bad"	$\lambda_1 = 1$
	$8 < K_1 < 8{,}5$	"Very bad"	$\lambda_1 = 2 \cdot (K_1 - 8)$
	$8 < K_1 < 8{,}5$	"Bad"	$\lambda_2 = 1 - \lambda_1$
	$6{,}5 \le K_1 \le 8$	"Bad"	$\lambda_2 = 1$
	$6 < K_1 < 6{,}5$	"Bad"	$\lambda_2 = 2 \cdot (K_1 - 6)$
	$6 < K_1 < 6{,}5$	"Average"	$\lambda_3 = 1 - \lambda_2$
	$4{,}5 \le K_1 \le 6$	"Average"	$\lambda_3 = 1$
	$4 < K_1 < 4{,}5$	"Average"	$\lambda_3 = 2 \cdot (K_1 - 4)$
	$4 < K_1 < 4{,}5$	"Good"	$\lambda_4 = 1 - \lambda_3$
	$2{,}5 \le K_1 \le 4$	"Good"	$\lambda_4 = 1$
	$2 < K_1 < 2{,}5$	"Good"	$\lambda_4 = 2 \cdot (K_1 - 2)$
	$2 < K_1 < 2{,}5$	"Very good"	$\lambda_5 = 1 - \lambda_4$
	$0 \le K_1 \le 2$	"Very good"	$\lambda_5 = 1$

For all other input indicators $X_i\ (i = \overline{2,n})$ it is necessary to develop similar algorithms for determining the corresponding terms and values of membership functions $\lambda_{ij}\ (i = \overline{2,n}, j = \overline{1,5})$ on the basis of a survey of experts.) of their fuzzy sets $D_{ij}\ (i = \overline{2,n}, j = \overline{1,5})$. At the same time, immediately consider whether the indicator is a stimulant or a disincentive.

The second step in determining the desired result is to determine the fuzzy number Q. It can be calculated by the formula [12]:

$$Q = (q_1, q_2, q_3, q_4, q_5) = \sum_{j=1}^{5} Z_j \otimes \gamma_j \quad (7)$$

in which the values of $Z_j\left(j = \overline{1,5}\right)$ are calculated by the following formulas:

$$Z_j = \left(\sum_{i=1}^{N} p_i \cdot \lambda_{ij}\right) / \left(\sum_{i=1}^{N} p_i\right) \ \left(j = \overline{1,5}\right) \quad (8)$$

where the sign "⊗" means the operation of multiplying the real number Z_j by the fuzzy number y_j, which is given by formula (5), the value of p_i is the priority factor of the indicator X_i $(i = \overline{1,\ N})$.

At the last stage of our calculations, we defuzzification the obtained fuzzy number $\underline{Q}$ into a "clear" format. The belonging of the trapezoidal interval Q to one of the fuzzy subsets {C} of the development level of socio-economic processes can be determined using the formulas of section and union of fuzzy sets. The degree of belonging S of the level of development to one of the states C_j is determined using the area Δ of a figure by the formula [12]:

$$S = [\Delta(Q \cap C_j)] / [\Delta(Q \cup C_j)] \quad (9)$$

where Δ is defined as the corresponding area bounded by trapezoidal curves of membership functions.

Given the significant difficulties in the calculations by formula (9), according to [12] we will find the approximate value of immediately "clear" generalized indicator of the level of development of socio-economic processes by the formula:

$$Z = 0{,}075 \cdot Z_1 + 0.3 \cdot Z_2 + 0.5 \cdot Z_3 + 0.7 \cdot Z_4 + \ +0.925 \cdot Z_5, \quad (10)$$

where $Z_j\ \left(j = \overline{1,5}\right)$ are determined by formulas (8).

If necessary, for the value of this generalized exponent Z found by formula (10), it is possible to determine the corresponding terms and values of the membership function of fuzzy sets that correspond to them. To do this, use the algorithm given in the table. 2, which is constructed using formulas (5).

Table 2

Algorithm for finding terms and corresponding values of membership functions of the generalized indicator of the level of development of socio-economic processes

The name of the indicator	Value range	Level of development (classification group, term)	Degree of assessed confidence (membership function)
The development level of socio-economic processes	$0 \le Z \le 0{,}15$	"low"	$\gamma_1 = 1$
	$0{,}15 < Z < 0{,}25$	"low"	$\gamma\_1 = 10 \cdot (0{,}25 - Z)$
	$0{,}15 < Z < 0{,}25$	"satisfactory"	$\gamma_2 = 1 - \gamma_1$
	$0{,}25 \le Z \le 0{,}35$	"satisfactory"	$\gamma_2 = 1$
	$0{,}35 < Z < 0{,}45$	"satisfactory"	$\gamma_2 = 10 \cdot (0{,}45 - Z)$
	$0{,}35 < Z < 0{,}45$	"average"	$\gamma_3 = 1 - \gamma_2$
	$0{,}45 \le Z \le 0{,}55$	"average"	$\gamma_3 = 1$
	$0{,}55 < Z < 0{,}65$	"average"	$\gamma_3 = 10 \cdot (0{,}65 - Z)$
	$0{,}55 < Z < 0{,}65$	"good"	$\gamma_4 = 1 - \gamma_3$
	$0{,}65 \le Z \le 0{,}75$	"good"	$\gamma_4 = 1$
	$0{,}75 < Z < 0{,}85$	"good"	$\gamma_4 = 10 \cdot (0{,}85 - Z)$
	$0{,}75 < Z < 0{,}85$	"high"	$\gamma_5 = 1 - \gamma_4$
	$0{,}85 \le Z \le 1$	"high"	$\gamma_5 = 1$

Results of numerical calculations

We use the considered method of constructing a generalized indicator to assess the development of socio-economic processes in Ukraine during 2013-2020. As primary indicators, we take 8 (n=8) indicators: force majeure (X_1), gross domestic product (GDP) (X_2), the volume of sold innovative products to the total volume of sold industrial products (X_3), average wages (X_4), unemployment rate (X_5), consolidated budget expenditures on education (X_6), consolidated budget expenditures on health care (X_7), and consumer price index (X_8), All these primary indicators are given in annual terms. Moreover, X_1, X_5, X_8 are destimulators, and all others are stimulators. Experts on a ten-point scale determine the value of the first of these indicators. All other indicators are statistical.

GDP and consolidated budget expenditures on education and health care are presented in UAH billion. The third and fourth indicators are expressed in parts and UAH, respectively. The unemployment rate and consumer price index are given as a percentage. The values of all cost indicators are indicated in the prices of 2013. As a result, the following initial data were used for calculations (Table 3).

Table 3

The value of input indicators to determine the level of development of socio-economic processes in Ukraine for 2013-2020

Indiator	Year							
	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
X_1	4,5	9,8	8,7	7,8	7,5	6,9	7,1	9,2
X_2	1454,9	1351,8	1229,6	1264,2	1295,9	1339,8	1384,2	1327,3
X_3	3,3	2,5	1,4	1,05	0,7	0,8	1,3	1,9
X_4	3273,8	2994,0	2613,5	2751,7	3087,0	3338,3	3658,1	3669,8
X_5	7,7	9,7	9,5	9,7	9,9	9,1	8,6	9,9
X_6	101561	86376	70934	68661	77296	79071	83151	79836
X_7	61 568	61261	46842	54315	50330	48331	61222	55630
X_8	100,5	124,9	143,3	112,4	113,7	109,8	104,1	105

Based on the expert survey's processed results for each of the considered input indicators, the trapezoidal membership functions of the type (4) of the considered linguistic variables were determined, which are given in the table. 4. In the process of forming this table, the belonging of the primary indicators X_i $(i = \overline{1,\ 8})$ to stimulators or destimulators is immediately taken into account.

Table 4

Functions of belonging of input indicators

Indicator	Trapezoidal numbers for the values of the linguistic variable "The value of the indicator X_i"				
	"Very bad"	**"Bad"**	**"Average"**	**"Good"**	**"Very good"**
X_1	(8; 8,5; 10; 10)	(6; 6,5; 8; 8,5)	(4; 4,5; 6; 6,5)	(2; 2,5; 4; 4,5)	(0; 0; 2; 2,5)
X_2	(0; 0; 1200; 1250)	(1200; 1250; 1300; 1350)	(1300; 1350; 1400; 1450)	(1400; 1450; 1500; 1550)	(1500;1550; $+\infty$; $+\infty$)
X_3	(0; 0; 0,5; 1)	(0,5; 1; 1,5; 2)	(1,5; 2; 2,5; 3)	(2,5; 3; 3,5; 4)	(3,5; 4; 20; 20)
X_4	(0; 0; 2600; 2800)	(2600; 2800; 3000; 3200)	(3000; 3200; 3400; 3600)	(3400; 3600; 3800; 4000)	(3800;4000; $+\infty$; $+\infty$)
X_5	(9,5; 10; 100; 100)	(8; 8,5; 9,5; 10)	(6,5; 7; 8; 8,5)	(5; 5,5; 6,5; 7)	(0; 0; 5; 5,5)
X_6	(0; 0; 60; 65)	(60; 65; 75; 80)	(75; 80; 90; 95)	(90; 95; 105; 110)	(105; 110; $+\infty$; $+\infty$)
X_7	(0; 0; 42,5; 45)	(42,5; 45; 50; 52,5)	(50; 52,5; 57,5; 60)	(57,5; 60; 65; 67,5)	(65; 67,5; $+\infty$; $+\infty$)
X_8	(130; 135; 500; 500)	(120; 125; 130; 135)	(110; 115; 120; 125)	(100; 105; 110; 115)	(0; 0; 100; 105)

Based on the table. 4 for each of the primary indicators X_i $(i = \overline{2,8})$ were developed similar to the table. 1 for indicator X_1 algorithms for fuzzification of these indicators. That is, algorithms for determining the terms and membership functions of the corresponding fuzzy subsets were developed for each primary indicator. Using these algorithms based on those given in the table. 3 initial data for all primary indicators X_i $(i = \overline{1,8})$ and each of the studied years from 2013 to 2020 were calculated specific values of membership functions $\lambda_j(x_i) = \lambda_{ij}$ $\left(i = \overline{1,8}, j = \overline{1,5}\right)$ fuzzy subsets of D_{ij} $\left(i = \overline{1,8}, j = \overline{1,5}\right)$. Using the formula (8), the values of Z_j $\left(j = \overline{1,5}\right)$ are also determined for each of these years. The calculations were performed under the same priority of all primary indicators. The obtained results of these calculations showed that for 2020 these variables have the following values (see table 5).

Table 5

Values {λ} і z_j for the level of the generalized indicator of socio-economic processes in 2020

$\{\theta\}$	λ_{i1}	λ_{i2}	λ_{i3}	λ_{i4}	λ_{i5}
X_1	1	0	0	0	0
X_2	0	0,454	0,546	1	0
X_3	0	0,2	0,8	0	0
X_4	0	0	1	0	0
X_5	0,8	0,2	0	0	0
X_6	0	0,08	0,92	0	0
X_7	0	0	1	0	0
X_8	0	0	0	1	0
Z_j	0,225	0,11675	0,53325	0,125	0

Using formula (10) and the calculated values of $Z_j\ (j=\overline{1,5})$ from table. 5 we get the value of the generalized indicator Z for 2020,

$$Z = 0{,}075 \cdot 0{,}225 + 0.3 \cdot 0{,}11675 + 0.5 \cdot 0{,}53325 + 0.7 \cdot 0{,}125 + 0.925 \cdot 0 = 0{,}406$$

We use the calculated value of this integral exponent to find the number of the corresponding fuzzy set $C_j\ (j=\overline{1,5})$ and its membership function $\gamma_j(j=\overline{1,5})$. Using the algorithm given in table. 2 we obtain that $\gamma_2 = 0{,}44\ \gamma_3 = 0{,}56$, and $\gamma_j = 0\ \ (j=1,4,5)$. That is, with a high degree of compliance it can be argued that in 2020 the level of development of socio-economic processes in Ukraine was "average" and with a lower degree - satisfactory.

The considered values for all studied years are given in the table. 6.

Table 6

The complex indicator values are the level of development of socio-economic processes in Ukraine *Z* and the membership function $y_j(Z)$ of the set of its states in 2013-2020

Year	Z	$\gamma_j(Z)$				
		γ_1	γ_2	γ_3	γ_4	γ_5
2013	0,6503	0	0	0	1	0
2014	0,3361	0	1	0	0	0
2015	0,1956	0,544	0,456	0	0	0
2016	0,2918	0	1	0	0	0
2017	0,3089	0	1	0	0	0
2018	0,3884	0	0,616	0,384	0	0
2019	0,4517	0	0	1	0	0
2020	0,4060	0	0,440	0,560	0	0

This table shows that the value of the integrated indicator, the level of development of socio-economic processes in Ukraine, was the highest in 2013. However, in the next two years, this value decreased and then began to grow slowly. The increase lasted until 2019, and in 2020 the value of this indicator decreased again. In our opinion, the main reason for the decrease in this indicator in 2014 and 2015 was Russia's annexation of the Autonomous Republic of Crimea and the war in eastern Ukraine, and the 2020 - COVID-19 pandemic.

Conclusions

The study showed that to assess the development of socio-economic processes, it is advisable to use the theory of fuzzy sets. Moreover, the problem can be solved by a simplified method, which consists of constructing a generalized indicator of the level of development of socio-economic processes, measured in the order scale. The performed calculations show that the level of development of socio-economic processes in Ukraine was the highest in 2013. Over the next two years, this level decreased, and then gradually increased until 2019. In 2020, the value of this indicator decreased again.

References

1. Kachnyj O.S. Formation and implementation of effective mechanisms of public administration of socio-economic processes at the regional level. *Investments: practice and experience*. № 5. 2018. pp. 64-66.
2. Skyba M.V. Conceptual principles of public management of socio-economic development of the region / *Regional economy*, 2020. №3. pp. 35-45.
3. Tofanjuk O. V. State regulation of socio-economic development of the regions of Ukraine / O.V. Tofanjuk, A.S. Zinchuk // *Black Sea Economic Studies*, 2018. Vol. 33, pp. 159-165.
4. Kutsenko T., Pintelei, I. "Management of socio-economic development of the municipality: strategy, analysis and tools for efficiency evaluation (on the example of kalush city in ivano-frankivsk oblast)", *Efficient economy.* [Online], 2021. vol. 7, available at: http://www.economy.nayka.com.ua/?op=1&z=9049 (Accessed 21 Jun 2022). DOI: 10.32702/2307-2105-2021.7.72)
5. Korchynsjka O. O. Management of socio-economic development of united territorial communities / *Visnyk of the National University "Lviv Polytechnic*". Series Problems of Economics and Management. Vol. 7, № 4(2). 2019. pp. 81-88.
6. Ghordej O. D., Myrkun B. A. The current state of social and economic development of Ukraine / Business Inform № 10. 2019. pp. 151-157.
7. Malina A. Analiza przestrzennego zróżnicowania poziomu rozwoju społeczno-gospodarczego województw Polski w latach 20052017 / *Social Inequalities and Economic Growth*, №61 (1/2020), pp. 138-155.
8. Roszkowska E. Measuring Sustainable Development Using an Extended Hellwig Method: A Case Study of Education / Ewa Roszkowska, Marzena Filipowicz-Chomko. *Social Indicators Research.* January 2021. 153(3):1-24. URL: https://www.researchgate.net/publication/344348425
9. Olenych I., Nechypor Yu., Olenych Yu., Gukaliuk A. Optimization of Transport Routes by Fuzzy Logic Methods, International Scientific and Practical Conference "*Electronics and Information Technologies*", 2018. Issue 10, Pages: A-147-A-150.
10. Siavavko M.S., Rybytska O.M. Mathematical modeling under uncertainty. Lviv: SPF "Ukrainian Technologies". 2000. 319 p.
11. Hrytsenko K.H. Using the theory of fuzzy sets for assessing the level. Visnyk bank security from cyber frauds) / *Priazovsky Economic Herald* . Vol. 1(12) 2019. pp. 214-219.
12. Pryimak V. I., Holubnyk O. R., Ucieklak-Jeż P., Kubicka J., Urbańska K., Babczuk A. Fuzzy Technologies Modeling the Level of Welfare of the Population in the System of Effective Management. *EuropeanResearch Studies Journal.* Piraeus, Greece. 2021. Volume XXIV, Issue 3. P. 749-762.

https://doi.org/10.31891/csit-2022-2-11
UDC 004.3

Yelyzaveta HNATCHUK, Vitalii BASHUK, Denys KVASNITSKYI
Khmelnytskyi National University

RESEARCH OF METHODS AND MEANS OF ENSURING THE RELIABILITY OF A SPECIALIZED COMPUTER VOICE VEHICLE CONTROL SYSTEM

The methods and means of protection of reliability in modern specialized computer systems of voice control of the car are investigated in the work. The evaluation of the characteristics and properties of the system is carried out. The basic principles of work and various possibilities of constructions of voice control of the car are considered. The methods and means of detecting dangers are shown, and the shortcomings and vulnerabilities of the car systems "Android Auto" and "Apple CarPlay" regarding the impact of malicious software based on modern methods of cyberattacks are identified. Preparation for cyber attack by ultrasonic and light commands on car voice control systems is shown. Methods and means for increasing the degree of protection of the voice authentication system of specialized computer systems "Android Auto" and "Apple CarPlay" are proposed.

To solve this problem, a hardware and software product of an additional biometric automotive user authentication system was created and developed. The system was created to ensure the reliability of cyber attacks on the voice control system.

Experimental studies confirm the effectiveness of the biometric authentication system as a proposed solution to provide an additional method of protection.

Keywords: voice control systems, method, protection, ultrasonic attacks

Єлизавета ГНАТЧУК, Віталій БАШУК, Денис КВАСНІЦЬКИЙ
Хмельницький національний університет

ДОСЛІДЖЕННЯ МЕТОДІВ ТА ЗАСОБІВ ЗАБЕЗПЕЧЕННЯ НАДІЙНОСТІ СПЕЦІАЛІЗОВАНОЇ КОМПЮТЕРНОЇ СИСТЕМИ ГОЛОСОВОГО КЕРУВАННЯ АВТОМОБІЛЕМ

В роботі досліджено методи та засоби захисту забезпечення надійності в сучасних спеціалізованих комп'ютерних системах голосового керування автомобілем. Проведена оцінка характеристик та властивостей системи. Розглянуто основні принципи роботи та різні можливості конструкцій голосового керування автомобілем. Показано методи і засоби виявлення небезпек, і визначено недоліки та вразливості автомобільних систем «Android Auto» та «Apple CarPlay» щодо впливу зловмисного програмного забезпечення на основі сучасних способів кібератак. Показано підготовку для здійснення кібератаки ультразвуковими та світловими командами на системи голосового керування автомобілем. Запропоновано методи та засоби для підвищення ступеня захисту системи голосової аутинтефікації спеціалізованих комп'ютерних систем «Android Auto» та «Apple CarPlay».

Для вирішення даної проблеми було створено та розроблено апаратно-програмний продукт додаткової біометричної автомобільної системи аутентифікації користувача. Система була створена для забезпечення надійності від кібератак на систему голосового керування.

Проведені експериментальні дослідження підтверджують ефективність системи біометричної аутентифікації як запропонованого рішення щодо забезпечення додаткового методу захисту.

Ключові слова: системи голосового керування, метод, захист, ультразвукові атаки

Introduction

Today, voice-controlled systems are widely used in various industries, including automotive engineering. They are increasingly used by people of all ages, because they are quite easy to operate, and most importantly effective. This is due to the fact that such systems help the user to solve various types of problems due to the wide range of functionality. Despite such widespread popularity among users, there are a number of problems when using such systems. Modern algorithms for recognizing voice commands are not yet perfect and do not always clearly understand a given user command, so there is a problem that such systems can be subjected to different types of cyberattacks. But, thanks to the development of neural network and cloud computing technologies, and the use of modern hardware and software and methods to ensure reliability, this problem can be minimized.

Therefore, the study of methods and means to ensure the reliability of a specialized computer system for voice control of the car is an urgent task.

Subject area analysis and relevant decisions

Problems solved in computer information systems have a number of characteristic features that affect the technology of automated data processing.

The computer system has the ability to integrate with other engineering technologies, expand capabilities and create a unified management environment, using the diversity and unification of computer equipment [1].

Dedicated computer voice control system helps you with voice commands to control functions such as navigating the route in the navigator, using climate control and its functionality, controlling the multimedia system, it also has the ability to interact with the user. With the help of voice assistants, the system can respond to voice commands and display various information on the screen of the driver's multimedia device.

In modern cars, voice control is performed by uttering the appropriate commands, which by undergoing certain transformations are converted into control signals for the respective systems. Today, you can use voice control to control the following systems in the car (Table 1).

Table 1

Voice control systems

System type	Execution of functions
Climate control	With the help of the climate control system, the user can change the temperature, turn on the seat heating, change the fan speed and more.
Multimedia	Provides the ability to receive, transmit, video and audio information.
Navigation	Perform voice control of the car navigation system
On-board computer	Determining the parameters of the car

In the voice control system, one of the main functions is voice recognition, which allows you to control the mobile phone connected to it, use the various features of the multimedia system, use the radio, navigation system and much more.

Entering voice commands greatly reduces your time and control, which in turn helps you focus on the road and driving. It is also possible to use voice commands to interact with navigation systems, ie paving or changing routes, etc. Voice control systems support a variety of languages, including the unpopular ones.

The process of voice recognition in specialized computer voice control systems (Figure 1) takes place in several stages. At each stage, a number of different methods are used to process the material signal. The process of voice recognition can be divided into three stages:

- receiving a voice signal and processing commands;
- recognition of phonemes and words;
- understanding of the voice command.

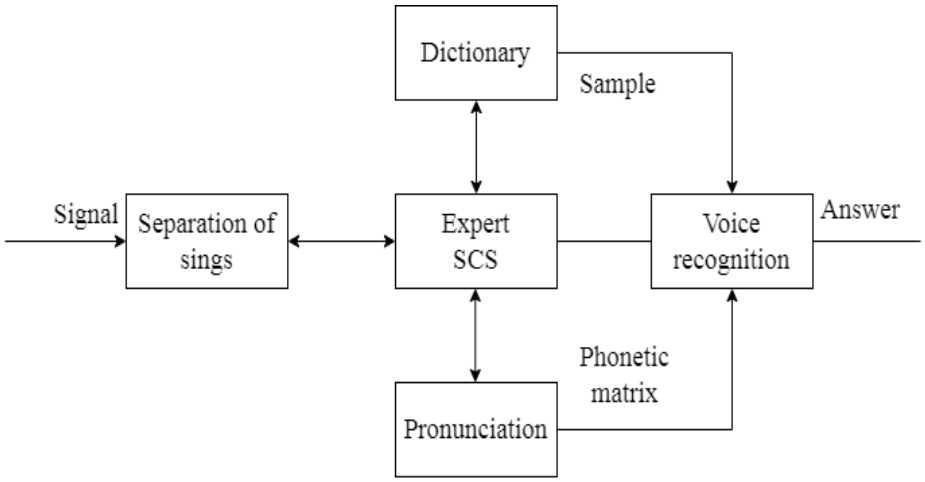

Fig 1. Scheme of the voice recognition process

The process of automatic determination of "who speaks" is performed on the basis of individual information input to the voice signal [2]. When driving a vehicle, human voice and gesture commands are entered as input data of the vehicle [3].

Today, the most popular car voice control systems are Android Auto [4] and Apple CarPlay [5]. To use the car's voice control functions, these systems use voice assistants.

After analyzing the operation of the car system "Android Auto", we can conclude that one of the problems is the voice control system - it's voice authentication. Due to this problem, criminals can perform cyber attacks, so-called inaudible or ultrasonic commands (DolphinAttak), on voice control systems.

To ensure the reliability of this system, you can use the method of "Hidden Markov model" [6]. There are two ways to send a voice signal to your device. They use the phonetic and whole word approach. The method is to identify the speaker and authorize it next to the voice database. First, the system learns with the help of certain voices, then it is tested with an unknown voice and then the system recognizes the user who owns the unknown voice. The recognition system is divided into two subsystems, such as text-dependent and text-independent.

Also, the car system "Android Auto" with voice control is vulnerable to other types of cyberattacks, such as attack by light commands performed by, giving a light command to the microphone of the voice control system with a special device for example, tinting car windows and more.

Analyzing the work of the car system "Apple CarPlay", we can conclude that the problem of voice assistant is often cloud data processing and dependence on the quality of Internet connection. So you need a quality and fast internet connection to ensure the reliability of your voice control system. To do this, you can buy a 3G / 4G WI-FI raster in the car, which will ensure the speed of your system with cloud data processing. The router can be connected to the car's cigarette lighter, to the USB port of your car, depending on your choice and characteristics of the car. But we should not forget that it is impossible to connect and configure devices from other manufacturers often enough, or they will work with limited functionality [7].

High-quality and stable Internet connection will also help to solve another shortcoming of the use of voice control, namely the malfunction, various system failures due to untimely software updates. To resolve this issue, you need to update your device to the latest available software version.

Another very serious problem that is often encountered with the voice assistant in the system "Apple CarPlay" is that it can read voice commands that were not assigned to it, ie respond to different types of noise, also due to noise voice control system may misunderstand and perform your voice team. To solve this problem, you can use the development of a system from Bose.

The company has developed a "QuietComfort Road Noise Control" system that can be installed in your car to reduce noise levels, which will ensure reliable voice control. The system consists of microphones and a set of accelerators, using acoustics installed in the car, filtering background noise, the system increases the clarity of voice commands and expands the possibilities of voice control.

You can also use the method of speech enhancement integrates the display of characteristics, time domain in a unified structure using the GAN network, it processes voice command waves and separates speech and noise signals coming into two one-dimensional layers of Fourier transform convolution, which reflect signal shapes in speech and noise spectrograms, which in turn are used to calculate losses. This method is superior to methods for improving voice commands, based on the DNN neural network.

One of the significant shortcomings of automotive systems is the voice authentication mechanism, for example, a criminal can bypass the security function of the voice assistant by pretending to be the owner by attacking light commands, thereby gaining unauthorized access to the vehicle. The study clearly demonstrated [8] how you can secretly and remotely enter voice commands with your own voice, in various ways without even attracting the attention of users.

To ensure reliability, you can use the method of dynamic time scale transformation (DTW) [9]. This method allows you to find the proximity, for two measurement sequences, in a certain period of time. It can be used to recognize a voice command if two speech signals represent the same output voice command, even at different speeds and lengths. One of the advantages of this method is ease of implementation.

Apple CarPlay, like Android Auto, is also vulnerable to cyberattacks, such as light commands. Next, the example of the threat model will show how such an attack occurs. The purpose of the thief is to remotely enter commands that pose a threat to the user's device, using a special device (laser). For example, an offender does not have physical access to a user's device, so he cannot change settings that are not available by voice, but he can gain remote access to the target device and its microphone by entering light commands. It should also be noted that remote access to the target device allows you to monitor the LEDs of the device, which in turn shows him how they react (light up) after recognizing the voice command and allows remote use as feedback. to determine the success of the attack attempt. To protect the reliability of the voice control device in the car as protection, you can use both hardware and software protection.

Methods of ensuring the reliability and protection against modern methods of cyber attacks on the voice control system

Today, Apple's Siri or Google Assistant voice assistants, used for voice control on Apple CarPlay [4] and Andoid Auto [5], respectively, are becoming popular. the method of human interaction with the car through voice control. With the advent of these systems, there has also been a need to provide protection for them. As previously described, these systems have a common vulnerability to the voice authentication system.

Next, we will discuss the methods of protection and reliability against cyber attacks by ultrasonic and light commands on a specialized computer system of voice control of the car. It should be understood that the voice control system, which depends directly on the speaker, is performed locally, and not the dependent voice control system is performed through the cloud service [10].

When a user uses a cloud service, signals that have been pre-processed are sent to servers where these signals will be recognized by machine algorithms. If the SCS recognizes the command, it will run the program to perform the operation. All commands and actions are system dependent and defined. Dedicated computer voice control systems have a wide range of functions and voice commands that are quite difficult to activate. Most security research for voice control systems focuses on cyberattacks, voice recognition algorithms [11], or malicious software.

In order to have access to control of the voice control system, Dolphin Attack must generate activation commands before the general introduction of voice control commands. Next, on the example of the voice assistant "Siri" who works in "Apple CarPlay" [4], how exactly is the generation of voice commands. Siri Voice Assistant

works in two modes, namely activation and recognition. Before executing voice commands, you need to activate it, so you need to generate two types of voice commands, for activation and basic control commands. Activation is considered successful if the voice command meets the requirements: has wake-up words "Hello, Siri" and mimics the user's voice under which the voice assistant was trained. For a thief, creating an activation team is quite difficult, unless of course he is able to record the words of the user's activation.

Generating a certain voice in "Hello, Siri" using the current speech methods and functions extracted from the recordings [12] is extremely difficult, and sometimes not possible at all, because it is unclear what set of functions is required for voice identification. Therefore, you can use two methods to create activation commands for the voice assistant.

DolphinAttax can use different voice command activation kits, with different voice tones, using speech synthesis systems. The method is used when the offender has the ability to write words or phrases of the user, with the possibility of further breaking them into phonemes and combining them into different words, including those necessary for activation.

After undergoing activation, the offender may have access to general voice control commands. it is possible to select the text of the control command and create it using language synthesis systems. The voice recognition system does not verify the identity of the control commands.

To ensure the reliability of voice control of the car, from ultrasonic (inaudible) cyber attack, I want to offer my own method of protection, which will use hardware protection and contain a device for future use.

The specialized computer voice control systems "Android Auto" [5] and "Apple Car Play" [4] have shortcomings with voice authentication. That is, the criminal for a successful cyber attack (DolphinAttack) must first remotely send an inaudible (ultrasonic) signal to wake up the system.

Such a signal, the offender can receive by recording the voice of the user on whose device the cyberattack will be directed, to further break the signal into words that are necessary for activation. The proposed method will help increase the protection of the system in the first stage of preparation for a successful cyber attack.

The essence of the method is to reduce the possibility of the offender to obtain a recording of the user's voice. To do this, use an ultrasonic microphone recording blocker. To date, there are many different types, with different characteristics and capabilities. Next (Figure 2), will show the use of an ultrasonic blocker using the algorithm of this device.

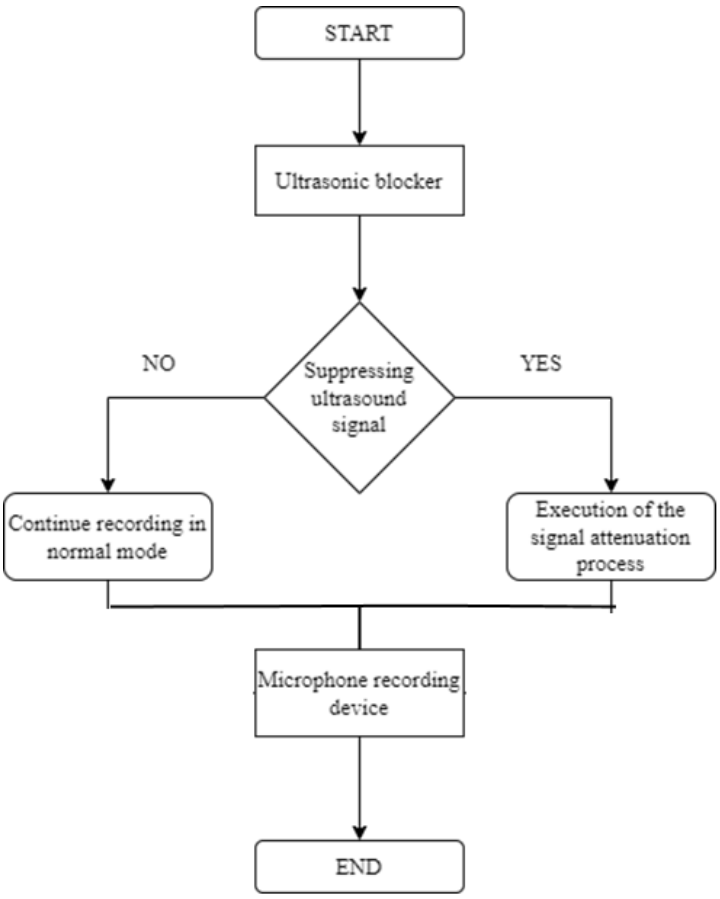

Fig. 2. Scheme of the algorithm of ultrasonic recording blocker

The size of the device is quite small and comfortable, and most importantly invisible, which makes it easy to install in the car showroom or anywhere else.

The next step will be to assess the effectiveness of Dolphin Attack's impact on various factors and methods of protection for them. For a cyberattack, the speed of recognition of different types of voice commands will not differ. Voice assistants such as Siri or Google Assistant are recommended for use in car voice control systems with minimal background noise, as SGCs are sensitive and can lead to incorrect analysis and execution of user-defined voice commands.

As the cyberattack is performed remotely, the level of background noise increases as the distance increases, as previously described, which can lead to incorrect recognition of the voice command. Next, the methods and means of protection in cyber attacks on the voice control system will be evaluated. Both hardware and software methods can be used for protection.

Hardware protection is to improve the SGC microphone and its characteristics. The main reason for a successful cyber attack is that the microphone can receive acoustic commands above 20 kHz, although ideally it should not.

In general, most microphones allow signals above 20 kHz [13], so the microphone should be extended and designed to curb acoustic signals in which the frequency is in the range of ultrasonic commands.

You can add a low-pass filter module to the microphone to detect modulated voice commands and cancel the bandwidth using modulated voice commands. This allows you to detect signals in the frequency range of the ultrasound, showing the modulation characteristics and where to modulate these signals to obtain the main frequency band.

To provide software, you need to use the unique properties of voice commands that distinguish them from the real thing. Figure 3 shows a demodulated cyberattack signal that differs from the original signal and that recorded at high frequencies in the range of 800-2400 Hz.

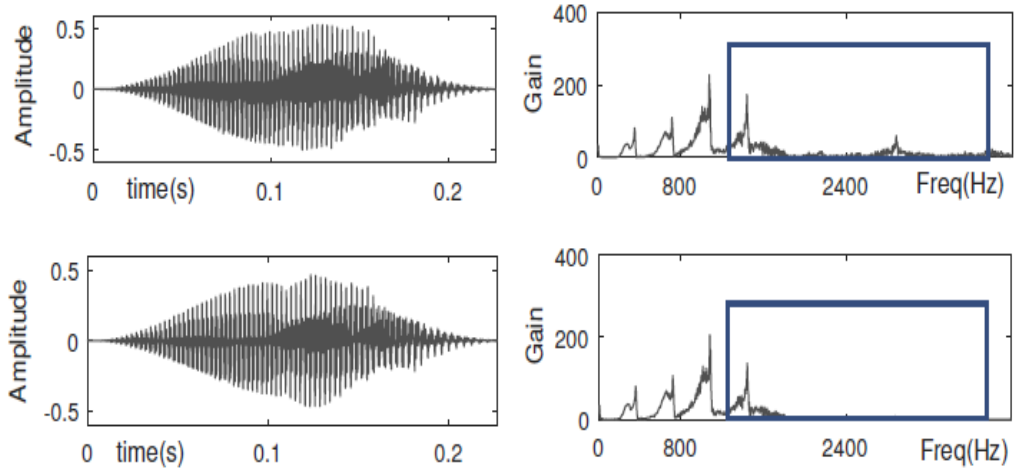

Fig. 3. Difference of demodulated signal from original [14]

The original signal produced by the Google TTS engine has a frequency of 25 kHz for modulation, so it is possible to detect "Dolphin Attack" by performing a frequency analysis in the range from 800 to 2400 Hz. To confirm the feasibility of detecting a cyber attack, the method of reference vectors as a classifier and extraction from audio functions in the frequency and time domain.

Using the created voice commands "Hello, Siri", with the help of special programs for converting text into a voice command, two samples of voice commands were obtained, in which one was recorded and the other was played. In order to teach the classifier on the method of reference values, to detect malicious voice commands, it is necessary to use several recorded audio samples, other samples can be used for testing. The classifier can distinguish restored audio recordings from those recorded with a true positive result and a negative value of one hundred percent.

The result of using a classifier made by the method of reference vectors, shows that this software method of protection can be detected for malicious cyberattacks.

The next method of protection against inaudible cyberattacks will be to search for and detect signs of non-linearity of the signal that is transmitted to the microphone of the voice control system. To do this, you need to understand whether it is possible to identify traces of non-linearity, which the offender will not be able to get rid of. But first you need to understand exactly how acoustic nonlinearity works.

In general, microphones and speakers are designed as linear systems, which means that the output signals are linear combinations of input signals. In the power amplifier used in microphones and speakers, the input audio signal is s (t), then the output signal should ideally be:

$$S_{out}(t) = A_1 s(t), \quad (1)$$

where A_1 is the gain of the amplifier;

In practice, components in microphones can usually be linear only in audible frequency ranges, ie greater than 20 kHz. In ultrasonic bands where the frequency is less than 25 kHz, they do not show linearity [15]. It follows that for ultrasonic signals the output of the amplifier is calculated:

$$s_{out}(t) = \sum_{i=1}^{\infty} A_i s^i(t) = A_1 s(t) + A_2 S^2(t) + A_3 S^3 \cdots \approx A_1 s(t) + A_2 S^2(t) \quad (2)$$

[16] shows how it is possible to reproduce ultrasonic signals that can be recorded by a microphone, but they will be inaudible to humans. In the ultrasonic speaker there is a possibility of reproduction of two inaudible tones:

$s_1(t) = \cos(2\pi f_1 t)$ with frequency $f_1 = 38$ kHz і $s_2(t) = \cos(2\pi f_2 t)$ with $f_1 = 40$ kHz.

When the combined signal passes through a nonlinear microphone at the output it becomes:

$$\begin{aligned} s_{out}(t) = {} & A_1 s_{hi}(t) + A_2 s_{hi}^2(t) = A_1(s_1(t) + s_2(t)) + A_2(s_1(t) + s_2(t))^2 \\ & = A_1 \cos(2\pi f_1 t) + A_1 \cos(2\pi f_2 t) + A_2 \cos^2(2\pi f_1 t) + A_2 \cos^2(2\pi f_2 t) \\ & + 2A_2 \cos(2\pi f_1 t) \cos(2\pi f_2 t), \end{aligned} \quad (3)$$

This signal has frequency componentsv f_1, f_2, $2f_1$, $2f_2$, $f_2 + f_1$ and $f_2 - f_1$. The microphone before, digital processing and recording uses a low-pass filter to remove components higher than 24 kHz. So frequencies

$$s_{low}(t) = A_2 + A_2 \cos(2\pi(f_2 - f_2)t), \quad (4)$$

In general,, $f_2 - f_1 = 2$ kHz з recorded by the microphone, this shows a property that allows you to send an inaudible signal, with the ability to generate a copy of the sound in the middle of the microphone.

Thus we mark the signal of the voice command: "Siri, pave the route…", which was pronounced by the user - $v(t)$, when he will say this command, the expression will be executed:

$$s_h = v(t) + n(t), \quad (5)$$

where $n(t)$- microphone noise;

Let the offender reproduce this voice command using ultrasound, recorded signal s_{atk} look like:

$$s_{atk} = \frac{A_2}{2}(1 + 2v(t) + v^2(t)) + n(t), \quad (6)$$

Figure 4 shows the spectrum of the voice command for the s_h and s_{atk}, as these signals are almost similar in structure, which means that the text converter outputs the same text for the s_h and s_{atk}.

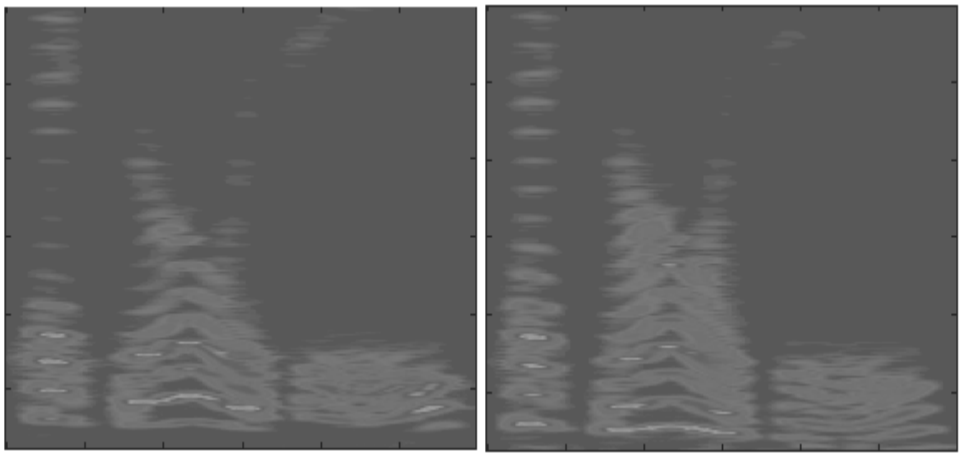

Fig. 4. Spectrogram for signal s_h and s_{atk}k, voice command "Siri, pave route…"

Based on this, we can conclude that for protection you need to check any signal (input) and determine whether it is a low-frequency user-specified, or a copy of the high-frequency cyber attack.

Attack by light commands is similar to cyberattacks by ultrasonic commands, the difference is that it uses a special device (laser) to attack. Hardware and software protection methods are used to protect against this type of attack. The software method of protection is to apply an additional level of authentication to the voice control system. In [17], the authors use an additional step of user authentication, thus trying to protect against the execution of unauthorized conference commands. The method is to use an additional authentication step before executing critical commands and reduce attempts to enter the wrong password if the system supports this feature.

This method can also help if the offender is unable to hear the response of the voice control system because it is far from the attacked device. For example, the system will ask any random question before executing a voice command, to which the offender will not be able to answer, thus stopping the attack.

The next method of protection is to use the operation of sensor algorithms, and use methods of merging them to detect commands entered on the basis of light [18]. Voice assistants often have and use multiple microphones. The essence of the method is that the offender uses one special device (laser) to attack light

commands and uses only one microphone that receives the signal, while other microphones do not receive anything. So you can try to detect the attack using comparisons of signals from multiple microphones, ignoring voice commands that are entered using a special device (laser). This method can be effective only when one attacking device is running.

The imperfections of specialized computer systems with voice control of the car, namely with the problems of authentication or lack thereof in general, allow criminals to access various functions and capabilities of the car. Using various types of cyberattacks, such as attacks by light commands, or the introduction of inaudible (ultrasonic) commands, with the help of special devices aimed at microphones of voice control of the car, will allow criminals to bypass the imperfections of the driver authentication system.

The previously described protection methods allow to ensure the reliability of the car's voice control systems, quite effectively, but not as much as possible. As a result of the review of the original sources of the authentication system of different cars, the material obtained, the analysis of which led to the conclusion that to ensure the reliability of the car voice control system, you need to use an additional authentication system that will allow only the driver or proxies. system management and more. To solve the problem of additional authentication, an autonomous security system will be created, with the possibility of direct authentication using the fingerprints of the car owner or proxies. The main components of the system will be the Arduino UNO board, fingerprint scanner, LCD display, servomotor. The Arduino IDE will be used to download code and program the Arduino UNO board. The algorithm of the system (Figure 5) shows that if the user is not authenticated correctly, the system will not be able to provide access to the car's functions.

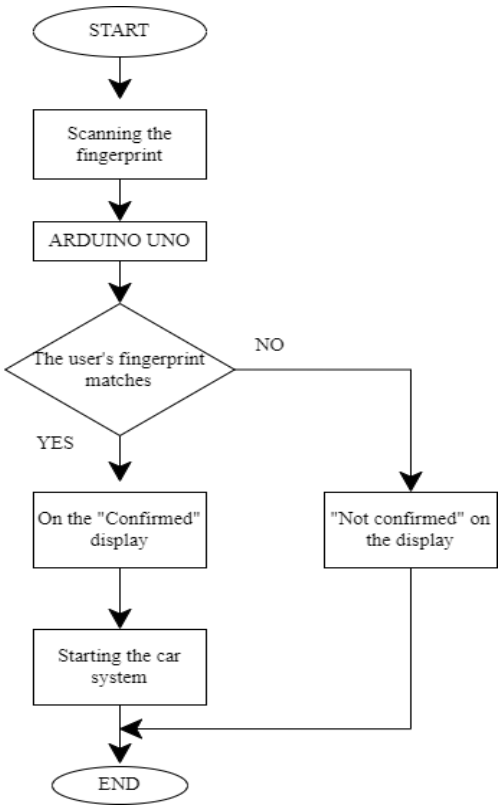

Fig. 5. Block diagram of the biometric authentication system algorithm

Biometric fingerprint recognition technology is a new and modern method for ensuring reliability and protection for security systems. This method uses the physical presence of the user to authenticate the user. Today, fingerprint recognition is widely used in various biometric systems, such as telephones, smart devices, biometric locks, bank payments and more [19]. The use of biometric authentication as a personal code as a personal code is considered a traditional method.

The use of this biometric authentication system will be quite reliable and will allow the user to provide reliability for specialized computer systems of the car, and block the criminal's access to the voice control system, which in turn will prevent various methods of cyberattacks

Known approaches to solving this problem are based on the work [20], which states that biometric security technologies are one of the most effective protection systems, and are increasingly becoming everyday attributes in the lives of ordinary people. In recent years, these systems have become widespread in the production of mobile technology, ie smartphones are built-in fingerprint scanners, voice recognition and more.

Particular attention is paid to the problem of authentication, related to the development of methods and tools to ensure the reliability of the car's SCS. That is, through the process of authentication of a person, using the comparison of its characteristics with the characteristics that were previously entered into the system, it is possible to determine as accurately as possible whether the user has appropriate access to the requested information or not. This makes it possible to ensure the reliability of the current problem of information security.

An important point for research in automotive authentication systems is the fact that in today's world there is a high demand for reliable and safest systems in vehicles. Thus, the design and development of hardware and software biometric security system using fingerprint technology to prevent unauthorized access to the car is simple and useful to use. The hardware and software implementation of the system will use an additional method of user authentication, which will be based on the ability to start the car's ignition system, which in turn will use its functionality and prevent its use in case of incorrect biometric authentication with fingerprint.

Fingerprint sensor, allows you to match the image of the user's fingerprint with what is stored in the system memory of the sensor. The research program focuses on a tool to obtain answers and follow instructions according to the results obtained, using the Android Uno microcontroller and includes the following security issues. Who use the analysis of the obtained results and check whose fingerprints can get access rights to turn on a specialized computer system of the car.

Experimental results and analysis of an additional user authentication system

Checking access to the car's biometric authentication system can be considered successful if the user turns on the car's system using their own fingerprint, which is registered in the device's memory. If the user registered in the system is unable to do so, the system may be considered defective.

The experiment is performed by detecting fingerprints for the system, aimed at finding the value of the success rate of other fingerprints that have not been entered into the scanner database. The experiment is performed by setting the fingerprint of the user, and then continuing to establish the fingerprint of the second user. To determine the percentages, changes in the ten right fingerprints of users that were used using another person's scan pattern that was not registered in the fingerprint sensor scanner will be checked.

An important point for the study is the position that the scanner is very sensitive to the placement of the user's fingerprints. The location of the fingerprint should be exactly on the layer of the scanner glass so that the fingerprint is read clearly and in accordance with the input and stored in the system. The results of the study of the fingerprint module are shown in table 2.

Table 2 shows the matrix of test results. The matrix displays the actual and incorrect number of predictions in the matrix test data. The input data of the matrix have the following values:

- positive (true) - is the number of fingerprints of users using a scanner;
- negative (true) for the number of prints of other users that are detected incorrectly;
- false-positive fingerprint results of another user being entered, verified and correct;
- false-negative when the fingerprint scanner module indicates that the car startup system could not be accessed.

Table 2

Study of the fingerprint scan of the user of the car's biometric system

Fingerprint	Positive (real)	Negative (real)	False-positive	False-negative	Car system
1	1	0	0	0	included
2	1	0	0	0	included
3	1	0	0	1	excluded
4	1	0	0	0	included
5	1	0	0	0	included
6	1	0	0	0	included
7	1	0	0	1	excluded
8	1	0	0	0	included
9	1	0	0	0	included
10	1	0	0	0	included

Based on the above results, it can be concluded that the success rate of the fingerprint of the user who can access the car is 90 percent.

Similar works

There are many articles on this topic, for example, a scientific article [21] presented the results of user interaction with a specialized computer system "Android Auto". The study examined the interaction of drivers with the functions of the voice control system and the safety of their control on the road. The results of the study showed that using the "Android Auto" system is quite safe.

McAfee and its partners have published a report called "Precautionary Software" [22], in which they analyzed the new threats and risks in the automotive specialized computer system that are present in modern cars. In

[23-24], the authors show a comprehensive approach to show that the safety of modern cars may be compromised due to interference and interference with the passage of Bluetooth and Wi-Fi signals. In some articles, such as [25], security and privacy issues in car voice control systems are solved using different cryptographic methods, or using different secure development environments [26].

Conclusions

As a result of summarizing the literature to ensure the reliability of a specialized computer voice control system, a number of problems have been identified, the main of which is the imperfection of the user authentication system.

To implement the solution to this problem, a hardware and software product of an additional biometric automotive user authentication system was created and developed. The system was created to ensure the reliability of cyber attacks on the voice control system.

An experiment with a biometric authentication system found that the success rate for a registered fingerprint user who can access the car is ninety percent.

References

1. What is a computer system? - definition from technopedia. URL: https://uk.theastrologypage.com/computer-system (Accessed on: 06.11.2021).
2. Theology voice control. URL: http://tehnology.com (Accessed on: 22.11.2021).
3. S.Mohith., S.Santhanalakshmi., M.Sudhakaren. Gesture and Voice Controlled Robotic Car using Arduino.2018., pp 3392-3396.
4. What is Android Auto? And how it works. URL: http://www.rcd330.com.ua/ chto-takoe-android-auto (Accessed on: 05.12.2021).
5. How to set up Apple CarPlay in your car (manual). URL: https://uk.vemprarua.org/how-setup-apple-carplay-your-car (Accessed on: 15.12.2021).
6. A Voice Identification System using Hidden Markov Model T. K. Das., Khalid M.O., Nahar SITE, VIT University, Vellore – 632014, Tamil Nadu, India; Department of Computer Science, Yarmouk University, Irbid – 21163, Jordan.
7. V. N. Shmatkov, P. Bonkowski, D. S. Medvedev [et al.] Interact with IOT devices using the voice interface // Scientific and technical Bulletin of information technologies, mechanics and optics, 2019
8. Nicholas Carlini., Pratyush Mishra., Tavish Vaidya., Yuankai Zhang., Micah Sherr.,Clay Shields., DavidWagner.,Wenchao Zhou. 2016. Hidden Voice Commands. In 25th USENIX Security Symposium (USENIX Security 16). USENIX Association, Austin, TX, 513–530.
9. Collection of scientific works. Center for Strategic Studies of the National University of Defense of Ukraine named after Ivan Chernyakhovsky. 2018. № 3(64). C. 149.
10. Chaouki Kasmi and Jose Lopes Esteves. 2015. IEMI threats for information security: Remote command injection on modern smartphone
11. Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. 2016. Hidden voice commands. In Proceedings of the USENIX Security Symposium.
12. Dibya Mukhopadhyay, Maliheh Shirvanian, and Nitesh Saxena. 2015. All your voices are belong to us: Stealing voices to fool humans and machines. In Proceedings of the European Symposium on Research in Computer Security. Springer, 599–621.
13. STMicroelectronics. 2016. MP34DB02 MEMS audio sensor omnidirectional digital microphone. http://www.mouser.com/ds/2/389/mp34db02-955149.pdf. 2016.
14. Yitao He, Junyu Bian, Xinyu Tong, Zihui Qian, Wei Zhu, Xiaohua Tian, Xinbing Wang. Canceling Inaudible Voice Commands Against Voice Control Systems. 2019. Article No.: 28. Pages 1-15.
15. DOBRUCKI, A. Nonlinear distortions in electroacoustic devices. Archives of Acoustics 36. 2 (2011). 437–460.
16. Nirupam Roy, Sheng Shen, Haitham Hassanieh, Romit Roy Choudhury University of Illinois at Urbana-Champaign. Inaudible Voice Commands: The Long-Range Attack and Defense.
17. Takeshi Sugawara, The University of Electro-Communications; Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu, University of Michigan. Light Commands: Laser-Based Audio Injection Attacks on Voice-Controllable Systems August 12–14. 2020.
18. D. Davidson, H. Wu, R. Jellinek, T. Ristenpart, and V. Singh, “Controlling UAVs with sensor input spoofing attacks,” in USENIX WOOT, 2016.
19. Omidiora E. O., Fakolujo O. A., Arulogun O. T., Aborisade D. O. 2011. A Prototype of a Fingerprint Based Ignition Systems in Vehicles. 62(2): 164-171.
20. Tomas Trainys, Algimantas Venčkauskas. Encryption Keys Generation Based on Bio- Cryptography Finger Vein Method. CEUR Workshop Proceedings 2145 (2018) 106-111
21. R Ramnath., N Kinnear., S Chowdhury., THyatt. Interacting with Android Auto and Apple CarPlay when driving: The effect on driver performance. 2020. pp 12-17.
22. Stuart McClure. Caution: malware ahead. Vision Zero International. 2013.
23. Stephen Checkoway, Damon McCoy, Brian Kantor, Danny Anderson, Hovav Shacham, Stefan Savage, Karl Koscher, Alexei Czeskis, Franziska Roesner, Tadayoshi Kohno, et al. 2016.
24. Charlie Miller., Chris Valasek. Remote exploitation of an unaltered passenger vehicle. Black Hat. 2015.
25. Ramon de Graaff. 2015. Controlling your Connected Car. 2015.
26. Yunhan Jack Jia., Ding Zhao., Qi Alfred Chen.,Z Morley Mao.Towards Secure and Safe Appified Automated Vehicles. 2017.

https://doi.org/10.31891/csit-2022-2-12
UDC 004.8

Andrii POPOVYCH, Vitalii YAKOVYNA
Lviv Polytechnic National University

COVID-19 MORTALITY PREDICTION USING MACHINE LEARNING METHODS

The paper reports the use of machine learning methods for COVID-19 mortality prediction. An open dataset with large number of features and records was used for research. The goal of the research is to create the efficient model for mortality prediction which is based on large number of factors and enables the authorities to take actions to avoid mass spread of virus to and reduce the number of cases and deaths. Feature selection was conducted in order to remove potentially irrelevant input variables and improve performance of machine learning models. The classic machine learning models (both linear and non-linear), ensemble methods such as bagging, stacking and boosting, as well as neural networks, is used. Comparison of efficiency of ensemble methods and neural networks compared to classic ML methods such as linear regression, Support Vector Machines, K-nearest neighbors etc. is conducted. Ensemble methods and neural networks show much greater efficiency than classical ones. Feature selection does not significantly affect the prediction accuracy.

The scientific novelty of this paper is the large number of machine learning models trained on the large-scale dataset with significant number of features related to different factors that can potentially affect COVID-19 mortality, as well as further analysis of their efficiency. This will assist to select the most valuable features and to become a basis for creating a software designed for tracking the dynamics of the pandemic.

The practical significance of this paper is that present study can be useful for authorities and international organizations in prevention of COVID-19 mortality increase by taking proper preventive measures.

Keywords: machine learning, COVID-19, mortality prediction, ensemble methods, neural networks, feature selection.

Андрій ПОПОВИЧ, Віталій ЯКОВИНА
Національний університет «Львівська політехніка»

ПРОГНОЗУВАННЯ СМЕРТНОСТІ ВІД COVID-19 МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Дана стаття описує використання методів машинного навчання для передбачення рівня смертності від COVID-19. Для дослідження було використано відкритий набір даних з великою кількістю ознак та записів. Метою даного дослідження є створення ефективної моделі для передбачення рівня смертності, що базується на великій кількості чинників та дозволить компетентним органам вжити превентивні заходи для запобігання масовому поширенню COVID-19 та зменшення кількості хворих та померлих від хвороби. Проведено відбір ознак з метою усунення потенційно нерелевантних вхідних змінних та покращення продуктивності моделей машинного навчання. Було використано класичні моделі машинного навчання (як лінійні, так і нелінійні), ансамблеві методи, зокрема беггінг, стекінг та бустинг, а також нейронні мережі. Виконано порівняння ефективності ансамблевих методів порівняно з класичними методами машинного навчання, такими як лінійна регресія, методи опорних векторів, K найближчих сусідів та інші. Ансамблеві методи та нейронні мережі показують значно більшу ефективність, ніж класичні. Відбір ознак не має значного впливу на точність передбачення.

Наукова новизна даної роботи полягає в великій кількості моделей машинного навчання, натренованих на великому наборі даних, що містить значну кількість ознак, які стосуються різноманітних чинників, які потенційно можуть вплинути на смертність від COVID-19, та в подальшому аналізі їх ефективності. Це може допомогти відібрати найбільш значущі ознаки та стати основою у створенні програмних засобів, призначених для відстеження динаміки хвороби.

Практичне значення даної роботи полягає в тому, що наявні в ній дослідження можуть бути корисні для дослідників, закладів охорони здоров'я, державних органів та міжнародних організацій в запобіганні зростання смертності від COVID-19 шляхом вжиття відповідних запобіжних заходів.

Ключові слова: машинне навчання, прогнозування смертності від COVID-19, ансамблеві методи, нейронні мережі, відбір ознак.

Introduction

The COVID-19 pandemic caused by SARS-CoV-2 strain, which started in December 2019 in Wuhan (Hubei province, China), triggered severe global social and economic outcomes around the world. As of May 29, 2022, more than 528 million cases have been registered worldwide, including more than 6.28 million deaths. By the late 2020 - early 2021 when the mass production of vaccines and the mass vaccination started, in order to reduce morbidity and mortality the governments were forced to take strict preventive measures such as lockdowns, social distancing, travel restrictions, wearing masks, quarantines, curfews, workplace hazard controls, postponing or cancelling the events, testing systems, etc.

To mitigate the effects of pandemic and reduce the number of casualties it is crucial to have an instrument which considers different factors that can significantly affect the course of the pandemic, in particular demographic, economic, geographical, etc. This will enable researchers and authorities to better understand dynamics of the pandemic and take proper preventive actions.

The paper describes research and efficiency comparison of different machine learning models using large-size dataset with many features which will potentially improve mortality prediction accuracy.

Related works

In more than two years since the outbreak of the pandemic, a large number of studies have been conducted to predict the COVID-19 mortality rate. Most of them use the clinical and laboratory results of hospitalized patients as input data. These studies used different models of machine learning, feature selection methods, as well as metrics and indicators, which assessed the effectiveness of the models and the quality of their predictions.

Early mortality prediction using machine learning based on based on typical laboratory results and clinical data registered on the day of intensive care unit admission is considered in [1]. Such machine learning algorithms as Random Forest, logistic regression, gradient boosting classifier, Support Vector Machine classifier, and artificial neural network algorithms were used to build classification models. The impact of each marker on the RF model predictions was studied by implementing the LIME-SP technique. The study [2] aimed to compare several ML algorithms to predict the COVID-19 mortality using the patient's data at the first time of admission. An Information GainRatio Attribute evaluation (GA) method was used to select the features. Seven ML algorithms including the J48 decision tree, Random Forest, K-nearest neighborhood, multi-layer perceptron, Naïve Bayes, eXtreme gradient boosting (XGBoost), and logistic regression were applied. Random Forest had better performance than other ML algorithms.

In the study [3] inspired modification of partial least square (SIMPLS)-based model was developed to predict hospital mortality. Latent class analysis (LCA) was carried to cluster the patients with COVID-19 to identify low- and high-risk patients. SIMPLS-based model was able to predict hospital mortality with moderate predictive power and high accuracy. Clustering analysis identified high- and low-risk patients among COVID-19 survivors. The aim of the next study [4] was the development and prospective validation of a state-of-the-art machine learning model to provide mortality prediction within 72 hours after confirmation of SARS-CoV-2 infection. Traditional machine learning models were evaluated independently as well as in a stacked learner and various recurrent neural network architectures were considered. The GRU-D recurrent neural network achieved peak cross-validation performance.

The study [5] aims to train several ML algorithms to predict the COVID-19 in-hospital mortality and compare their performance to choose the best performing algorithm. Six feature scoring techniques and nine well-known ML algorithms were used. To evaluate the models' performances, the metrics derived from the confusion matrix calculated. Experimental results indicated that the Bayesian network algorithm has been more successful in predicting mortality. This study [6] was conducted to develop a machine learning model to predict prognosis based on sociodemographic and medical information. The least absolute shrinkage and selection operator (LASSO), linear Support Vector Machine, SVM with radial basis function kernel, Random Forest and K-nearest neighbors were tested. LASSO and linear SVM demonstrated high sensitivities and specificities while maintaining high specificities, as well as high area under the receiver operating characteristics curves.

Prediction of in-hospital mortality for COVID-19 patients treated with steroid and remdesivir was conducted in [7]. The important variables associated with in-hospital mortality were identified using LASSO and SHAP (SHapley Additive exPlanations) through the light gradient boosting model (GBM). Six important variables were selected. Additionally, the light GBM had high predictability for the latest data (AUC: 0.881). This study [8] aimed to develop a predictive model to predict patients' mortality from the basic medical data on the first day of admission. From different ML models the naive Bayes demonstrated the best performance with an AUC of 0.85. The ensemble model from the naive Bayes and neural network combination had slightly better performance.

The study [9] aimed to develop and compare prognosis prediction machine learning models based on invasive laboratory and noninvasive clinical and demographic data from patients' day of admission. Three SVM models were developed and compared using invasive, non-invasive, and both groups. The results suggested that non-invasive features could provide mortality predictions that are similar to the invasive. The next study [10] experimentally verified that some anti-cancer drugs can be regarded as potential treatments against COVID-19. A broad panel of time-to-event machine learning models was implemented and compared, such as Elastic net penalized Cox proportional hazards regression and Weibull accelerated failure time regression, DeepSurv neural network approach, Random Survival Forests and XGBoost Survival Embeddings.

The purpose of study [11] is to predict new cases and deaths rate one, three and seven-day ahead during the next 100 days. Three methods (LSTM, Convolutional LSTM, and GRU) and their bidirectional variants were used. The results show that the bidirectional models have lower errors than other models. The next study [12] is about development and testing of machine learning-based models for COVID-19 severity prediction. In this research, a new feature engineering method based on topological data analysis called Uniform Manifold Approximation and Projection (UMAP) were used. UMAP has 100% accuracy, specificity, sensitivity, and ROC curve in conducting a prognostic prediction using different machine learning classifiers.

In the study [13] authors developed, verified, and deployed a stacked generalization model to predict mortality by combining 5 previously validated scores and additional novel variables reported to be associated with COVID-19-specific mortality. A ridge regularized logistic regression was chosen as the top-level model to limit overfitting and to address correlation between the component models. The objective of the next study [14] was to develop and validate models that predict mortality of patients diagnosed with COVID-19 admitted to the hospital. A

linear logistic regression and non-linear tree-based gradient boosting algorithm were used. Both models outperformed age-based decision rules used in practice.

The objective of study [15] was to identify prognostic serum biomarkers in patients at greatest risk of mortality. The developed Support Vector Machine model achieved 91% sensitivity and 91% specificity (AUC 0.93) for predicting patient expiration status on held-out testing data. The next study [16] aimed to develop risk scores based on clinical characteristics at presentation to predict ICU admission and mortality in COVID-19 patients. Logistic regression was used to identify independent clinical variables predicting the two outcomes. The risk score model yielded good accuracy for predicting ICU admission and for predicting mortality for the testing dataset.

The next study [17] leverages a database of blood samples to identify crucial predictive biomarkers of disease mortality. For this purpose, multi-tree XGBoost classifier selected three biomarkers that predict the mortality of individual patients more than 10 days in advance with more than 90% accuracy. The aim of next study [18] was to develop an accurate model for predicting COVID-19 mortality using epidemiolocal and clinical variables and for identifying a high-risk group of confirmed patients. Risk scores for COVID-19 mortality prediction model were developed by logistic regression analysis.

This study [19] seeks to develop and validate a data-driven personalized mortality risk calculator for hospitalized COVID-19 patients. The COVID-19 Mortality Risk tool was developed using the XGBoost algorithm to predict mortality. In the last study [20] a bootstrap averaged ensemble of Bayesian networks was also learned to construct an explainable model for discovering actionable influences on mortality and days to outcome. XGboost and logistic regression model yielded the best performance on risk stratification and mortality prediction respectively.

As we can see, the vast majority of studies related to the COVID-19 mortality prediction of from focus on predicting the survival of individual patients who have been hospitalized with a confirmed diagnosis. These studies are based on data provided by health facilities. So, the aim of this study to predict the COVID-19 mortality rate among the population on the basis of a large number of potentially relevant factors that may affect the pandemic. This task involves the selection of the appropriate set of input data, as well as the selection of the optimal prediction method and the factors influencing its results.

Dataset description and exploratory data analysis

An open dataset [21] which contains data related to COVID-19 outbreak in the US, including data from 3142 counties of 49 US states from the beginning of the outbreak (January 2020) to June 2021, was used for study given in this paper.

This data was collected from many public scientific, governmental and other online databases and include daily number of COVID 19 confirmed cases and deaths and features, as well as features that may be relevant to the dynamics of the pandemic: demographic, geographic, climatic, social, etc.

The dataset consists of 992266 records and 64 features. The target variable is daily number of COVID-19 deaths in each county.

The dataset is essentially an aggregation of big amount of data collected from large number of open sources. The data in the dataset were preliminarily prepared by its authors. In particular, KNNimputer was used to impute missing data, and the records about counties with values of both fixed features and temporal features missed for all dates were deleted.

The correlation matrices for some features are presented in Fig. 1. We can see that significant correlation between them and target variable is absent.

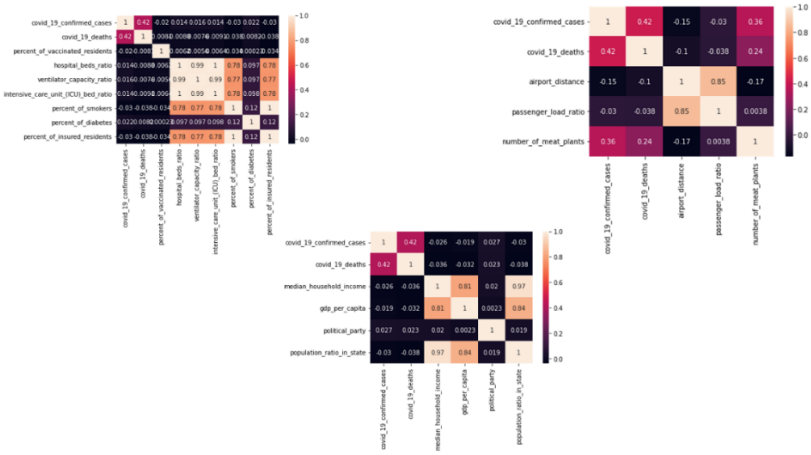

Fig.1. Pearson correlation coefficients matrices for some features

Feature selection

As the dataset contains large number of features, it is necessary conduct feature selection to select a set of input variables that are the relevant to target variable. This will potentially reduce the dimensionality of the training

set, improve model performance and reduce it fitting time. As it is unknown what set of features will be optimal, the following algorithms were used:

1) Boruta [22]. This algorithm based on Random Forest creates random shuffled shadow copies for each feature and determines their Z-scores. Feature is removed if its score is lower than maximum score of its shadow copies. 6 features were selected by this algorithm (9.52% of total number of features).

2) Recursive Feature Elimination (RFE) [23]. This algorithm uses an external estimator to assign some weight coefficients to initial set of features, then features with the lowest weights are pruned. Procedure is recursively repeated until the desired number of features is reached. 32 features were selected by this algorithm (50.7% of total number of features).

3) Recursive Feature Elimination with cross-validation (RFECV) [24] which allows to get the optimal set of features. 22 features were selected by this algorithm (34.9% of total number of features.

Comparison of efficiency of different machine learning models

The first step is applying linear machine learning models to both the entire dataset and the selected features. Such models as linear [25], logistic [26], ridge [27] and ElasticNet [28] regression, as well as stochastic gradient descent [29], were used. For model evaluation, metrics such as mean absolute error (MAE), mean squared error (MSE), its root (RMSE) and coefficient of determination (R^2 score) were used. Data was split with ratio: 75% - training set, 25% - test set. Results are presented in Table 1.

Table 1.

Comparison of efficiency of linear models for different sets of features

Model/metric	MAE	MSE	R^2 score	RMSE
For all features				
Linear	0.674	8.265	0.329	2.875
Logistic	0.480	10.163	0.174	3.188
Ridge	0.674	8.265	0.329	2.875
ElasticNet	0.691	9.198	0.253	3.033
SGD	0.628	8.366	0.320	2.892
For features selected by Boruta algorithm				
Linear	0.637	8.367	0.320	2.893
Logistic	0.480	11.052	0.102	3.324
Ridge	0.637	8.367	0.320	2.893
ElasticNet	0.691	9.198	0.253	3.033
SGD	0.641	8.502	0.309	2.916
For features selected by RFE algorithm				
Linear	0.674	8.265	0.329	2.875
Logistic	0.480	10.163	0.174	3.188
Ridge	0.674	8.265	0.329	2.875
ElasticNet	0.691	9.198	0.253	3.033
SGD	0.628	8.366	0.320	2.892
For features selected by RFECV algorithm				
Linear	0.675	8.293	0.326	2.880
Logistic	0.479	10.545	0.143	3.247
Ridge	0.674	8.293	0.326	2.880
ElasticNet	0.693	9.267	0.247	3.044
SGD	0.660	8.709	0.292	2.951

The next step is the analysis of efficiency of some non-linear machine learning models. The following methods were used: K-nearest neighbors [30], Support Vector Machine [31], decision tree [32]. Results are presented in Table 2.

Table 2.

Comparison of efficiency of non-linear models for different sets of features

Model/metric	MAE	MSE	R^2 score	RMSE
For all features				
DecisionTree	0.606	11.942	0.030	3.456
SVR	0.487	9.589	0.221	3.097
KNeighbors	0.602	9.820	0.202	3.137
For features selected by Boruta algorithm				
DecisionTree	0.630	14.910	0.021	3.860
SVR	0.484	9.733	0.210	3.120
KNeighbors	0.612	10.010	0.187	3.164
For features selected by RFE algorithm				
DecisionTree	0.606	11.942	0.030	3.456
SVR	0.487	9.589	0.221	3.097
KNeighbors	0.604	10.846	0.193	3.293
For features selected by RFECV algorithm				
DecisionTree	0.611	14.930	0.021	3.864
SVR	0.486	9.659	0.215	3.108
KNeighbors	0.612	9.905	0.206	3.147

In general, non-linear models with selected features show slightly worse results than with entire dataset.

The next step is to compare ensemble methods, in particular:

1) Bootstrap aggregation (bagging) [33] - algorithm is trained on random data subsets several times, then the results are averaged. In this study decision tree and Random Forest [34] are used.

2) Boosting [35] - several algorithms are trained consistently; each subsequent algorithm focuses on samples misclassified by previous ones. Gradient boosting [36] (based on decision tree), AdaBoost [37] and XGBoost [38] were used.

3) Stacked generalization (stacking) [39] - several algorithms are trained using the available data, then the results are used as inputs by final estimator which makes the final decision. Gradient boosting, decision tree and Random Forest were used to create ensemble. Results are presented in Table 3.

Table 3.

Comparison of efficiency of ensemble models for different sets of features

Model/metric	MAE	MSE	R^2 score	RMSE
For all features				
AdaBoost	0.451	6.280	0.490	2.506
Bagging	0.505	5.390	0.562	2.322
Gradient Boosting	0.549	5.716	0.536	2.391
XGB	0.508	5.390	0.562	2.322
Random Forest	0.505	5.419	0.560	2.328
Stacking	0.497	5.145	0.582	2.268I
For features selected by Boruta algorithm				
AdaBoost	0.456	6.286	0.489	2.507
Bagging	0.512	5.583	0.546	2.362
Gradient Boosting	0.551	5.906	0.520	2.430
XGB	0.516	5.700	0.537	2.387
Random Forest	0.511	5.182	0.579	2.276
Stacking	0.506	5.091	0.586	2.256
For features selected by RFE algorithm				
AdaBoost	0.451	6.280	0.490	2.506
Bagging	0.505	5.390	0.562	2.321
Gradient Boosting	0.549	5.716	0.536	2.391
XGB	0.508	5.390	0.562	2.322
Random Forest	0.505	5.419	0.560	2.328
Stacking	0.497	5.145	0.582	2.268
For features selected by RFECV algorithm				
AdaBoost	0.465	6.845	0.444	2.616
Bagging	0.508	6.140	0.501	2.478
Gradient Boosting	0.550	5.890	0.521	2.427
XGB	0.514	6.284	0.489	2.507
Random Forest	0.508	6.168	0.499	2.484
Stacking	0.516	5.559	0.548	2.358

We can see that results of ensemble models are much better than results of models mentioned above.

Finally, let's compare efficiency of some deep learning models. For comparison, two neural networks with experimentally selected topologies were created.

The first one is multilayer perceptron [40] neural network, it has four fully connected layers (one input layer and three hidden ones), each of then consists of 256, 128, 64 and 32 nodes respectively. A Rectified Linear Unit (ReLU) activation function is applied to each layer. After every layer we use Dropout layer, which is used for network regularization using neurons exclusion with certain rate (0.2 in our case) to prevent overfitting. Adam optimizer was selected and number of epochs is 100.

The second one is convolutional neural network [41], which contains one input layer with 64 nodes and one hidden layer with 32 nodes. The Flatten layer designed for converting input data into one-dimensional vector, as well as ReLU activation function and Adam optimizer is used.

Table 4.

Comparison of efficiency of neural networks for different sets of features

Model/metric	MAE	MSE	R^2 score	RMSE
For all features				
MLP	0.588	5.641	0.542	2.375
CNN	0.522	5.782	0.530	2.405
For features selected by Boruta algorithm				
MLP	0.528	5.526	0.551	2.351
CNN	0.608	5.975	0.514	2.444
For features selected by RFE algorithm				
MLP	0.567	6.072	0.507	2.464
CNN	0.603	6.171	0.499	2.484
For features selected by RFECV algorithm				
MLP	0.558	6.723	0.454	2.593
CNN	0.515	5.741	0.534	2.396

As it is shown above, the performance of neural networks is slightly lower than ensembles.

Discussion

The comparison graphs, where efficiency of studied machine learning models for both all and selected features is displayed, are shown in Fig. 2-5.

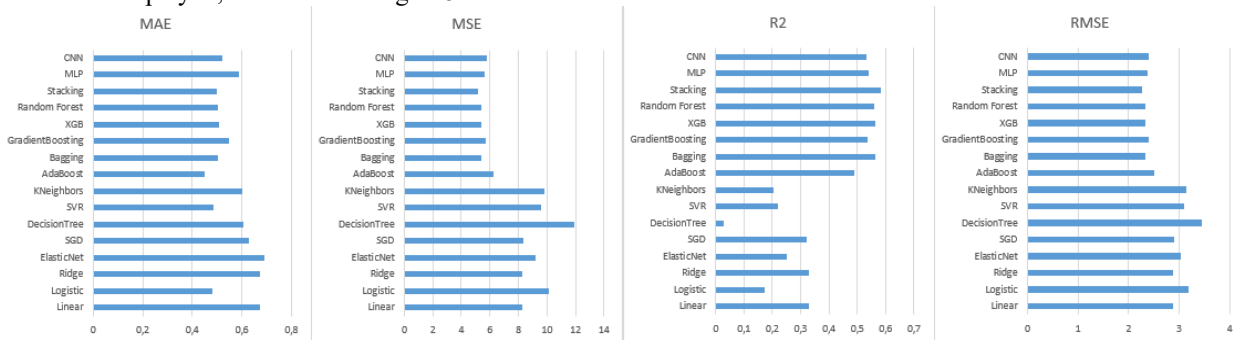

Fig.2. Comparison of efficiency of models (all features)

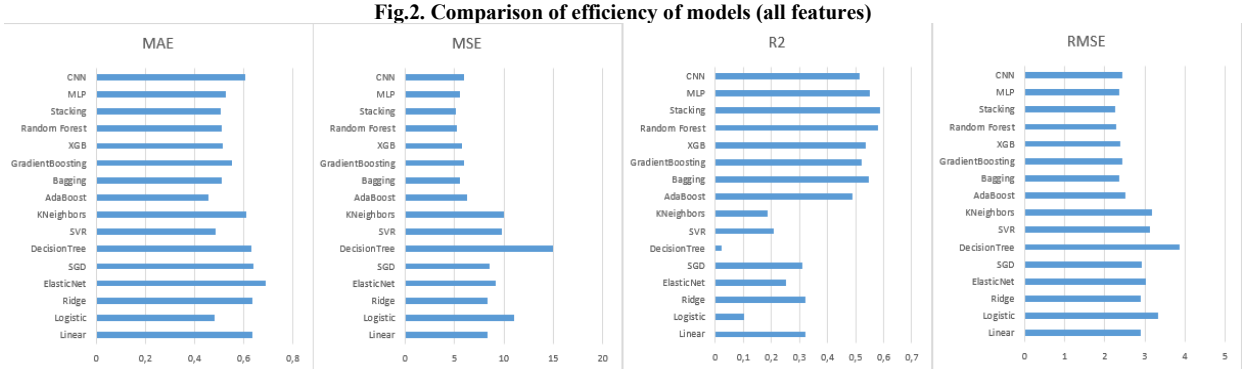

Fig.3. Comparison of efficiency of models (features selected by Boruta algorithm)

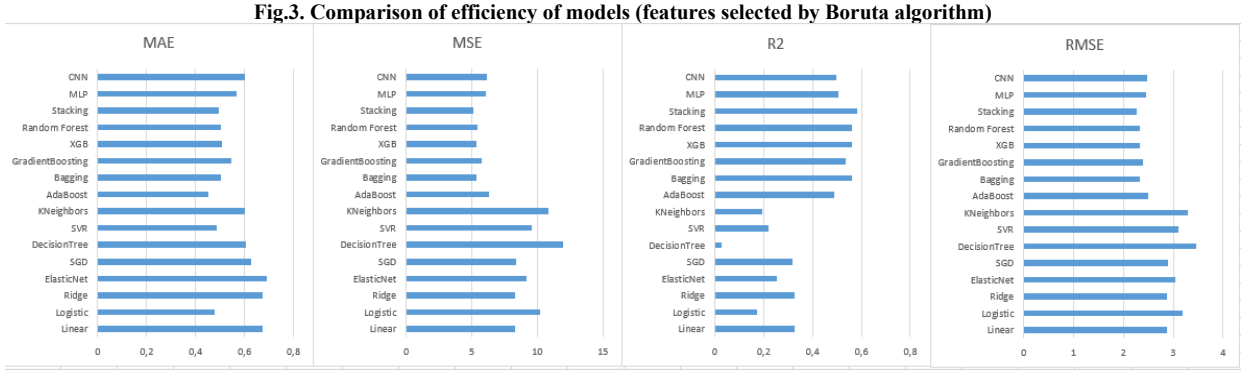

Fig.4. Comparison of efficiency of models (features selected by RFE algorithm)

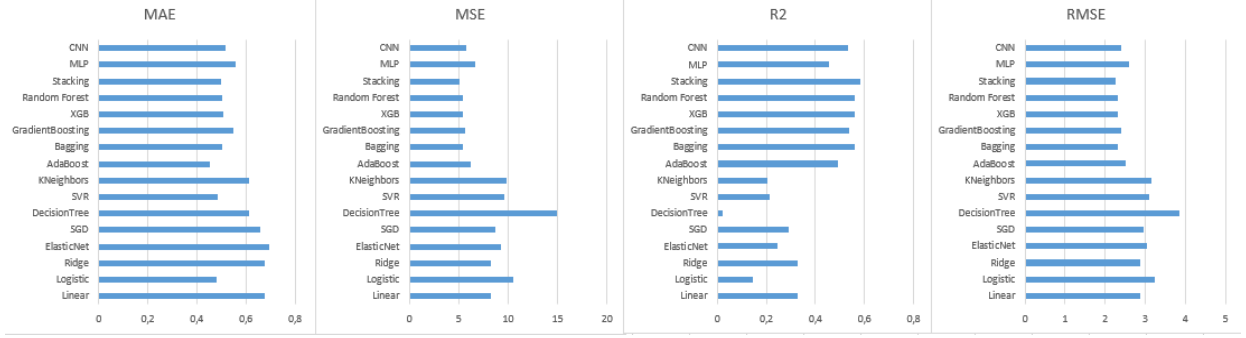

Fig.5. Comparison of efficiency of models (features selected by RFECV algorithm)

Ensemble methods and neural networks give better results compared to classic methods. Developed method improved generalization abilities.

Ensemble methods combine predictions of multiple trained models. The drawback of this approach is that contribution every model makes to ensemble is the same and does not depend on performance of model. The modification of this approach is a weighted average ensemble [42] that weighs contribution of every ensemble member by the expected performance of the model on a holdout dataset. This means that model contribution depends on its performance. This improves average weighted ensemble over average model ensemble.

The main problem related to usage of neural networks is impossibility to select architecture optimal to solve specific task in advance. Selection of suitable configuration is conducted experimentally, such methods as random search, heuristic search, grid search, etc. is often used.

Developed methods for solving the COVID-19 mortality prediction showed significant increase of accuracy compared to existing approaches (decision trees, K-nearest neighbors, Support Vector Machines, linear regression, etc.).

The results are presented both for the entire dataset and selected features, and the results of the metrics in all cases differ slightly.

Conclusions

The subject of this paper is creation of optimal machine learning designed for COVID-19 mortality prediction task, which can be useful for researchers, governments and international organizations to take preventive actions.

The dataset used for study was analyzed, feature selection was conducted, selected models were trained and their efficiency was compared.

Ensemble methods (stacking, bagging and boosting) as well as neural networks were found to be the most efficient. Prediction accuracy may be improved in future studies.

It was discovered that addition of a new predictor can increase the accuracy of prediction, because the output data of the base predictors are input data for the final predictor. In this case, these features are probably correlated, as all basic predictors try to predict the same result.

References

1. Jamshidi E, Asgary A, Tavakoli N, Zali A, Setareh S, Esmaily H, Jamaldini SH, Daaee A, Babajani A, Sendani Kashi MA, Jamshidi M, Jamal Rahi S and Mansouri N (2022) Using Machine Learning to Predict Mortality for COVID-19 Patients on Day 0 in the ICU. Front. Digit. Health 3:681608. doi: 10.3389/fdgth.2021.681608.

2. Moulaei, K., Shanbehzadeh, M., Mohammadi-Taghiabad, Z. et al. Comparing machine learning algorithms for predicting COVID-19 mortality. BMC Med Inform Decis Mak 22, 2 (2022). https://doi.org/10.1186/s12911-021-01742-0.

3. Banoei, M.M., Dinparastisaleh, R., Zadeh, A.V. et al. Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. Crit Care 25, 328 (2021). https://doi.org/10.1186/s13054-021-03749-5.

4. Sankaranarayanan S, Balan J, Walsh JR, Wu Y, Minnich S, Piazza A, Osborne C, Oliver GR, Lesko J, Bates KL, Khezeli K, Block DR, DiGuardo M, Kreuter J, O'Horo JC, Kalantari J, Klee EW, Salama ME, Kipp B, Morice WG, Jenkinson G COVID-19 Mortality Prediction From Deep Learning in a Large Multistate Electronic Health Record and Laboratory Information System Data Set: Algorithm Development and Validation J Med Internet Res 2021;23(9):e30157 doi: 10.2196/30157 PMID: 34449401 PMCID: 8480399.

5. Shanbehzadeh M, Orooji A, Kazemi-Arpanahi H. Comparing of Data Mining Techniques for Predicting in-Hospital Mortality Among Patients with COVID-19. JBE. 2021;7(2):154-173.

6. An, C., Lim, H., Kim, DW. et al. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. Sci Rep 10, 18716 (2020). https://doi.org/10.1038/s41598-020-75767-2.

7. Kuno, T, Sahashi, Y, Kawahito, S, Takahashi, M, Iwagami, M, Egorova, NN. Prediction of in-hospital mortality with machine learning for COVID-19 patients treated with steroid and remdesivir. J Med Virol. 2022; 94: 958- 964. doi:10.1002/jmv.27393.

8. Tabatabaie M, Sarrami A, Didehdar M, et al. (October 14, 2021) Accuracy of Machine Learning Models to Predict Mortality in COVID-19 Infection Using the Clinical and Laboratory Data at the Time of Admission. Cureus 13(10): e18768. doi:10.7759/cureus.18768.

9. Mahdavi M, Choubdar H, Zabeh E, Rieder M, Safavi-Naeini S, et al. (2021) A machine learning based exploration of COVID-19 mortality risk. PLOS ONE 16(7): e0252384. https://doi.org/10.1371/journal.pone.0252384.

10. Thomas Linden, Frank Hanses, Daniel Domingo-Fernández, Lauren Nicole DeLong, Alpha Tom Kodamullil, Jochen Schneider, Maria J.G.T. Vehreschild, Julia Lanznaster, Maria Madeleine Ruethrich, Stefan Borgmann, Martin Hower, Kai Wille, Torsten Feldt, Siegbert Rieg, Bernd Hertenstein, Christoph Wyen, Christoph Roemmele, Jörg Janne Vehreschild, Carolin E.M. Jakob, Melanie Stecher, Maria Kuzikov, Andrea Zaliani, Holger Fröhlich, Machine Learning Based Prediction of COVID-19 Mortality Suggests Repositioning of Anticancer Drug for Treating Severe Cases, Artificial Intelligence in the Life Sciences, Volume 1, 2021, 100020, ISSN 2667-3185, https://doi.org/10.1016/j.ailsci.2021.100020

11. Nooshin Ayoobi, Danial Sharifrazi, Roohallah Alizadehsani, Afshin Shoeibi, Juan M. Gorriz, Hossein Moosaei, Abbas Khosravi, Saeid Nahavandi, Abdoulmohammad Gholamzadeh Chofreh, Feybi Ariani Goni, Jiří Jaromír Klemeš, Amir Mosavi, Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods, Results in Physics, Volume 27, 2021, 104495, ISSN 2211-3797, https://doi.org/10.1016/j.rinp.2021.104495.

12. Laatifi, M., Douzi, S., Bouklouz, A. et al. Machine learning approaches in Covid-19 severity risk prediction in Morocco. J Big Data 9, 5 (2022). https://doi.org/10.1186/s40537-021-00557-0.

13. Peter D Sottile, David Albers, Peter E DeWitt, Seth Russell, J N Stroh, David P Kao, Bonnie Adrian, Matthew E Levine, Ryan Mooney, Lenny Larchick, Jean S Kutner, Matthew K Wynia, Jeffrey J Glasheen, Tellen D Bennett, Real-time electronic health record mortality prediction during the COVID-19 pandemic: a prospective cohort study, Journal of the American Medical Informatics Association, Volume 28, Issue 11, November 2021, Pages 2354–2365, https://doi.org/10.1093/jamia/ocab100.

14. Ottenhoff MC, Ramos LA, Potters W on behalf of The Dutch COVID-PREDICT research group, et alPredicting mortality of individual patients with COVID-19: a multicentre Dutch cohortBMJ Open 2021;11:e047347. doi: 10.1136/bmjopen-2020-047347.

15. Booth, A.L., Abels, E. & McCaffrey, P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. Mod Pathol 34, 522–531 (2021). https://doi.org/10.1038/s41379-020-00700-x.

16. Zhao Z, Chen A, Hou W, Graham JM, Li H, et al. (2020) Prediction model and risk scores of ICU admission and mortality in COVID-19. PLOS ONE 15(7): e0236618. https://doi.org/10.1371/journal.pone.0236618.

17. Yan, L., Zhang, HT., Goncalves, J. et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2, 283–288 (2020). https://doi.org/10.1038/s42256-020-0180-7.

18. Jee, Y., Kim, YJ., Oh, J. et al. A COVID-19 mortality prediction model for Korean patients using nationwide Korean disease control and prevention agency database. Sci Rep 12, 3311 (2022). https://doi.org/10.1038/s41598-022-07051-4.

19. Bertsimas D, Lukin G, Mingardi L, Nohadani O, Orfanoudaki A, et al. (2020) COVID-19 mortality risk assessment: An international multi-center study. PLOS ONE 15(12): e0243262. https://doi.org/10.1371/journal.pone.0243262.

20. Alle S, Kanakan A, Siddiqui S, Garg A, Karthikeyan A, et al. (2022) COVID-19 Risk Stratification and Mortality Prediction in Hospitalized Indian Patients: Harnessing clinical data for public health benefits. PLOS ONE 17(3): e0264785. https://doi.org/10.1371/journal.pone.0264785.

21. Arezoo Haratian, Hadi Fazelinia, Zeinab Maleki, Pouria Ramazi, Hao Wang, Mark A. Lewis, Russell Greiner, David Wishart, Dataset of COVID-19 outbreak and potential predictive features in the USA, Data in Brief, Volume 38, 2021, 107360, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2021.107360.

22. Boruta Feature Selection (an Example in Python) | by Aaron Lee | Towards Data Science. URL: https://towardsdatascience.com/simple-example-using-boruta-feature-selection-in-python-8b96925d5d7a (дата звернення: 29.05.2022).

23. Recursive Feature Elimination (RFE) for Feature Selection in Python. URL: https://machinelearningmastery.com/rfe-feature-selection-in-python/ (дата звернення: 29.05.2022).

24. Recursive feature elimination with cross-validation — scikit-learn 1.1.1 documentation. URL: https://scikit-learn.org/stable/auto_examples/feature_selection/plot_rfe_with_cross_validation.html (дата звернення: 29.05.2022).

25. ML | Linear Regression - GeeksforGeeks. URL: https://www.geeksforgeeks.org/ml-linear-regression/ (дата звернення: 29.05.2022).

26. Logistic Regression — Detailed Overview | by Saishruthi Swaminathan | Towards Data Science. URL: https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc. (дата звернення: 29.05.2022).

27. Ridge Regression Definition & Examples | What is Ridge Regression? URL: https://www.mygreatlearning.com/blog/what-is-ridge-regression/ . (дата звернення: 29.05.2022).

28. How to Develop Elastic Net Regression Models in Python. URL: https://machinelearningmastery.com/elastic-net-regression-in-python/ (дата звернення: 29.05.2022).

29. Stochastic Gradient Descent — Clearly Explained !! | by Aishwarya V Srinivasan | Towards Data Science. URL: https://towardsdatascience.com/stochastic-gradient-descent-clearly-explained-53d239905d31 (дата звернення: 29.05.2022).

30. Machine Learning Basics with the K-Nearest Neighbors Algorithm | by Onel Harrison | Towards Data Science. URL: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761 (дата звернення: 29.05.2022).

31. Support Vector Machine Algorithm - GeeksforGeeks. URL: https://www.geeksforgeeks.org/support-vector-machine-algorithm/ (дата звернення: 29.05.2022).

32. Decision Tree - GeeksforGeeks. URL: https://www.geeksforgeeks.org/decision-tree/ (дата звернення: 29.05.2022).

33. Bagging algorithms in Python | Engineering Education (EngEd) Program | Section. URL: https://www.section.io/engineering-education/implementing-bagging-algorithms-in-python/ (дата звернення: 29.05.2022).

34. Understanding Random Forest. How the Algorithm Works and Why it Is… | by Tony Yiu | Towards Data Science. URL: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (дата звернення: 29.05.2022).

35. A Quick Guide to Boosting in ML. This post will guide you through an… | by Jocelyn D'Souza | GreyAtom | Medium. URL: https://medium.com/greyatom/a-quick-guide-to-boosting-in-ml-acf7c1585cb5 (дата звернення: 29.05.2022).

36. A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. URL: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/ (дата звернення: 29.05.2022).

37. Understanding AdaBoost. Anyone starting to learn Boosting… | by Akash Desarda | Towards Data Science. URL: https://towardsdatascience.com/understanding-adaboost-2f94f22d5bfe (дата звернення: 29.05.2022).

38. A Gentle Introduction to XGBoost for Applied Machine Learning. URL: https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/ (дата звернення: 29.05.2022).

39. Stacking in Machine Learning - GeeksforGeeks. URL: https://www.geeksforgeeks.org/stacking-in-machine-learning/ (дата звернення: 29.05.2022).

40. Multilayer Perceptron Explained with a Real-Life Example and Python Code: Sentiment Analysis | by Carolina Bento | Towards Data Science. URL: https://towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141 (дата звернення: 29.05.2022).

41. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way | by Sumit Saha | Towards Data Science. URL: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 (дата звернення: 29.05.2022).

42. Simple Weighted Average Ensemble | Machine Learning | by Jinhang Jiang | Analytics Vidhya | Medium. URL: https://medium.com/analytics-vidhya/simple-weighted-average-ensemble-machine-learning-777824852426 (дата звернення: 29.05.2022).

https://doi.org/10.31891/csit-2022-2-13
UDC 004.3

Yelyzaveta HNATCHUK, Anita BOIKO, Alina HNATCHUK
Khmelnytskyi National University

RESEARCH OF METHODS OF SEAT DISTRIBUTION IN PUBLIC TRANSPORT

In large cities, especially during a pandemic, the problem of allocating seats in transport and building the most successful route is an urgent task. The study of four methods of distribution of seats in public transport. Their advantages and disadvantages are investigated. According to the results of the research, it became clear that the existing methods alone do not allow to effectively solve the problem of distribution of seats in transport, so it is necessary to either improve existing methods or develop new methods. Improving existing methods is possible by either combining several methods into one or adding certain elements that eliminate existing shortcomings or minimize their impact on the effectiveness of the method. According to the results of the research, the method of electronic seat selection has been improved in terms of adding seat weight sensors, which makes it possible to record the number of occupied seats and helps to provide information to passengers using a mobile application for free seats. The method was also further developed with the help of a mobile application that allows you to pave a route and helps you choose a convenient place to travel to your destination.

Keywords: distribution of seats in transport, infrared sensors, touch sensors, mobile application

Єлизавета ГНАТЧУК, Аніта БОЙКО, Аліна ГНАТЧУК
Хмельницький національний університет

ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПОДІЛУ МІСЦЬ В ГРОМАДСЬКОМУ ТРАНСПОРТІ

У великих містах, особливо під час пандемії, проблема розподілу сидячих місць в транспорті та складання найбільш вдалого маршруту є актуальною задачею. В роботі проведено дослідження чотирьох методів розподілу місць в громадському транспорті. Досліджено їх переваги та недоліки. За результатами проведених досліджень виявилось, що існуючі методи окремо не дозволяють ефективно вирішити задачу розподілу місць в транспорті, тому необхідно або удосконалювати існуючі методи або розробляти нові методи. Удосконалення існуючих методів можливе за рахунок або об'єднання декількох методів в одному або додавання певних елементів, що дозволяють усунути наявні недоліки або звести їх вплив на ефективність методу до мінімуму. За результатами проведених досліджень удосконалено метод електронного вибору місця в частині додавання вагових датчиків сидіння, що дає можливість фіксувати кількість зайнятих місць та допомагає надавати інформацію пасажирам за допомогою мобільного додатку про вільні місця. Також набув подальшого розвитку метод за допомогою мобільного додатку, що дозволяє прокласти маршрут та допомагає обрати зручне місце для поїздки в пункт призначення.

Ключові слова: розподіл місць в транспорті, інфрачервоні датчики, сенсорні датчики, мобільний додаток

Introduction

The main task of the organization of urban public transport is to ensure the quality of traffic, which is assessed mainly by the average waiting time of passengers at the stop or by the average interval of traffic on the route. The quality of passenger traffic is determined, in addition, by the actual duration of the trip, the cost of travel, the speed of arrival and the fullness of the vehicle [1, 2]. It is these indicators that determine the attractiveness of using a particular type of public transport. Uneven distribution of passenger traffic by periods of the day significantly affects the organization of traffic and this, in turn, affects the level of efficiency of public transport.

The efficiency of passenger traffic has been studied by many scientists [1]. The vast majority of studies concerned the analysis of passenger traffic in the city, the probability of passenger traffic in transport interchanges, but the issue of determining the availability of vacancies in transport was paid almost no attention. Therefore, the study of existing methods of allocation of seats in transport and the choice of an effective method of allocation of seats for convenient use of transport and ensuring the arrival of citizens in time to the destination is an urgent task, which this study is devoted to.

Analysis of the peculiarities of passenger traffic in public transport

As an example of public transport, consider bus transport. On weekdays, business trips predominate, so the peak passenger flows fall on the period from 07.30-09.00 and 17.00-19.00. The inter-peak period is characterized by a decrease in the efficiency of vehicle use due to the increase in the intervals of their movement [3]. This increases the waiting time at stops. The flow of people through urban infrastructure has a major impact on several areas, such as tourism and transport. In particular, the ability to accurately count the number of passengers is one of the most important components of the transit service, as it provides a key indicator of the efficiency of public transport companies and is key to effective transit network planning, both long-term and short-term. Indeed, long-term route planning and related schedules make it possible, through the analysis of departure-destination matrices, to obtain information about travel during peak hours [2, 3]. Moreover, such matrices provide guidance on congested travel time and appropriate routes, which affects short-term planning strategies. Thus, long- and short-term planning promotes efficient use of resources and ensures that buses run on the right routes.

To obtain information on the number of passengers, transport campaigns usually use traditional mechanisms, ranging from non-automatic visual counting performed by a person to automatic counting of passengers. Automatic methods based on various data collection technologies (eg matte sensors, infrared sensors, video cameras. These systems need to be installed on vehicles and they are usually quite expensive.

Consider in more detail the following methods: method of electronic location selection, infrared technologies of distribution of places in transport, method of allocating seats in transport using a mobile application and method is based on the detection of the maximum intensity of passenger traffic.

Method of electronic location selection

The essence of this method is that when entering public transport, the passenger must pay with an electronic ticket, select a place on the screen and then go to the selected place [3]. The distribution of seats in public transport through electronic seat selection will help to record the number of occupied seats and warn subsequent passengers about free seats in the vehicle. The method is effective provided moderate passenger traffic. For rush hour, you should improve this method: add weight sensors to the seat. The strain gauges are connected to the information board by means of a strain gauge cable and an analog-to-digital converter to illuminate up-to-date information. To pay for travel on the back of the seat, attach an NFC-sensor (Fig. 1), which will be programmed for non-cash payment, which will be an advantage for passengers.

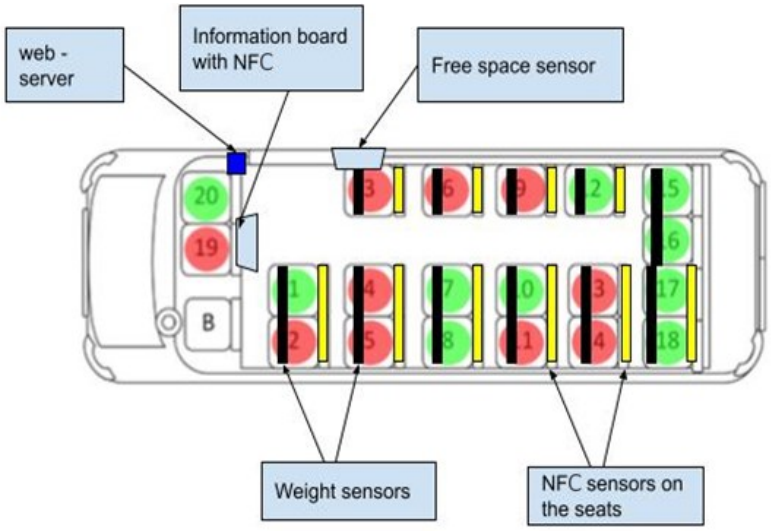

Fig 1. Diagram of a vehicle using the electronic seat selection method with a weight sensor on each seat and an NFC sensor on the back of each seat

Infrared technologies of distribution of places in transport

Cameras with infrared sensors measure the number of passengers with light rays (Fig. 2). When the beam distance is reduced, the occupied position is registered. The sequence in which the rays are broken determines the direction of movement of the passenger. Infrared technology is most common in buses and widespread in retail chains. This method is effective, but during heavy passenger traffic there is a possibility of incorrect data. Therefore, the error of incorrect data is present here. Video imaging technology measures the number of passengers using appropriate cameras in the bus that recognize the passenger. They use several algorithms to:

a) motion detection;

b) assessment of its direction;

c) confirmation of the existence of a moving passenger.

The cameras send a signal through the router to the web-server, where the data on free seats in the vehicle are processed and displayed on an additional board for information to subsequent passengers.

In addition to the task of fixing free passenger seats, infrared sensors can serve as cameras that provide control in the vehicle during heavy passenger traffic [3, 4]. This helps drivers avoid dangers on the road and transport companies improve service.

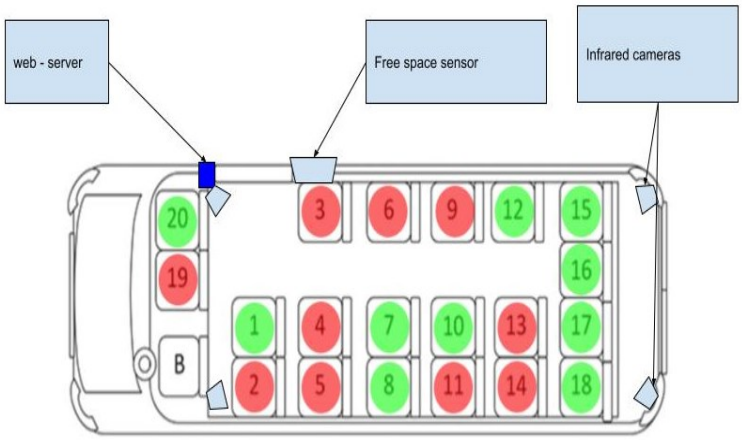

Fig 2. Diagram of a vehicle using the Infrared technologies of distribution of places

Method of allocating seats in transport using a mobile application

This method is based on the method of electronic seat selection and is advanced in terms of adding weight sensors to the seat, which allows you to record the number of occupied seats and helps to provide information to passengers through a mobile application for free seats. Thanks to the installed weight sensors in the bus, passengers can see the number of occupied seats and find out which route to take next. Thanks to the built-in artificial intelligence system, the information is quickly updated and provides only accurate data. When developing a computer system for automatic allocation of seats in transport, it is necessary to take into account that during peak hours there will be a large load on the server. In addition, a potential passenger will be warned that during rush hour seats are quickly occupied.

With the help of the developed application on the smartphone, the passenger can enter the start and end points of the route. A similar system can be seen in Google Maps, which selects the fastest route. In megacities, the function of viewing the employment of minibuses based on passenger feedback is available. But a small percentage of public transport users leave feedback, so Google Maps may not always be relevant. The block diagram of such an appendix is presented in Figure 3.

The application works on the following principle:

1. The passenger chooses a route by number or enters points A and B.
2. The system automatically searches for options for the appropriate route.
3. Check of free places in transport:

a. if there are more than or equal to 3 seats, then a message about the waiting time for the vehicle;

b. if there are less than 3 places, another route is laid and the set conditions are checked again;

c. we repeat cyclically until we find a suitable option.

4. After finding the desired route, the application informs in which part of the vehicle it is better to take a place and the approximate time of arrival at the station.

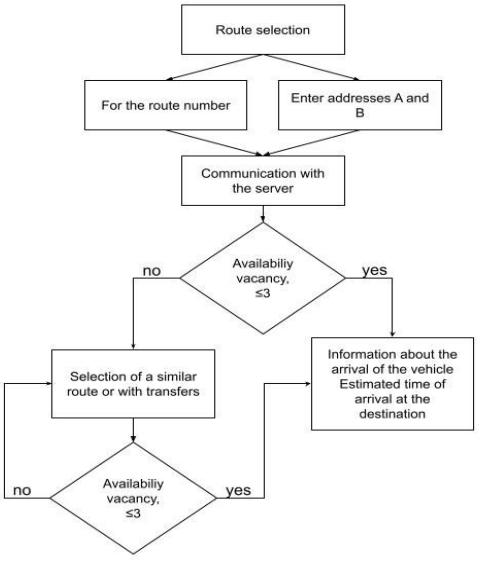

Fig 3. The logic of the application to find the appropriate route

Each bus carrying citizens of the city must have appropriate weight sensors that will record the number of occupied seats. They are placed under the seat. The sensor has measuring limits from 20 kg to 200 kg. A strain gauge is a device that translates mechanical effects into electrical signals and transmits them to a suitable connected device. They come in different types. This work uses a cantilever strain gauge Keli SQB-A 260kg, which when loaded creates resistance, gives the desired signal and fixes it on the screen. The information about the occupied places will be transferred to the server, which is synchronized with the mobile application to provide up-to-date information (Fig. 4).

The characteristics of such a load cell are as follows:

1. 10kg - 250kg.
2. Material: Stainless steel.
3. Protection class: IP68.

Fig 4. Synchronization with servers to update information

In addition to weight sensors, at the entrance to the car is a POS-terminal for payment by bank card (Fig. 4), as well as QR-codes for payment using Internet banking applications, including payment systems Apple Pay, Google Pay, Pay Pass. Adding such payment methods increases the efficiency of drivers, because they are not distracted by payment.

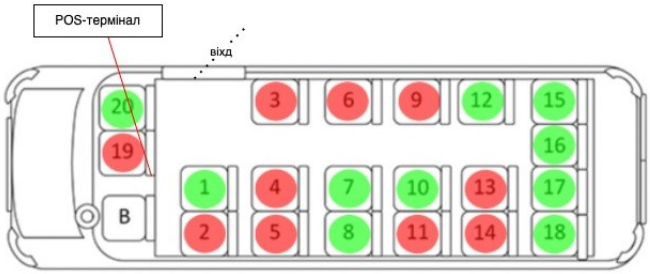

Fig 5. Location of the POS-terminal in the Volkswagen bus

Method is based on the detection of the maximum intensity of passenger traffic

This method of determination is performed by estimating the change in total intensity relative to the X and Y axes and by recording their maxima. This method allowed to observe the trajectory of the object and to estimate the duration of the object's entry into the vehicle [5]. The full projection on the X-axis only helps to find a person relative to the X-axis. The full projection on the Y-axis changes when moving to / from the tire, so the continuity of the Y-axis change is much more important than the continuity of the X-axis change. there are jumps of total intensity (the reason - steel accessories of an entrance ladder, fig. 6).

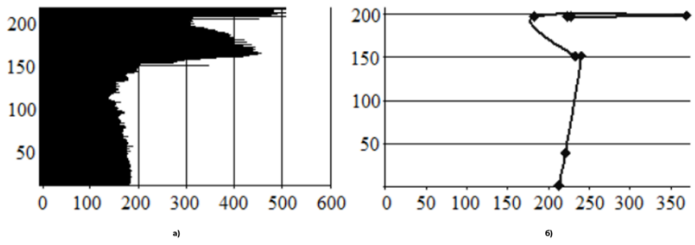

Fig 6. Graph of total intensity: a - graph of the sum of the intensity of the projection Y; b - defined trajectory of one person

Analysis of the trajectory showed that there was an improvement, although due to a number of shortcomings still inaccuracy prevails. The linear moving average filter is represented by formula 1. As the results of the experiment showed, it is rational to choose the filter parameter a to 1/4.

$$y(n) = ax(n) + ax(n-1) + ax(n-2) + ax(n-3), \quad (1)$$

where *a* is the weighting factor,

n is the sum of the intensity of the line number of the projection of the *Y* axis.

This filter allowed to reduce the impact of background noise and correctly indicate the trajectory of passengers (Fig. 7).

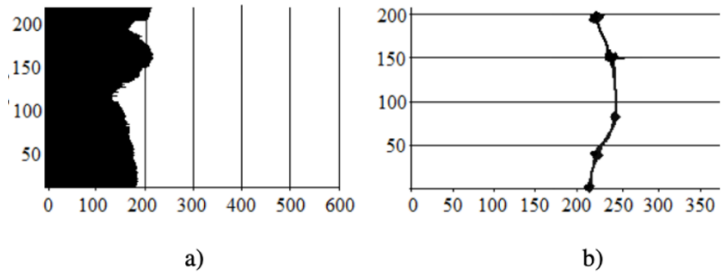

Fig 7. Graph of the total intensity in the projection Y and the obtained trajectory: a - graph of the sum of the intensity of the projection Y; b - defined trajectory of one person

An experiment using this method was conducted on the example of people entering and exiting the vehicle one by one. Thus after qualitative estimation of a method in 70 such situations accuracy of 90% was revealed. The disadvantage of this method is that it does not work for situations where more than one person enters / exits at the same time.

Advantages and disadvantages of the presented methods

Consider the advantages and disadvantages of these methods.

The advantage of the method of electronic selection of the place is the presence of an electric board for quick selection of the place, also thanks to the POS-terminal or the presence of the NFC-sensor it is possible to pay the fare. Having an information screen for the number of free seats is also an advantage, because passengers can see how many next citizens can enter the vehicle.

The biggest disadvantage of this method is that in the presence of a sufficiently large passenger flow, the choice of free space yourself takes a lot of time, which affects the route schedule. Another disadvantage is the cost of all strain gauges. The cost of 1 sensor is 350 UAH. A bus with 20 seats requires at least 12 of them, which amounts to UAH 4,200 per 1 bus, not including two information boards, switches and routers to connect to the server. The cost of installing all sensors for 1 car will be approximately UAH 15,000. If there are 7-10 cars on the route, it will be a big expense for the carrier.

Infrared distribution technologies have the ability to distinguish the thermal radiation of objects in the middle IR range. The recognition results are transmitted to a computer system, where the number of vacancies or occupied is already recorded.

The disadvantage of this method is that there are no ready-made specialized cameras with infrared sensors, they are only to order and are developed according to the scheme of the vehicle, because the routes are buses with different passenger capacity (from 19 to 53). Another disadvantage of this method is the data delay, because the system needs to compare the scheme of the car and the actual data from infrared sensors.

The method of allocating seats in transport using a mobile application, which paves the route and helps to choose the right place to travel to the destination, has the following advantages:

1. Route selection system.
2. Information on available seats in the vehicle.
3. Ability to pay by card using POS-terminals placed in accordance with the number of passenger capacity of the vehicle.
4. If there are no seats available in the bus, the system automatically searches for other similar routes.

The disadvantages of the system are - technical support of the application and updating the version according to the needs and tasks of the system. The application does not work on older versions of Android and iOS. There may be technical issues with the system that are related to the servers. The distribution method does not help to ensure 100% accuracy, for any operating conditions.

Conclusions

In large cities, especially during a pandemic, the problem of allocating seats in transport and building the most successful route is an urgent task. The study of four methods of distribution of seats in public transport. Their advantages and disadvantages are investigated. According to the results of the research, it became clear that the existing methods alone do not allow to effectively solve the problem of distribution of seats in transport, so it is necessary to either improve existing methods or develop new methods. Improving existing methods is possible by either combining several methods into one or adding certain elements that eliminate existing shortcomings or

minimize their impact on the effectiveness of the method. According to the results of the research, the method of electronic seat selection has been improved in terms of adding seat weight sensors, which makes it possible to record the number of occupied seats and helps to provide information to passengers using a mobile application for free seats. The method was also further developed with the help of a mobile application that allows you to pave a route and helps you choose a convenient place to travel to your destination.

References

1. Demchuk I., Ye. Fornalchyk, A. Bilous The Model of Correspondence of Passenger Transportation on the Basis of Fuzzy Logic. Econtechmod : an international quarterly journal on economics in technology, new technologies and modelling processes. Lublin; Rzeszow, 2015. - Volume 04, number 2. P. 59-64.
2. Erik Jenelius Personalized predictive public transport crowding information with automated data sources. Transportation Research Part C: Emerging Technologies. Volume 117. 2020. ISSN 0968-090X. https://doi.org/10.1016/j.trc.2020.102647.I.
3. A. Kara Determination of passenger flows on urban routes using fuzzy logic and transactions of cellular subscribers. 2017. pp.85-94.
4. Giffinger, R.; Christian, F.; Hans, K.; Kalasek, R.; Pichler-Milanovic´, N.; Evert, M. Smart cities: ranking of medium-sized cities in Europe. URL: https://ec.europa.eu/digital-agenda/en/smart-cities
5. Yap, M. D., O. Cats, N. Van Oort, and S. P. Hoogendoorn. Reliable transmission conclusion algorithm for public transport trips during interruptions «Transport research procedure».2017. pp.1042–1049.
6. Xiaochen Liu, Lingshan Li, Xiaohua Liu, Tao Zhang. Analysis of passenger flow and its influences on HVAC systems: An agent based simulation in a Chinese hub airport terminal, Building and Environment.Volume 154. 2019. Pages 55-67. ISSN 0360-1323. https://doi.org/10.1016/j.buildenv.2019.03.011.

Full requirements for the design of the manuscript
Повні вимоги до оформлення рукопису
http://csitjournal.khmnu.edu.ua/

No editorial responsibility is required for the content of messages sub.
За зміст повідомлень редакція відповідальності не несе

To print 30.06.2022. Mind. Printing. Arch. 9,43. Obl.-vid. Arch. 9,04
Format 30x42 / 4, offset paper. Another risography.
Overlay 100, deputy. №

Підп. до друку 30.06.2022. Ум. друк. арк. 9,43. Обл.-вид. арк. 9.04
Формат 30х42/4, папір офсетний. Друк різографією.
Наклад 100, зам. №

Replication is made from the original layout, made edited
by the magazine "Computer Systems and Information Technology"

Тиражування здійснено з оригінал-макету, виготовленого
редакцією журналу "Комп'ютерні системи та інформаційні технології"

Editorial and publishing center of khmelnytsky national university
29016, Khmelnytskyi, street Institutska, 7/1, tel. (0382) 72-83-63

Редакційно-видавничий центр Хмельницького національного університету
29016, м. Хмельницький, вул. Інститутська, 7/1, тел. (0382) 72-83-63