ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

ХАБАРЛАРЫ

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Казахский национальный
университет имени аль-Фараби

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

SERIES
PHYSICO-MATHEMATICAL

3 (343)

JULY – SEPTEMBER 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **No 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: ақпараттық технологиялар
Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*
Мерзімділігі: *жылына 4 рет.*
Тиражы: *300 дана.*
Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

«Известия НАН РК. Физика-математическая».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*
Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **No 16906-Ж** выданное 14.02.2018 г.
Тематическая направленность: *серия информационные коммуникационные технологии.*
В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*
Периодичность: *4 раз в год.*
Тираж: *300 экземпляров.*
Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

News of the National Academy of Sciences of the Republic of Kazakhstan. Physical-matematical series.
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж,** issued 14.02.2018
Thematic scope: *series information technology.*
Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*
Periodicity: *4 times a year.*
Circulation: *300 copies.*
Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022),18-51
https://doi.org/10.32014/2022.2518-1726.138
UDC 621.39:004.05
IRSTI 81.93.29

Zh. Avkurova[1*], S. Gnatyuk[2], B. Abduraimova[3], L. Kydyralina[4]

[1]Karaganda Industrial University NJSC, Kazakhstan, Temirtau;
[2]National Aviation University, Ukraine, Kyiv;
[3]L.N.Gumilyov ENU, Kazakhstan, Astana;
[4]"ShakarimUniversity in Semey" NJSC, Kazakhstan, Semey.
E-mail: *zhadyra.avkurova.83@mail.ru*

MODELS OF STANDARDS AND GOVERNING RULES FOR THE SYSTEMS OF EARLY DETECTION OF APT-ATTACKS AND IDENTIFICATION OF VIOLATORS IN CYBERSPACE

Abstract. The violator of cybersecurity acting on the system changes some of its parameters, initializes or blocks the processes inherent in it. By evaluating these parameters, it is possible to detect the fact of intrusion into the system. On this principle the work of modern systems for the early detection of attacks and the identification of violators is based. In previous works, the authors described the parameters by which the intruder is identified - these are host and network parameters. Since the process of detecting and identifying an intruder takes place under conditions of uncertainty, and a number of certain parameters of systems for early detection of attacks are fuzzy, the functioning of such a system should be based on fuzzy logic.

Thus, in this work, on the basis of the proposed parameters, the linguistic variables were introduced and models of parameter standards were created. Membership functions were calculated for each variable and graphs of their terms were plotted. Also, standards have been formed that are necessary for the development of a system of logical rules to ensure the functioning of the system for early detection of attacks.

The results obtained will be further used to create an IDS / IPS system based on honeypot technology. In addition, examples of rules were developed

to identify the activities of different categories of cybersecurity intruders: disinformer, spammer, cracker, hacker, spam bot and hacker bot. These results can be used to improve existing IDS / IPS system or develop a new security system for early detection of APT-attacks directed on the critical information infrastructure of the state or other important objects.

Key words: intruder identification, cybersecurity, attack detection, linguistic variables, security policy.

Ж.С. Авкурова[1*], С.А. Гнатюк[2], Б.К. Абдураимова[3], Л.М. Кыдыралина[4]

[1]Қарағанды индустриялық университеті, КеАҚ, Қазақстан, Теміртау;
[2]Ұлттық авиациялық университеті, Украина, Киев;
[3]Л.Н.Гумилев атындағы Еуразия ұлттық университеті,
Қазақстан, Астана;
[4]Семей қаласының Шәкәрім атындағы университеті КеАҚ,
Қазақстан, Семей.
E-mail: *zhadyra.avkurova.83@mail.ru*

КИБЕРКЕҢІСТІКТЕГІ АРТ-ШАБУЫЛДАРДЫ ЕРТЕ АНЫҚТАУ ЖӘНЕ БҰЗУШЫЛАРДЫ СӘЙКЕСТЕНДІРУ ҮШІН ЭТАЛОН МОДЕЛЬДЕРІ АНЫҚТАУШЫ ЕРЕЖЕЛЕР

Аннотация. Жүйеде әрекет ететін киберқауіпсіздікті бұзушы оның кейбір параметрлерін өзгертеді, өзіне тән процестерді бастайды немесе блокқа қояды. Осы параметрлерді бағалау арқылы жүйеге ену фактісін анықтауға болады. Дәл осы принцип шабуылдарды ерте анықтаудың және құқық бұзушыларды сәйкестендірудің заманауи жүйелерінің жұмысына негізделген. Алдыңғы жұмыстарда авторлар бұзушыны анықтайтын параметрлерді сипаттады - бұл хост және желі параметрлері. Құқық бұзушыны анықтау және нақтылау процесі белгісіздік жағдайында жүретіндіктен және шабуылдарды ерте анықтау жүйелерінің бірқатар параметрлері анық емес болғандықтан, мұндай жүйенің жұмыс істеуі анық емес логикаға негізделуі керек. Осылайша, бұл жұмыста ұсынылған параметрлер негізінде лингвистикалық айнымалылар енгізіліп, параметрлер стандарттарының модельдері жасалды. Әрбір айнымалы үшін тиістілік функциялары есептелді және олардың графиктері салынды. Сондай-ақ, шабуылдарды ерте анықтау жүйесінің жұмыс істеуін қамтамасыз етуге мүмкіндік беретін логикалық

ережелер жүйесін әзірлеуге қажетті стандарттар қалыптастырылды. Алынған нәтижелер одан әрі honeypot технологиясы негізінде IDS / IPS жүйесін құру үшін пайдаланылатын болады. Сонымен қатар, киберқауіпсіздікті бұзушылардың әртүрлі санаттарының қызметін сәйкестендіру және анықтау үшін ережелер мысалдары әзірленді: теріс ақпарат беруші (дезинформатор), спаммер, крекер, хакер, спам-бот және крекер бот. Бұл нәтижелерді қолданыстағы IDS/IPS жүйелерін жақсарту немесе мемлекеттің және басқа маңызды объектілердің маңызды ақпараттық инфрақұрылымына жасалған APT шабуылдарын ерте анықтауға бағытталған жаңа қауіпсіздік жүйелерін әзірлеу үшін пайдаланылуы мүмкін.

Түйін сөздер: шабуылдаушыны сәйкестендіру, киберқауіпсіздік, шабуылды анықтау, лингвистикалық айнымалылар, қауіпсіздік саясаты.

Ж.С. Авкурова[1*], С.А. Гнатюк[2], Б.К. Абдураимова[3], Л.М. Кыдыралина[4]

[1]НАО Карагандинский индустриальный университет,
Казахстан, Темиртау;
[2]Национальный авиационный университет, Украина, Киев;
[3]ЕНУ им.Л.Н.Гумилева, Казахстан, Астана;
[4]НАО «Университет имени Шакарима города Семей»,
Казахстан, Семей.
E-mail:*zhadyra.avkurova.83@mail.ru*

МОДЕЛИ ЭТАЛОНОВ И ОПРЕДЕЛЯЮЩИЕ ПРАВИЛА ДЛЯ СИСТЕМ РАННЕГО ВЫЯВЛЕНИЯ APT-АТАК И ИДЕНТИФИКАЦИИ НАРУШИТЕЛЕЙ В КИБЕРПРОСТРАНСТВЕ

Аннотация. Нарушитель кибербезопасности, действуя на систему, меняет некоторые ее параметры, инициализирует или блокирует присущие ей процессы. Оценивая эти параметры, можно провести выявления факта вторжения нарушителя в систему. Именно на таком принципе основывается работа большинства существующих современных систем раннего выявления атак и идентификации нарушителей. В предыдущих работах авторами были описаны параметры, по которым осуществляется идентификация нарушителя – это хостовые и сетевые параметры. Поскольку процесс выявления и

идентификации нарушителя происходит в условиях неопределенности, а ряд определенных параметров систем раннего выявления атак носят нечеткий характер, то функционирование такой системы должно основываться на нечеткой логике.

Таким образом, в этой работе на основе предложенных параметров, были введены лингвистические переменные и построены модели эталонов параметров. Для каждой переменной были рассчитаны функции принадлежности и построены графики их термов. Также сформированы стандарты, необходимые для разработки системы логических правил, позволяющих обеспечить функционирование системы раннего выявления атак.

Полученные результаты в дальнейшем могут быть использованы для построения IDS / IPS систем на базе технологии honeypot. Кроме того, авторами были разработаны примеры правил для выявления и идентификации деятельности разных категорий нарушителей кибербезопасности: дезинформатора, спаммера, крекера, хакера, спам-бота и бота взломщика. Эти результаты могут использоваться для усовершенствования существующих IDS / IPS систем или для разработки новых систем безопасности, которые направлены на ранее выявление APT-атак на объекты критической информационной инфраструктуры государства и других важных объектов.

Ключевые слова: идентификации нарушителей, кибербезопасность, выявления атак, лингвистические переменные, политика безопасности.

Introduction. During the attack, the intruder, acting on the system, changes some of its parameters, creates or stops its inherent processes, and the like. All these actions affect the state of the system. By evaluating these parameters, it is possible to detect the fact of intrusion into the system. On this principle the work of modern systems for early detection of attacks (SEDA) and identification of violators is based (Avkurova, et all, 2020). This is how the NIDES system audits such processes as logging in, working with files and processes, administration and fixing errors and failures. The work (Iashvili, et all, 2021) describes the parameters by which the intruder is identified by the developed system. These include (host parameters):

- Username when logging in, UID;
- Login time, Tlog;
- Frequency of login requests, Nlog;
- Time taken to login, TSlog;
- Intensity of actions, I;
- CPU time/CPU load, CPU;

- The amount of loaded RAM, Muse;
- Number of executable files, NEF;
- The type of files used in the attack, AtEF;
- Number of failures and errors, NEr;
- Process/file execution time, RTPr/F;
- Unusual processes, UPr;
- File transfer to the system, TrFin;
- Modifying files, ModF,
- Copy/transfer files from the system, TrFout;
- Keyboard keystrokes, KS;
- Characteristics of ARP-, IP-, ICMP- and TCP- packets (network parameters).

Since the process of detecting and identifying an intruder takes place under conditions of uncertainty, and a number of the given parameters of the SEDA are fuzzy, the operation of such a system should be based on fuzzy logic.

Analysis of modern research and problem statement. In order to identify the intruder, we can use the logical-linguistic approach and the basic model of parameters, partially described in (Iashvili, et all, 2021), which will be the basis for creating of the developed SEDA. For example, to detect port scans, papers (Zhang, et all, 2020) use linguistic variables (LV) «Number of virtual channels» and «Age of virtual channels», and in paper (Zuzčák, et all, 2019) LV «Number of simultaneous connections», «Request processing speed», «Delay between requests» and «Number of packets with the same source and destination address» - to detect DDOS attacks and spoofing.

The process of detecting and identifying an intruder requires determining the necessary parameters and their properties. In this regard, the main goal of this work is to create models of parameter standards and a system of defining rules (DR) necessary for the effective functioning of SEDA and the identification of violators in a poorly formalized environment.

Materials and methods: Rationale for the approach. Let us consider the method of linguistic terms using statistical data (MLTS), where as a measure of membership of an element in a set, an estimate of the frequency of using a concept, which is given by a fuzzy set to characterize the element, is taken. To do this, the value of LV X={x1, x2, . . ., xn} is placed on the universal scale [0,1]. The method is based on the condition that the same number of experiments fall into each interval of the scale, but in practice this is usually not observed. In real conditions, there is an empirical table in which experiments can be unevenly distributed over intervals. Some of them

may not be involved at all, then the data is processed using a hint matrix. Let it be necessary to estimate in LV values the deviations of the parameter Δ BO [0, B] (B - the maximum possible deviation), which characterizes the current measurements. Next, for n = 5, we determine the value of the LV {x1, x2, x3, x4, x5}. The interval [0, B] and B/B (estimated ratio) are divided into k segments (for example, 5), according to which statistics are collected that characterize the frequency of using the LV value by an expert to display his conclusions. Further, the data is entered into the table and processed in such a way as to reduce the errors introduced during the experiment: individual elements are removed from the table, on the left side and on the right side of which there are zeros in the line. The hint matrix is a line whose elements are calculated by the formula:

$$k_j = \sum_{i=1}^{n} b_{ij} = \sum_{i=1}^{5} b_{ij},\ j = \overline{1,5}. \qquad (5)$$

Next, in the resulting row of the matrix, the maximum element $k_{\max} = \max k_j$ is selected, and then all elements of the table are converted by the expression

$$c_{ij} = b_{ij} k_{max} / k_j,\ i = \overline{1,5};\ j = \overline{1,5}, \qquad (6)$$

and for columns where $k_j = 0$ linear approximation is applied $c_{ij} = (c_{ij-1} + c_{ij+1})/2,\ i = \overline{1,5};\ j = \overline{1,5}$. Next, the value of the LV is calculated by the formula

$$\mu_{ij} = c_{ij} / c_{imax},$$

where $c_{imax} = \max\limits_{j} c_{ij},\ i = \overline{1,5};\ j = \overline{1,5}$. (7)

The described method uses data from statistical studies. Their processing is quite laborious, since to construct the FV of one term, it is necessary to carry out statistical studies of all terms of the LP (Gizun et all, 2013).

Let us create a model of standards of linguistic variables for fuzzy parameters of intruder identification from the set of parameters defined in (Iashvili, et all, 2021).

Login time, Tlog. This parameter is based on the fact that the activity of IS and users of these systems depends on the time of day. Usually, a lot of user activity when logging into the system is manifested in the daytime, less - at night, but other statistics are possible, which is determined by the mode of operation of the organization, which includes IS. The nature of this parameter is unclear, because it is impossible to unequivocally draw a conclusion about the illegal activity of the violator. So in organizations with

working hours from 08.00 to 16.00, the probability that the user who logged in is the intruder is the lowest at 08.00 and increases over time, reaching a maximum in the hours after 16.00. However, it should be noted that in the concept of honeypot technologies, this parameter loses its weight somewhat, since any activity on them is considered malicious.

Let's evaluate the LV «The level of legitimacy over time». Let's define the value of the linguistic variable $\{x_1, x_2, x_3\}$ corresponding to {legitimate, suspicious, illegitimate}. That is $T_{T\log} = \bigcup_{i=1} T^i_{T\log}$ = {legitimate, suspicious, illegitimate}

We use statistics on B = 24 hours. It is advisable to divide the total interval into 4 intervals [00:00;06:00], [06:00;12:00], [12:00;18:00], [18:00;24:00].

Data for LV Tlog Table 1

LV value	Interval			
	№1	№2	№3	№4
High	0	8	6	1
Medium	2	1	2	3
Low	6	1	1	4

Using expression (5), we determine $k_j = \|8\ 10\ 9\ 8\|$, where $k_{max} = 10$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 0 & 8 & 6{,}66 & 1{,}25 \\ 2{,}5 & 1 & 2{,}22 & 3{,}75 \\ 7{,}5 & 1 & 1{,}11 & 5 \end{Vmatrix}.$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 0 & 1 & 0{,}83 & 0{,}16 \\ 0{,}66 & 0{,}26 & 0{,}59 & 1 \\ 1 & 0{,}13 & 0{,}15 & 0{,}66 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1} \Delta B_i / B =$ $= \{0{,}25;\ 0{,}5;\ 0{,}75;$ and get the following fuzzy numbers:

$$Л = \{0/0{,}25;\ 1/0{,}5;\ 0{,}83/0{,}75;\ 0{,}16/1\},$$
$$П = \{0{,}66/0{,}25;\ 0{,}26/0{,}5;\ 0{,}59/0{,}75;\ 1/1\},$$
$$Н = \{1/0{,}25;\ 0{,}13/0{,}5;\ 0{,}15/0{,}75;\ 0{,}66/1\}.$$

Graph of FP for the LP terms. The login time is shown on Fig. 1.

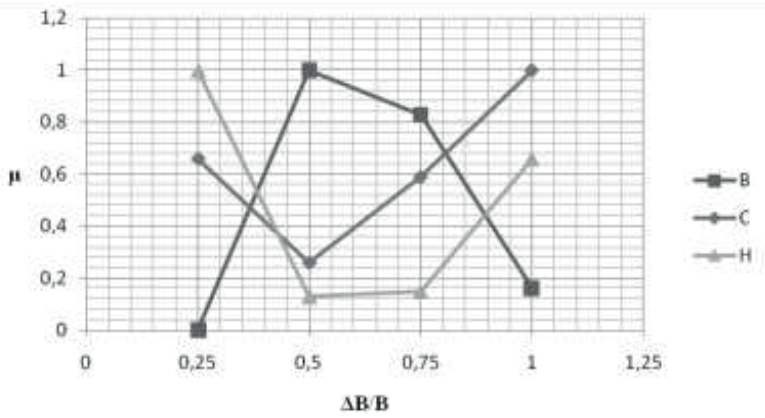

Fig. 1. Linguistic patterns of fuzzy numbers for Tlog

The frequency of login requests, Nlog. It is clear that a high frequency of login requests will be noted when the system is attacked by bots (in particular by hacking bots, since spammers do not require a login). A human attacker is also characterized by an increased frequency of requests as a result of an attempt to bypass the protection and the theoretical assumption that he does not have a legitimate login and password, so he will be forced to make at least several attempts. Moreover, the greater the number of attempts, the more likely it is that the intruder is really trying to enter the IS. It is clear that this parameter is also fuzzy.

Let's estimate the LV «Frequency of requests to enter the system.» Let's define the value of the linguistic variable {x1, x2, x3, x4, x5} corresponding to {low, below average, average, above average, high}. That is

$$T_{N\log} = \bigcup_{i=1}^{5} T_{N\log}^{i} = \{low, belowaverage, average, aboveaverage, high\}$$

The frequency of login requests for an ordinary user is usually minimal (more often, a legitimate user enters a login and password once), and modern password guessing programs are able to take 5310986 passwords /s (Golub et all, 2009). However, to determine the terms of this LV, it will be sufficient to limit ourselves to the value B = 100 requests/s, because a person is not able to go through the authentication procedure more than 10-15 times per minute. It is advisable to divide the general interval into 5 intervals [0, 1], [1, 10], [10, 40], [40, 80], [80, 100]

Data for LV Nlog Table 2

LV value	Interval				
	№1	№2	№3	№4	№5
Low	8	0	0	0	0
Below average	5	2	0	0	0
Average	1	6	4	0	0
Above average	0	2	8	1	0
High	0	0	1	6	6

Using expression (5), we determine $k_j = \|14 \quad 10 \quad 13 \quad 7 \quad 6\|$, where $k_{max} = 14$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 8 & 0 & 0 & 0 & 0 \\ 5 & 2{,}8 & 0 & 0 & 0 \\ 1 & 8{,}4 & 4{,}31 & 0 & 0 \\ 0 & 2{,}8 & 8{,}62 & 2 & 0 \\ 0 & 0 & 1{,}08 & 12 & 16 \end{Vmatrix}.$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0{,}56 & 0 & 0 & 0 \\ 0{,}12 & 1 & 0{,}51 & 0 & 0 \\ 0 & 0{,}32 & 1 & 0{,}23 & 0 \\ 0 & 0 & 0{,}07 & 0{,}75 & 1 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B = \{0{,}01;\ 0{,}1;\ 0{,}4;\ 0{,}8;\ 1\}$ and get the following fuzzy numbers:

$$H = \{1/0{,}01;\ 0/0{,}1;\ 0/0{,}4;\ 0/0{,}8;\ 0/1\},$$
$$HC = \{1/0{,}01;\ 0{,}56/0{,}1;\ 0/0{,}4;\ 0/0{,}8;\ 0/1\},$$
$$C = \{0{,}12/0{,}01;\ 1/0{,}1;\ 0{,}51/0{,}4;\ 0/0{,}8;\ 0/1\},$$
$$BC = \{0/0{,}01;\ 0{,}32/0{,}1;\ 1/0{,}4;\ 0{,}23/0{,}8;\ 0/1\},$$
$$B = \{0/0{,}01;\ 0/0{,}1;\ 0{,}07/0{,}4;\ 0{,}75/0{,}8;\ 1/1\}.$$

Graph of FP for the LP terms. The frequency of login requests is shown on Fig. 2

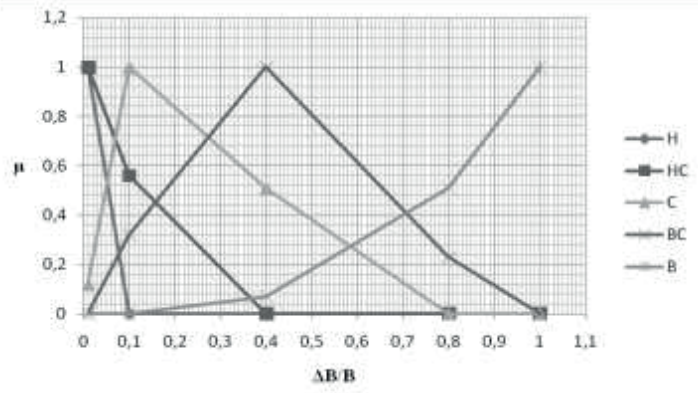

Fig. 2. Linguistic patterns of fuzzy numbers for Nlog

Time taken to login, TSlog. A parameter that is closely related to the previous one. The time spent by the intruder is in most cases greater than the time spent by the legitimate user. But it is fuzzy, because it does not allow an unambiguous identification.

Let's estimate the LV «Time spent on logging in». Let's define the value of the linguistic variable {x1, x2, x3, x4, x5} corresponding to {very small, small, medium, large, very large}. That is

$$T_{S\log} = \bigcup_{i=1}^{5} T_{S\log}^{i} = \{verysmall, small, medium, l\arg e, veryl\arg e\}$$

A legitimate user in a password-protected IS spends from several seconds to several minutes on identification. But the time spent by illegitimate users to break the password system is relatively large. So modern password guessing systems for cracking an 8-character password, consisting of a combination of letters, numbers and special characters, spend up to 61 days (Niu, et all, 2017). Therefore, we take the value of B = 60 days = 5184000 s. It is advisable to divide the total interval into 5 intervals [0 s; 30 s], [30 s; 5 min], [5 min; 1 h], [1 h; 1 day], [1 day; 60 days].

Data for LV TSlog Table 3

LV value	Interval				
	№1	№2	№3	№4	№5
Very small	9	3	0	0	0
Small	5	10	1	0	0
Medium	1	7	5	0	0
Large	0	1	2	9	2
Very large	0	0	1	6	9

Using expression (5), we determine $k_j = \|14 \quad 21 \quad 9 \quad 15 \quad 11\|$, where $k_{max} = 21$ and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 13,5 & 3 & 0 & 0 & 0 \\ 7,5 & 10 & 2,33 & 0 & 0 \\ 1,5 & 7 & 11,67 & 0 & 0 \\ 0 & 1 & 4,67 & 12,6 & 3,82 \\ 0 & 0 & 2,33 & 8,4 & 17,18 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,22 & 0 & 0 & 0 \\ 0,75 & 1 & 0,23 & 0 & 0 \\ 0,13 & 0,6 & 1 & 0 & 0 \\ 0 & 0,08 & 0,37 & 1 & 0,3 \\ 0 & 0 & 0,14 & 0,49 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B = \{5,79 \cdot 10^{-6};\ 5,79 \cdot 10^{-5};\ 6,94 \cdot 10^{-4};\ 0,02;\ 1\}$ and get the following fuzzy numbers:

$$ОМ = \{1/5,79 \cdot 10^{-6};\ 0,22/5,79 \cdot 10^{-5};\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0/1\},$$
$$М = \{0,75/5,79 \cdot 10^{-6};\ 1/5,79 \cdot 10^{-5};\ 0,23/6,94 \cdot 10^{-4};\ 0/0,02;\ 0/1\},$$
$$С = \{0,13/5,79 \cdot 10^{-6};\ 0,6/5,79 \cdot 10^{-5};\ 1/6,94 \cdot 10^{-4};\ 0/0,02;\ 0/1\},$$
$$Б = \{0/5,79 \cdot 10^{-6};\ 0,08/5,79 \cdot 10^{-5};\ 0,37/6,94 \cdot 10^{-4};\ 1/0,02;\ 0,3/1\},$$
$$ОБ = \{0/5,79 \cdot 10^{-6};\ 0/5,79 \cdot 10^{-5};\ 0,14/6,94 \cdot 10^{-4};\ 0,49/0,02;\ 1/1\}.$$

The graph of the FP for the LP terms is shown on Fig.3.

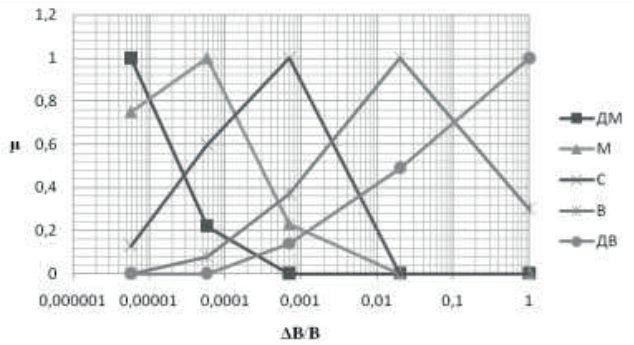

Fig. 3. Linguistic patterns of fuzzy numbers for TSlog

Intensity of actions, I. This refers to the number of any user actions, including logging in/out of the system, transferring, changing, copying files, starting/stopping processes, etc., per unit of time. The intensity may not differ for a human intruder and for a legitimate user, however, for bots it is much higher, therefore, it is the most significant for identifying and distinguishing between human-robot categories. Although a significant excess of the norm indicates the activity of unauthorized automatic intruder systems (bots), however, I is a fuzzy parameter, since it is very difficult to determine the normal value of the intensity indicator.

Let's evaluate the LV «Intensity of actions». Let's determine the value of the linguistic variable {x1, x2, x3} corresponding to {low, medium, high}.

That is $T_I = \bigcup_{i=1}^{i} T_I^i = \{low, medium, high\}$

The intensity of the actions of an ordinary person is very low, usually from 3 to 10 actions per minute. The intensity of the bots is ten times greater. For research, we will take the upper limit of 100 actions/min, although this figure may be higher for robots. It is advisable to divide the total interval into 4 intervals [0; 5], [5, 10], [10, 50], [50, 100].

Data for LV I Table 4

LV value	Interval			
	№1	№2	№3	№4
Low	7	5	1	0
Medium	0	7	4	0
High	0	1	5	7

Using expression (5), we determine $k_j = \|7\ 13\ 10\ 7\|$, where $k_{max} = 13$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 13 & 5 & 1,3 & 0 \\ 0 & 7 & 5,2 & 0 \\ 0 & 1 & 6,5 & 13 \end{Vmatrix}.$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,38 & 0,1 & 0 \\ 0 & 1 & 0,74 & 0 \\ 0 & 0,08 & 0,5 & 1 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,05;\ 0,1;\ 0,5;\ 1\}$ and get the following fuzzy numbers:

$$H = \{1/0,05;\ 0,38/0,1;\ 0,1/0,5;\ 0/1\},$$
$$C = \{0/0,05;\ 1/0,1;\ 0,74/0,5;\ 0/1\},$$
$$B = \{0/0,05;\ 0,08/0,1;\ 0,5/0,5;\ 1/1\}.$$

The graph of the FP for the LP terms. The intensity of actions is shown on Fig.4.

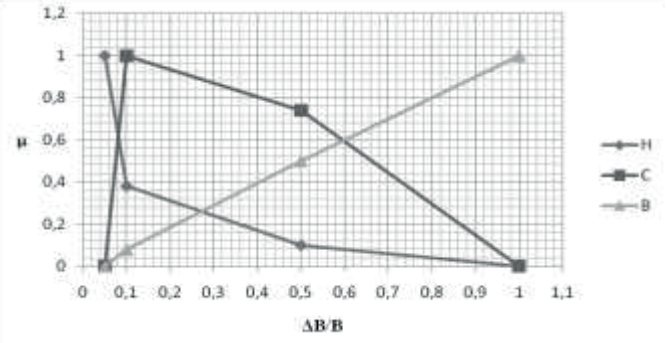

Fig. 4. Linguistic patterns of fuzzy numbers for I

Processor time / processor load, CPU. Since the number of active processes on honeypot systems should be kept to a minimum, any increase in load is a sign of intruder activity on the system. In real ISs, the probability

that the activity is caused by the intruder is somewhat lower, and, of course, the normal amount of processor time is higher. However, this parameter can still be effectively used to identify the fact of a violation in intrusion detection systems and SEDA. Since it is impossible to give an unambiguous answer about the intruder for this parameter, primarily due to the possible activity of viruses, the processor CPU is a fuzzy parameter.

Let's estimate the LV «Processor time/processor load». Let's determine the value of the linguistic variable {x1, x2, x3} corresponding to {low, medium, high}.

That is $T_{CPU} = \bigcup_{i=1}^{3} T^{i}_{CPU} = \{low, medium, high\}$

Under normal conditions of average use of PC capacities and in the absence of intruders or malware, the average processor load is 20-30%. Of course, the norm may vary somewhat depending on the OS, installed software and the production tasks of the organization. The maximum possible percentage of CPU load is B = 100%. It is advisable to divide the total interval into 4 intervals [0; 20], [20, 50], [50; 75], [50, 100].

Data for LV CPU Table 5

LV value	Interval			
	№1	№2	№3	№4
Low	9	6	0	0
Medium	3	8	1	0
High	0	1	5	8

Using expression (5), we determine $k_j = \|12\ 15\ 6\ 8\|$, where $k_{max} = 15$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 11,25 & 6 & 0 & 0 \\ 3,75 & 8 & 2,5 & 0 \\ 0 & 1 & 12,5 & 15 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,53 & 0 & 0 \\ 0,47 & 1 & 0,31 & 0 \\ 0 & 0,07 & 0,83 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,2;\ 0,5;\ 0,75;\ 1\}$, and get the following fuzzy numbers:

$$H = \{1/0,2;\ 0,53/0,5;\ 0/0,75;\ 0/1\},$$
$$C = \{0,47/0,2;\ 1/0,5;\ 0,31/0,75;\ 0/1\},$$
$$B = \{0/0,2;\ 0,07/0,5;\ 0,83/0,75;\ 1/1\}.$$

The graph of the FP for the LP terms. Processor time / processor load is shown on Fig.5.

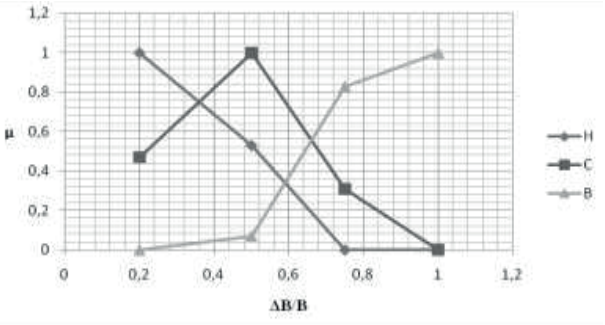

Fig. 5. Linguistic patterns of fuzzy numbers for CPU

The amount of loaded RAM, Muse. Similar in essence to the previous one and is also fuzzy. The FV for this parameter is almost identical to the LV «Processor time/processor load».

Number of executable files/processes, NEF. Also included in the group of fuzzy parameters. The fact of an attacker's actions on this parameter is determined by a deviation from the norm. This parameter includes only user processes and files, and system files are not taken into account. So in each organization, in accordance with the security policy and job responsibilities, each legitimate user can use certain files at a certain moment, and the simultaneous use of several files or processes at once is practically excluded. This allows to identify both external and internal Intruders, but with a certain probability.

Let's estimate the LV «Number of executable files». Let's define the value of the linguistic variable {x1, x2, x3} corresponding to {very small, small, normal, large, very large}.

That is $T_{NEF} = \bigcup_{i=1}^{5} T^{i}_{NEF} = \{\text{very small, small, normal, large, very large}\}$

The normal number of executable files, as already noted, is very dependent on the industry in which a particular IS operates, and on the set of security policy rules and job descriptions of the organization. Basically, this value ranges from 7 to 10 processes. Considering also the technological limitations of IS and typical security policies, we will take the maximum value of 20 user processes for research. It is advisable to divide the total interval into 4 intervals [0; 4], [4, 8], [8, 12], [12, 16], [16; 20].

Data for LV NEF Table 6

LV value	Interval				
	№1	№2	№3	№4	№5
Very small	8	5	1	0	0
Small	4	7	2	0	0
Normal	1	4	9	3	1
Large	0	0	4	8	2
Very large	0	0	1	5	8

Using expression (5), we determine $k_j = \|13 \quad 16 \quad 17 \quad 16 \quad 11\|$, where $k_{max} = 17$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 10,46 & 5,31 & 1 & 0 & 0 \\ 5,23 & 7,44 & 2 & 0 & 0 \\ 1,31 & 4,25 & 9 & 3,19 & 1,55 \\ 0 & 0 & 4 & 8,5 & 3,09 \\ 0 & 0 & 1 & 5,31 & 12,36 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,51 & 0,1 & 0 & 0 \\ 0,7 & 1 & 0,27 & 0 & 0 \\ 0,15 & 0,47 & 1 & 0,35 & 0,17 \\ 0 & 0 & 0,47 & 1 & 0,36 \\ 0 & 0 & 0,08 & 0,43 & 1 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B = \{0,2;\ 0,4;\ 0,6;\ 0,8;\ 1\}$ and get the following fuzzy numbers:

$$ОМ = \{1/0,2;\ 0,51/0,4;\ 0,1/0,6;\ 0/0,8;\ 0/1\},$$
$$М = \{0,7/0,2;\ 1/0,4;\ 0,27/0,6;\ 0/0,8;\ 0/1\},$$
$$Н = \{0,15/0,2;\ 0,47/0,4;\ 1/0,6;\ 0,35/0,8;\ 0,17/1\},$$
$$Б = \{0/0,2;\ 0/0,4;\ 0,47/0,6;\ 1/0,8;\ 0,36/1\},$$
$$ОБ = \{0/0,2;\ 0,0/0,4;\ 0,08/0,6;\ 0,43/0,8;\ 1/1\}.$$

The graph of the FP for the LP terms. The number of executable files/ processes is shown on Fig.6.

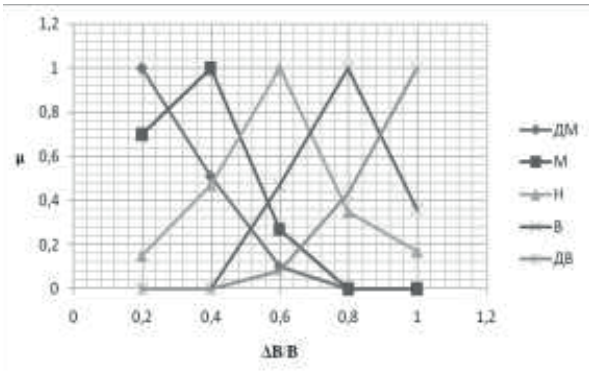

Fig. 6. Linguistic patterns of fuzzy numbers for NEF

Number of failures and errors, NEr. This parameter is fuzzy, since failures and errors can occur when working in the authorized user mode and intruder. However, with frequent repeated failures or errors, it can be concluded with a certain degree of probability that the system has been attacked. This group includes a wide range of events from authorization errors to failures when executing certain processes or files. With the active work of the intruder, regardless of his class and category, the frequency of occurrence of malfunctions will be somewhat higher. It should also be noted that it is quite possible when identifying an intruder-work, this frequency will be even higher.

Let's estimate the LV «Number of failures and errors». Let's determine the value of the linguistic variable {x1, x2, x3} corresponding to {low, medium, high}.

That is $T_{NEr} = \bigcup_{i=1}^{3} T^{i}_{NEr} = \{\text{low, medium, high}\}$

The functionality of an IS is considered normal if there are no errors or failures in its operation at all. However, the occurrence of a small number

of failures is still possible, mainly due to insufficient user qualifications, his negligence or the use of low-quality hardware / software. There are no official statistics on this issue, so it is difficult to determine the normal value of this parameter. As part of the study, we set the maximum number of errors and failures per day B = 10. It is advisable to divide the total interval into 4 intervals [0, 1], [1, 4], [4, 8], [8, 10].

Data for LV NEr Table 7

LV value	Интервал			
	№1	№2	№3	№4
Low	5	1	0	0
Medium	0	4	3	0
High	0	0	1	6

Using expression (5), we determine $k_j = \|5\ 5\ 4\ 6\|$, where $k_{max} = 6$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 6 & 1,2 & 0 & 0 \\ 0 & 4,8 & 4,5 & 0 \\ 0 & 0 & 1,5 & 6 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,2 & 0 & 0 \\ 0 & 1 & 0,94 & 0 \\ 0 & 0 & 0,25 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,1;\ 0,4;\ 0,8;\ 1\}$ and get the following fuzzy numbers:

$$H = \{1/0,1;\ 0,2/0,4;\ 0/0,8;\ 0/1\},$$
$$C = \{0/0,1;\ 1/0,4;\ 0,94/0,8;\ 0/1\},$$
$$B = \{0/0,1;\ 0/0,4;\ 0,25/0,8;\ 1/1\}.$$

The graph of the FP for the LP terms. The number of failures and errors is shown on Fig.7.

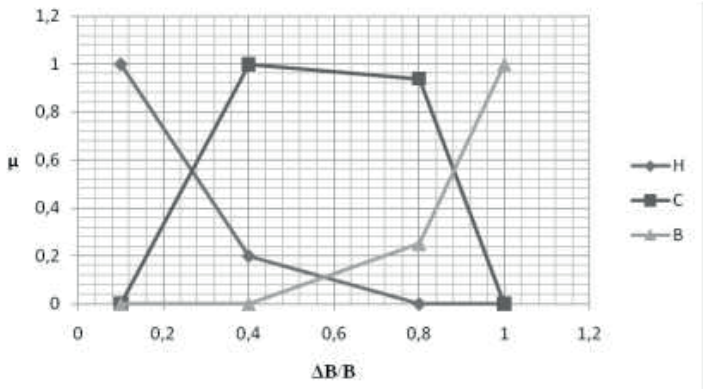

Fig. 7. Linguistic patterns of fuzzy numbers for NEr

Process/file execution time, RTPr/F. Examining the statistics of the work of IS of various enterprises and organizations, it is easy to notice that, depending on the specifics of the work, the time spent on performing a certain operation is approximately the same for the same type of IS and their tasks. Honeypot systems mainly run system processes, that support the operation of the honeypot itself, or administrator processes that run at a certain time for a certain period. Thus, when identifying such processes, it can be concluded that the system is attacked by an intruder. Since this can be caused by the negligence of the employee, the conclusion is ambiguous and, accordingly, the parameter is fuzzy.

Let's estimate the LV «Process/file execution time». Let's define the value of the linguistic variable {x1, x2, x3, x4, x5} corresponding to {very small, small, medium, large, very large}.

That $T_{RT\,\mathrm{Pr}/F} = \bigcup_{i=1}^{5} T^{i}_{RT\,\mathrm{Pr}/F} == \{\text{very small, small, medium, large, very large}\}$

A legitimate user in IS, in the course of performing his job duties, works with a certain file or process in the course of a certain time. So the average worker works with one file or process for a period of time from 30 minutes to 3 hours. If this indicator is significantly less or more, then this may indicate suspicious activity. It is advisable to set the maximum value of this variable B = 24 hours, the total interval is divided into 5 intervals [0 s; 1 min], [1 min; 30 min], [30 min 3 h], [3 h; 6 h], [6 h; 24h].

Data for LV RTPr/F Table 8

LV value	Interval				
	№1	№2	№3	№4	№5
Very small	9	4	1	0	0

Small	5	7	2	0	0
Medium	0	3	8	3	0
Large	0	0	3	9	6
Very large	0	0	1	4	9

Using expression (5), we determine $k_j = \|14 \quad 14 \quad 15 \quad 16 \quad 15\|$, where $k_{max} = 16$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 10,29 & 4,57 & 1,07 & 0 & 0 \\ 5,71 & 8 & 2,13 & 0 & 0 \\ 0 & 3,43 & 8,53 & 3 & 0 \\ 0 & 0 & 3,2 & 9 & 6,4 \\ 0 & 0 & 1,07 & 4 & 9,6 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,44 & 0,1 & 0 & 0 \\ 0,71 & 1 & 0,27 & 0 & 0 \\ 0 & 0,4 & 1 & 0,35 & 0 \\ 0 & 0 & 0,36 & 1 & 0,71 \\ 0 & 0 & 0,11 & 0,42 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B =$ $= \{6,94 \cdot 10^{-4};\ 0,02;\ 0,125;\ 0,25;\ 1\}$ and get the following fuzzy numbers:

$$ОМ = \{\ 1/6,94 \cdot 10^{-4};\ 0,44/0,02;\ 0,1/0,125;\ 0/0,25;\ 0/1\ \},$$
$$М = \{\ 0,71/6,94 \cdot 10^{-4};\ 1/0,02;\ 0,27/0,125;\ 0/0,25;\ 0/1\ \},$$
$$С = \{\ 0/6,94 \cdot 10^{-4};\ 0,4/0,02;\ 1/0,125;\ 0,35/0,25;\ 0/1\ \},$$
$$Б = \{\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0,36/0,125;\ 1/0,25;\ 0,71/1\ \},$$
$$ОБ = \{\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0,11/0,125;\ 0,42/0,25;\ 1/1\ \}.$$

The graph of the FP for the LP terms. The execution time of the process/ file is shown on Fig.8.

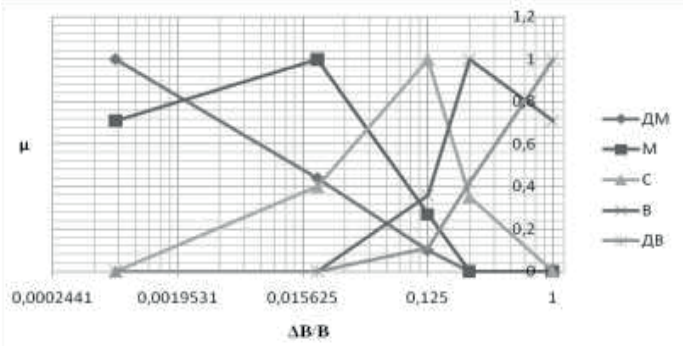

Fig. 8. Linguistic patterns of fuzzy numbers for RTPr/F

Thus, in the work with the use of SEDA, LV was introduced and models of the standards of parameters Tlog, Nlog, TSlog, I, CPU, Muse, NEF, NEr, RTPr/F were created. Also, for each LV, the FV were calculated and the graphs of their terms were plotted. Formed standards are necessary for the formation of logical rules to ensure the functioning of the SEDA. The results obtained will be further used to create an IDS system based on honeypot technology. Firstly, we will form in the next subsection of the DR to identify the fact of an IS violation and identify the identity of the attacker

DR system for efficient implementation of virtual honeypots.

To date, the following classes of SEDA are being developed:

 systems that detect the fact of an intrusion based on a comparison of the functional state of the IS with a set of specific signatures (templates) and

 systems based on used so-called DR (Gizun et all, 2013- Siddiqui, et all, 2016).

The main disadvantage of the first class of SEDA, which DR-based systems are deprived of, is the impossibility of their use in unknown attacks and, as a result, the impossibility of practical use in conditions of uncertainty and a fuzzy formalized environment. Therefore, despite the extremely low percentage of false positives of signature systems, the further development of second-class SEDA, in our opinion, is much more promising. The existing decision-principle SEDA are mainly focused on the use of complex mathematical models and require a lot of time for the formation of statistical data. However, expert approaches do not have such a requirement, which greatly simplifies the use of this method in the field of constructing SEDA. In this regard, an urgent task in the development of SOPs is to create models for detecting an anomalous state of IS caused by the activity of an intruder, based on the use of fuzzy logic methods, expert assessments and models of parameter standards necessary to identify an intruder. The use of these

models in the construction of the decisive type of SEDA is associated with the need to form rules aimed at identifying the intruder and his identification. That is why the purpose of this work is to develop a mathematical model that is used in the formation of the corresponding DR to identify the intruder.

In previous works (Avkurova, et all, 2020- Iashvili, et all, 2021), two groups of parameters for detecting and identifying an intruder are defined: with fuzzy and clear nature, respectively. So fuzzy parameters (Login time, Tlog; Frequency of login requests, Nlog; Time spent on login, TSlog; Intensity of action, I; Processor time / processor load, CPU; Amount of loaded RAM, Muse; Number of executable files, NEF; Number of failures and errors, NEr; Process/file execution time, RTPr/F) at the first stage of the SEDA operation make it possible to identify the presence of an intruder in the IS and to carry out their preliminary identification in a poorly formalized environment. Clear parameters should be used to confirm the fact of activity of the violation and the final assignment of the intruder to a certain category (Login user name, UID; Type of files used in the attack, AtEF; Unusual processes, UPr; Transfer file to the system, TrFin; modifying files, ModF; copy / transfer files from the system, TrFout; Keyboard keystrokes, KS) at the second stage of the system operation.

To solve the problem, it is necessary to create sets of DR, which are some statements that are based on the result of generalization of certain theoretical and experimental knowledge (data) and reflect the intuitive judgments of experts to ensure the search for a rational semantic solution to weakly formalized problems.

The construction of the DR can be carried out using the corresponding model, for the creation of which we introduce a set of linguistic identifiers:

$$LI = \bigcup_{i=1}^{d} LI = \{LI_1, LI_2, ..., LI_d\}, \qquad (8)$$

where d - the number of elements of the set required to detect the anomalous state, a LI_i (i = 1,d) – elements LI, each of which takes one of the text values that characterize in linguistic form the level of the anomalous state of the system, which can be generated by attacking actions. For example, for d = 5, expression (1) can be defined as follows:

$$LI = \bigcup_{i=1}^{5} LI = \{LI_1, LI_2, LI_3, LI_4, LI_5\} = \{Н, БНВ, БВН, В, К\}, \qquad (9)$$

where LI_1=Н, LI_2=БНВ, LI_3=ББН, LI_4=В and LI_5=К respectively reflected by the text values «Low», «More low than high», «More high than low», «High» and «Critical».

Further, based on the sets of identifiers LI and the set of linguistic or extended by the name - logical-linguistic connection LC, we will construct a set of PV:

$$ER = \left\{ \bigcup_{i=1}^{n} ER_i \right\} = \{ER_1, ER_2, ..., ER_n\}, \tag{10}$$

where ER_i (i=1,n) - a subset of possible rules for detecting the n-th anomalous state generated by the n-th attack, while

$$\begin{aligned} &\bigcup_{i=1}^{n} ER_i = \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_n} ER_{ij} \right\} = \{ER_{11}, ER_{12}, ..., ER_{1r_1}\}, \\ &\{ER_{21}, ER_{22}, ..., ER_{2r_2}\}, ..., \{ER_{n1}, ER_{n2}, ..., ER_{nr_n}\}, \end{aligned} \tag{11}$$

where ER_{ij} (i=1,n, j=1,r_n) - j-th rule of the n-th subset of possible rules, r_i (i=1,n) - the total number of possible rules aimed at detecting the n-th anomaly.

Note that for each ER_{ij} there is a corresponding DR:

$$\begin{aligned} &\{ER_{11} = (LC_{11} \to LI_{11}), ER_{12} = (LC_{12} \to LI_{12}), ..., \\ &ER_{1r_1} = (LC_{1r_1} \to LI_{1r_1})\}, \{ER_{21} = (LC_{21} \to LI_{21}), \\ &ER_{22} = (LC_{22} \to LI_{22}), ..., ER_{2r_2} = (LC_{2r_2} \to LI_{2r_2})\}, \\ &... \\ &\{ER_{n1} = (LC_{n1} \to LI_{n1}), ER_{n2} = (LC_{n2} \to LI_{n2}), ..., \\ &ER_{nr_n} = (LC_{nr_n} \to LI_{nr_n})\}. \end{aligned} \tag{12}$$

Generalizing expression (5), taking into account (10) and (11), we obtain

$$\begin{aligned} &ER = \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_i} ER_{ir_j} \right\} = \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_i} (LC_{ir_j} \to LI_{ir_j}) \right\} = \\ &\left\{ \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_i} ER_{ir_j} = (LC_{ir_j} \to LI_{ir_j}) \right\} \right\}, \end{aligned} \tag{13}$$

where ER_{ir_j} is r_j-th anomaly detection rule generated by the n-th attack, which is literally interpreted as: «If LC_{ir_j} is true, then the level of the abnormal state that can be generated by the n-th attack will be LI_{ir_j}».

The creation of rules is usually carried out on the basis of an expert approach, this is especially important in cases where it is necessary to give preference to one of the alternatives, for example, in which LC_{ir_j} (13) is the result associated with LI_{ir_j} and most objectively reflect the state of the system. Let consider the process of forming a choice for a set of alternatives using a specific example.

Let r_i logical-linguistic links and d (1) linguistic identifiers be used to create a set of rules, one of which can most objectively reflect the state of the environment about the presence of an anomaly. Therefore, the total number of possible alternative solutions is dxr_1, that is, for the assembly of each rule ER_{1j} ($j = 1,r_1$), it is necessary to consider d alternative variants of the rules, to select one of which we will use the methods of determining the coefficients of importance (CI). We will use the rank transformation method (RT), since it allows using the services of several experts, tabular forms are used as input, the initial function is linear, and the complexity is low.

Next, as an example, we define d=5, r_1=3, then

$$LC_I = \left\{ \bigcup_{j=1}^{r_1} LC_{1j} \right\} = \{LC_1, LC_2, LC_3\} = \{(t_{Tlog} \cong П,$$
$$t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B,$$
$$t_{RTPr/F} \cong ДМ), (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B,$$
$$t_{CPU} \cong B, t_{NEF} \cong ДВ, t_{NEr} \cong B, t_{RTPr/F} \cong ДВ),$$
$$(t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong ДВ, t_{CPU} \cong B, t_{NEF} \cong ДВ,$$
$$t_{NEr} \cong B, t_{RTPr/F} \cong ДВ)\},$$

and as values LI_{1k}(k = 1,..,5) use the data from the formula (9). Thus, for each LC_{1j}(j = 1,..,3) there are possible d=5 finals of detection of anomalies associated with specific values of linguistic identifiers in (9). The most objective result will be determined using the method of average ranks (AR) (Gornitska et all, 2012).

According to this method, as an example, we will use the judgments of 4 experts on d=5 of possible results ER^k_{1j}(k = 1,d, j = 1,r_1) for each j-th rule.

For example, for the first rule, the set of alternative solutions will be

$$\bigcup_{k=1}^{d} ER_{11}^{k} = \left\{ ER_{11}^{1}, ER_{11}^{2}, ER_{11}^{3}, ER_{11}^{4}, ER_{11}^{5} \right\} =$$

$$\{(t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B,$$

$$t_{NEr} \cong B, t_{RTPr/F} \cong ДМ) \rightarrow H, (t_{Tlog} \cong П, t_{Nlog} \cong BC,$$

$$t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong ДМ)$$

$$\rightarrow БНВ, (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B,$$

$$t_{NEF} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong ДМ) \rightarrow БВН, (t_{Tlog} \cong П,$$

$$t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B,$$

$$t_{RTPr/F} \cong ДМ) \rightarrow B, (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B,$$

$$t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong ДМ) \rightarrow K\}.$$

Further, on the basis of RT, we determine the CI, which are reflected by the parameter λ. Its minimum value indicates a greater advantage of the alternative, that is, its CI is higher. Fo the rule ER_{11} let's calculate the values x_{1j}^{k} i λ_{1j}^{k} for each of the possible results $ER^{k}{}_{11}$ (k = 1,5): x_{11}^{1}=(1+3+1+2)/4=1,75; x_{11}^{2}=(2+1+3+2)/4=2; x_{11}^{3}=(3+2+2+2)/4=2,25; x_{11}^{4}=(2+4+3+3)/4=3; x_{11}^{5} =(4+4+ 3+4)/4=3,75. The CI value is defined as $\lambda_{1j}^{k} = x_{1j}^{k} / N$, where N - sum of all ranks (N=10). According to the results listed in Table 1, it can be seen that the best result has ER_{11}^{1}, since $\bigwedge_{k=1}^{5} \lambda_{11}^{k} = \lambda_{11}^{k} = 0{,}18$.

Similarly, we will make calculations for ER_{1j}^{k}(j=2,3): for ER_{12}^{k} - x_{12}^{1} =(2+3+1+2)/4=2, x_{12}^{2}=(1+2+1+2)/4=1,5; x_{12}^{3}=(3+1+2+3)/4=2,25; x_{12}^{4} =(3+4+2+2)/4=2,75; x_{12}^{5}=(3+2+3+4)/4=3; для ER_{13}^{k} - x_{13}^{1}=(2+3+2+4)/4=2,75; x_{13}^{2}=(3+2+2+1)/4 =2; x_{13}^{3}=(2+3+1+1)/4=1,75; x_{13}^{4}=(3+4+3+4)/4= 3,5; x_{13}^{5} =(4+3+2 +4)/4=3,25.

Results and discussion. The calculation results (see Table 9) show that the best result for the rules ER_{12}, ER_{13}, have corresponding alternatives ER_{12}^{2}, ER_{13}^{3}.

The obtained data can be used as specific values in the construction of real rules in practical SEDA (Niu, et all, 2017- Ma, et all, 2019).

Ranks ER^k_{1j} and CI Table 9

ER^k_{1j}	j	k	Experts				λ^k_{1j}	λ^k_{1j}
			1	2	3	4		
ER^1_{11}	1	1	1	3	1	2	1,75	0,18
ER^2_{11}		2	2	1	3	2	2	0,2
ER^3_{11}		3	3	2	2	2	2,25	0,23
ER^4_{11}		4	2	4	3	3	3	0,3
ER^5_{11}		5	4	4	3	4	3,75	0,38
ER^1_{12}	2	1	2	3	1	2	2	0,2
ER^2_{12}		2	1	2	1	2	1,5	0,15
ER^3_{12}		3	3	1	2	3	2,25	0,23
ER^4_{12}		4	3	4	2	2	2,75	0,28
ER^5_{12}		5	3	2	3	4	3	0,3
ER^1_{13}	3	1	2	3	2	4	2,75	0,28
ER^2_{13}		2	3	2	2	1	2	0,2
ER^3_{13}		3	2	3	1	1	1,75	0,18
ER^4_{13}		4	3	4	3	4	3,5	0,35
ER^5_{13}		5	4	3	2	4	3,25	0,33

Thus the rule ER_{11} will take the form:

$ER_{11} = (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong$

$\cong ДМ) \rightarrow$ which can be verbally interpreted as follows: «If the current values of $t_{Tlog}, t_{Nlog}, t_{TSlog}, t_{CPU}, t_{NEF}, t_{NEr}, t_{RTPr/F}$ are closest to the values $П, BC, B, B, B, B, ДМ$ respectively, included in $T_{Tlog}, T_{Nlog}, T_{TSlog}, T_{CPU}, T_{NEF}, T_{NEr}, T_{RTPr/F}$, then the level of anomalous state generated by the attack of the intruder (in this case, a disinformer, cracker or hacker) will be LOW «.

Based on fuzzy parameters, we will form a set of DR ER_1 to identify such categories of the intruder as a disinformer, cracker, hacker and present it in the form of table 10, denoting the values of the parameters П - suspicious, Н - illegitimate, ВС - above average, В - high (large) , ОБ - very large, ОМ - very small, М – small:

Set of rules ER_i Table 10

Rule	t_{Tlog}	t_{Nlog}	t_{TSlog}	t_{CPU}	t_{NEF}	t_{Ner}	$t_{RTPr/F}$	Result
ER_{11}	П	ВС	В	В	В	В	ОМ	Н
ER_{12}	П	ВС	В	В	В	В	М	Н
ER_{13}	П	ВС	В	В	В	В	В	Н
ER_{14}	П	ВС	В	В	В	В	ОБ	Н
ER_{15}	П	ВС	В	В	ОБ	В	ОМ	БНВ
ER_{16}	П	ВС	В	В	ОБ	В	М	Н
ER_{17}	П	ВС	В	В	ОБ	В	В	Н
ER_{18}	П	ВС	В	В	ОБ	В	ОБ	БНВ
ER_{19}	П	ВС	ОБ	В	В	В	ОМ	БНВ
ER_{110}	П	ВС	ОБ	В	В	В	М	БНВ
ER_{111}	П	ВС	ОБ	В	В	В	В	БНВ
ER_{112}	П	ВС	ОБ	В	В	В	ОБ	БНВ
ER_{113}	П	ВС	ОБ	В	ОБ	В	ОМ	БВН
ER_{114}	П	ВС	ОБ	В	ОБ	В	М	БНВ
ER_{115}	П	ВС	ОБ	В	ОБ	В	В	БНВ
ER_{116}	П	ВС	ОБ	В	ОБ	В	ОБ	БВН
ER_{117}	П	В	В	В	В	В	ОМ	БНВ
ER_{118}	П	В	В	В	В	В	М	БНВ
ER_{119}	П	В	В	В	В	В	В	БНВ
ER_{120}	П	В	В	В	В	В	ОБ	БНВ
ER_{121}	П	В	В	В	ОБ	В	ОМ	БВН
ER_{122}	П	В	В	В	ОБ	В	М	БНВ
ER_{123}	П	В	В	В	ОБ	В	В	БНВ
ER_{124}	П	В	В	В	ОБ	В	ОБ	БВН
ER_{125}	П	В	ОБ	В	В	В	ОМ	БВН
ER_{126}	П	В	ОБ	В	В	В	М	БНВ
ER_{127}	П	В	ОБ	В	В	В	В	БНВ
ER_{128}	П	В	ОБ	В	В	В	ОБ	БВН
ER_{129}	П	В	ОБ	В	ОБ	В	ОМ	В
ER_{130}	П	В	ОБ	В	ОБ	В	М	БВН
ER_{131}	П	В	ОБ	В	ОБ	В	В	БВН
ER_{132}	П	В	ОБ	В	ОБ	В	ОБ	В
ER_{133}	Н	ВС	В	В	В	В	ОМ	БНВ
ER_{134}	Н	ВС	В	В	В	В	М	Н

ER_{135}	Н	ВС	В	В	В	В	В	Н
ER_{136}	Н	ВС	В	В	В	В	ОБ	БНВ
ER_{137}	Н	ВС	В	В	ОБ	В	ОМ	БВН
ER_{138}	Н	ВС	В	В	ОБ	В	М	БНВ
ER_{139}	Н	ВС	В	В	ОБ	В	В	БНВ
ER_{140}	Н	ВС	В	В	ОБ	В	ОБ	БВН
ER_{141}	Н	ВС	ОБ	В	В	В	ОМ	БВН
ER_{142}	Н	ВС	ОБ	В	В	В	М	БВН
ER_{143}	Н	ВС	ОБ	В	В	В	В	БВН
ER_{144}	Н	ВС	ОБ	В	В	В	ОБ	БВН
ER_{145}	Н	ВС	ОБ	В	ОБ	В	ОМ	В
ER_{146}	Н	ВС	ОБ	В	ОБ	В	М	БВН
ER_{147}	Н	ВС	ОБ	В	ОБ	В	В	БВН
ER_{148}	Н	ВС	ОБ	В	ОБ	В	ОБ	В
ER_{149}	Н	В	В	В	В	В	ОМ	БВН
ER_{150}	Н	В	В	В	В	В	М	БВН
ER_{151}	Н	В	В	В	В	В	В	БВН
ER_{152}	Н	В	В	В	В	В	ОБ	БВН
ER_{153}	Н	В	В	В	ОБ	В	ОМ	В
ER_{154}	Н	В	В	В	ОБ	В	М	БВН
ER_{155}	Н	В	В	В	ОБ	В	В	БВН
ER_{156}	Н	В	В	В	ОБ	В	ОБ	В
ER_{157}	Н	В	ОБ	В	В	В	ОМ	В
ER_{158}	Н	В	ОБ	В	В	В	М	БВН
ER_{159}	Н	В	ОБ	В	В	В	В	БВН
ER_{160}	Н	В	ОБ	В	В	В	ОБ	В
ER_{161}	Н	В	ОБ	В	ОБ	В	ОМ	К
ER_{162}	Н	В	ОБ	В	ОБ	В	М	В
ER_{163}	Н	В	ОБ	В	ОБ	В	В	В
ER_{164}	Н	В	ОБ	В	ОБ	В	ОБ	К

Based on fuzzy parameters, we will form sets of DRs to identify a spammer:

$$ER_2 = \{ER_{21} = (t_1 \cong C, t_{CPU} \cong B, t_{NEr} \cong B,$$
$$t_{RTPr/F} \cong ДМ) \rightarrow БНВ, ER_{22} = (t_1 \cong C,$$
$$t_{CPU} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong M) \rightarrow H,$$

$ER_{23} = (t_1 \cong B, t_{CPU} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong$
$ДМ) \rightarrow B, ER_{24} = (t_1 \cong B, t_{CPU} \cong B,$
$t_{NEr} \cong B, t_{RTPr/F} \cong M) \rightarrow БВН\}$

and ER_3 to detect and identify spam bots:

$ER_3 = \{ER_{31} = (t_1 \cong B, t_{CPU} \cong B, t_{NEr} \cong B,$
$t_{RTPr/F} \cong ДМ) \rightarrow B, ER_{32} = (t_1 \cong B, t_{CPU} \cong B,$
$t_{NEr} \cong B, t_{RTPr/F} \cong M) \rightarrow БВН\}.$

Based on fuzzy parameters, we will form a set of DRs for detecting and identifying a cracker bot and present it in the form of a table 11:

Set of rules ER_4 Table 11

Rule	t_{Tlog}	t_{Nlog}	t_{TSlog}	t_I	t_{CPU}	t_{NEF}	t_{Ner}	$t_{RTPr/F}$	Result
ER_{11}	П	ВС	В	В	В	В	В	ОМ	Н
ER_{12}	П	ВС	В	В	В	В	В	М	Н
ER_{13}	П	ВС	В	В	В	В	В	В	Н
ER_{14}	П	ВС	В	В	В	В	В	ОБ	Н
ER_{15}	П	ВС	В	В	В	ОБ	В	ОМ	БНВ
ER_{16}	П	ВС	В	В	В	ОБ	В	М	Н
ER_{17}	П	ВС	В	В	В	ОБ	В	В	Н
ER_{18}	П	ВС	В	В	В	ОБ	В	ОБ	БНВ
ER_{19}	П	ВС	ОБ	В	В	В	В	ОМ	БНВ
ER_{110}	П	ВС	ОБ	В	В	В	В	М	БНВ
ER_{111}	П	ВС	ОБ	В	В	В	В	В	БНВ
ER_{112}	П	ВС	ОБ	В	В	В	В	ОБ	БНВ
ER_{113}	П	ВС	ОБ	В	В	ОБ	В	ОМ	БВН
ER_{114}	П	ВС	ОБ	В	В	ОБ	В	М	БНВ
ER_{115}	П	ВС	ОБ	В	В	ОБ	В	В	БНВ
ER_{116}	П	ВС	ОБ	В	В	ОБ	В	ОБ	БВН
ER_{117}	П	В	В	В	В	В	В	ОМ	БНВ
ER_{118}	П	В	В	В	В	В	В	М	БНВ
ER_{119}	П	В	В	В	В	В	В	В	БНВ
ER_{120}	П	В	В	В	В	В	В	ОБ	БНВ
ER_{121}	П	В	В	В	В	ОБ	В	ОМ	БВН
ER_{122}	П	В	В	В	В	ОБ	В	М	БНВ
ER_{123}	П	В	В	В	В	ОБ	В	В	БНВ

ER_{124}	П	В	В	В	В	ОБ	В	ОБ	БВН
ER_{125}	П	В	ОБ	В	В	В	В	ОМ	БВН
ER_{126}	П	В	ОБ	В	В	В	В	М	БНВ
ER_{127}	П	В	ОБ	В	В	В	В	В	БНВ
ER_{128}	П	В	ОБ	В	В	В	В	ОБ	БВН
ER_{129}	П	В	ОБ	В	В	ОБ	В	ОМ	В
ER_{130}	П	В	ОБ	В	В	ОБ	В	М	БВН
ER_{131}	П	В	ОБ	В	В	ОБ	В	В	БВН
ER_{132}	П	В	ОБ	В	В	ОБ	В	ОБ	В
ER_{133}	Н	ВС	В	В	В	В	В	ОМ	БНВ
ER_{134}	Н	ВС	В	В	В	В	В	М	Н
ER_{135}	Н	ВС	В	В	В	В	В	В	Н
ER_{136}	Н	ВС	В	В	В	В	В	ОБ	БНВ
ER_{137}	Н	ВС	В	В	В	ОБ	В	ОМ	БВН
ER_{138}	Н	ВС	В	В	В	ОБ	В	М	БНВ
ER_{139}	Н	ВС	В	В	В	ОБ	В	В	БНВ
ER_{140}	Н	ВС	В	В	В	ОБ	В	ОБ	БВН
ER_{141}	Н	ВС	ОБ	В	В	В	В	ОМ	БВН
ER_{142}	Н	ВС	ОБ	В	В	В	В	М	БВН
ER_{143}	Н	ВС	ОБ	В	В	В	В	В	БВН
ER_{144}	Н	ВС	ОБ	В	В	В	В	ОБ	БВН
ER_{145}	Н	ВС	ОБ	В	В	ОБ	В	ОМ	В
ER_{146}	Н	ВС	ОБ	В	В	ОБ	В	М	БВН
ER_{147}	Н	ВС	ОБ	В	В	ОБ	В	В	БВН
ER_{148}	Н	ВС	ОБ	В	В	ОБ	В	ОБ	В
ER_{149}	Н	В	В	В	В	В	В	ОМ	БВН
ER_{150}	Н	В	В	В	В	В	В	М	БВН
ER_{151}	Н	В	В	В	В	В	В	В	БВН
ER_{152}	Н	В	В	В	В	В	В	ОБ	БВН
ER_{153}	Н	В	В	В	В	ОБ	В	ОМ	В
ER_{154}	Н	В	В	В	В	ОБ	В	М	БВН
ER_{155}	Н	В	В	В	В	ОБ	В	В	БВН
ER_{156}	Н	В	В	В	В	ОБ	В	ОБ	В
ER_{157}	Н	В	ОБ	В	В	В	В	ОМ	В
ER_{158}	Н	В	ОБ	В	В	В	В	М	БВН
ER_{159}	Н	В	ОБ	В	В	В	В	В	БВН

ER_{160}	H	B	ОБ	B	B	B	B	ОБ	B
ER_{161}	H	B	ОБ	B	B	ОБ	B	ОМ	К
ER_{162}	H	B	ОБ	B	B	ОБ	B	M	B
ER_{163}	H	B	ОБ	B	B	ОБ	B	B	B
ER_{164}	H	B	ОБ	B	B	ОБ	B	ОБ	К

At the second stage, after applying the DR, created on the basis of fuzzy parameters, the rules developed on the basis of clear parameters should be applied to verify the decision made and to carry out the final categorization of the intruder's personality. So they will look as follows:

$D = \{UID = 1, AtEF = PHP\ or\ Java\text{-}script,$

$UPr = 1, TrFin = 1, ModF = 1, TrFout = 0, KS = 1\}$ - for a disinformer,

$S = \{UID = 0, AtEF = PHP, UPr = 1, TrFin = 1,$

$ModF = 0, TrFout = 0, KS = 1\}$ - for a spammer,

$C = \{UID = 1, AtEF = .exe,\ .com, UPr = 1,$

$TrFin = 1, ModF = 1, TrFout = 1, KS = 1\}$ - for a cracker,

$H = \{UID = 1, AtEF = script, UPr = 1,$

$TrFin = 0, ModF = 1, TrFout = 1, KS = 1\}$ - for a hacker,

$SB = \{UID = 0, AtEF = PHP, UPr = 1,$

$TrFin = 1, ModF = 0, TrFout = 0, KS = 0\}$ - for a spam bot and

$B = \{UID = 0, AtEF = script, UPr = 1,$

$TrFin = 0, ModF = 1, TrFout = 1, KS = 0\}$ - for a hacker bot.

The obtained rules are used to develop PWIS based on expert methods and based on fuzzy logic (Khosravi et all, 2020).

Conclusions. Intruder identification in the cyberspace is not simple research task because cyberspace is complex and non-formalised space as well as security side faces with conditions of uncertainty. But there are some parameters in the cyberspace that can be monitored and analised. Intruder influences on these and the assessment process can explain the character of intruder action as well as his attack strategy. In their previous works, the authors described the parameters by which the intruder is identified – these are host and network parameters. Since the process of detecting and identifying an intruder takes place under conditions of uncertainty, and a number of certain parameters of systems for early detection of attacks are

fuzzy, the functioning of such a system should be based on fuzzy logic. This theory has chosen as the basis of research study.

In the work, based on the proposed parameters, using the MLTS, there were introduced LVs and were created models of the standards of parameters Tlog, Nlog, TSlog, I, CPU, Muse, NEF, NEr, RTPr/F. Also, for each LV, the FV were calculated and the graphs of their terms were plotted. Formed standards necessary for the formation of logical rules to ensure the functioning of the SEDA. The results obtained will be further used to create an IDS system based on honeypot technology (other technologies also can be used for attacks detection).

In addition, the proposed DR model using fuzzy logic for the first group of rules and conventional logic for the second group makes it possible to display the anomalous state by using the set "influence of the intruder-parameter", "influence of the intruder-set of logical-linguistic connections" and the universal model of parameter standards in IS, generated by the influence of an intruder of information security of a certain type. Based on this model, there were developed examples of rules to detect and identify the activities of a disinformer, spammer, cracker, hacker, spam bot and hacker bot, which can be used to improve existing or develop a new system for early detection of APT-attacks directed on the critical information infrastructure of the state or other important objects.

Information about authors:

Avkurova Zhadyra – Master of Technical Sciences, Senior Lecturer, Department of Artificial Intelligence Technology *https://www.scopus.com/authid/detail.uri? authorId=57226553845* ID *https://orcid.org/0000-0002-0706-6075*;

Sergiy Gnatyuk – Doctor of Technical Sciences, Professor, Faculty of Cybersecurity, Computer and Software Engineering, National Aviation University, Kyiv, Ukraine, *sergio.gnatyuk@gmail.com*, *https://www.scopus.com/authid/detail.uri?authorId=36184129600*, *https://orcid.org/0000-0003-4992-0564*;

Abduraimova Bayan – Candidate of technical sciences, associate professor ENU L.N.Gumilyov, abduraimova_bk@enu.kz, https://orcid.org/0000-0003-3913-1895;

Kydyralina Lazat – NAO "Shakarim University in Semey", PhD, acting associate professor, *lazat_75@mail.ru*, https://orcid.org/0000-0002-2836-0919.

REFERENCES

Avkurova Zh., Abduraimova B., Gnatyuk S., Gizun A. Analiz sovremennyh sistem obnaruzheniya atak na osnove tekhnologij virtual'noj primanki // №6(142) Vestnik KazNITU, s.654-659, noyabr', 2020. (in Kaz.).

Iashvili G., Avkurova Zh., Iavich M., Bauyrzhan M., Gagnidze A., Gnatyuk S. Content-Based Machine Learning Approach for Hardware Vulnerabilities Identification System, Lecture Notes on Data Engineering and Communications Technologies, Vol. 83, pp. 117-126, 2021. (in Eng.).

Zuzčák M. and Bujok P., "Causal analysis of attacks against honeypots based on properties of countries," in IET Information Security, vol. 13, no. 5, pp. 435-447, 9 2019, doi: 10.1049/iet-ifs.2018.5141. (in Eng.).

Zhang W., Zhang B., Zhou Y., He H. and Ding Z., "An IoT Honeynet Based on Multiport Honeypots for Capturing IoT Attacks," in IEEE Internet of Things Journal, vol. 7, no. 5, pp. 3991-3999, May 2020, doi: 10.1109/JIOT.2019.2956173. (in Eng.).

Golub V. Password protection [Electronic resource]: article / V. Golub // Relga. – 01.12. 2009. - No. 17 (197). - Access mode: *http://www.relga.ru/ Environ/WebObjects/tgu-www.woa/wa/Main?textid= 2516&level1= main&level2* (in Eng.).

Gizun A.I. Basic parameters for the identification of the violator of information security / A.I. Gizun, V.V. Volyanska, V.O. Риндюк, C.O. Hnatyuk // Information protection. - 2013. - Vol.15. - №1. - P. 66-75. (in Russ.).

Siddiqui S., Khan M.S., Ferens K. and Kinsner W., "Detecting advanced persistent threats using fractal dimension based machine learning classification", Proc. ACM Int. Workshop Secur. Privacy Anal. (IWSPA), pp. 64-69, 2016. (in Eng.).

Gornitska D.A. Determination of coefficients of importance for expert evaluation in the field of information security / D.A. Gornitska, Volyanska V.V., Korchenko A.O. // Information protection. - 2012. - №1 (54). - P. 108-121. (in Eng.).

Niu W., Zhang X., Yang G., Chen R. and Wang D., "Modeling attack process of advanced persistent threat using network evolution", IEICE Trans. Inf. Syst., vol. 100, no. 10, pp. 2275-2286, 2017. (in Eng.).

Ma Z., Li Q. and Meng X., "Discovering Suspicious APT Families Through a Large-Scale Domain Graph in Information-Centric IoT," in IEEE Access, vol. 7, pp. 13917-13926, 2019, doi: 10.1109/ACCESS.2019.2894509. (in Eng.).

Khosravi M. and Ladani B.T., "Alerts Correlation and Causal Analysis for APT Based Cyber Attack Detection," in IEEE Access, vol. 8, pp. 162642-162656, 2020, doi: 10.1109/ACCESS.2020.3021499. (in Eng.).

МАЗМҰНЫ

А.С.Ақанова, А.А.Макашев, С.А. Наурызбаева, Н.Н.Оспанова
ИНТЕРНЕТТЕН ТАҚЫРЫП БОЙЫНША ДЕРЕКТЕРДІ АЛУЫН МОДЕЛДЕУ.....5

Ж.С. Авкурова, С.А. Гнатюк, Б.К. Абдураимова, Л.М. Кыдыралина
КИБЕРКЕҢІСТІКТЕГІ АРТ-ШАБУЫЛДАРДЫ ЕРТЕ АНЫҚТАУ ЖӘНЕ БҰЗУШЫЛАРДЫ СӘЙКЕСТЕНДІРУ ҮШІН ЭТАЛОН МОДЕЛЬДЕРІ АНЫҚТАУШЫ ЕРЕЖЕЛЕР.....19

М.А. Болатбек, К.Б. Багитова, Ш.Ж. Мусиралиева
КИБЕРҚАУІПСІЗДІК МӘСЕЛЕЛЕРІН ТАБИҒИ ТІЛДІ ӨҢДЕУ ӘДІСТЕРІ АРҚЫЛЫ ШЕШУ ТАҚЫРЫБЫНА ЖҮЙЕЛІК ШОЛУ.....52

А.К. Жумадиллаева, М.Д. Кабибуллин, Б.Б. Оразбаев, К.Н. Оразбаева, Ж.Н. Тулеуов
КАТАЛИТИКАЛЫҚ РИФОРМИНГ ҚОНДЫРҒЫСЫ РИФОРМИНГТЕУ РЕАКТОРЛАРЫ ЖҰМЫС РЕЖИМДЕРІН КОМПЬЮТЕРЛІК МОДЕЛЬДЕУ НЕГІЗІНДЕ ОПТИМИЗАЦИЯЛАУ.....71

Ж.Д. Изтаев, Г.Т. Джусупбекова, Г.К. Ордабаева
УНИВЕРСИТЕТ ҮШІН АҚПАРАТТЫҚ ҚАУІПСІЗДІК ҚАТЕРЛЕРІНІҢ ЖЕКЕ МОДЕЛІН ӘЗІРЛЕУ.....91

Ж.С. Каженова, Ж.Е. Кенжебаева, А.М. Прудник
MQTT (ТЕЛЕМЕТРИЯ ХАБАРЛАМАЛАРЫ КЕЗЕГІН ТАСЫМАЛДАУ) ХАТТАМАСЫНЫҢ ҚАУІПСІЗДІК МЕХАНИЗМДЕРІ.....117

А.Ж. Картбаев, Г.С. Ыбытаева, О.Ж. Мамырбаев, К.Ж. Мухсина, Б.Ж. Жумажанов
АВТОМАТТЫ ҚЫЛМЫС ОНТОЛОГИЯСЫН ҚҰРУ ҮШІН ҚЫЛМЫС ЖАҢАЛЫҚТАРЫНДА СУБЪЕКТІЛЕРДІ ФОРМАЛЬДЫ КӨРСЕТУ ӘДІСТЕРІ.....136

А.Т. Мазақова, Қ.Б. Бегалиева, Т.Ж. Мазаков, Ш.А. Жомартова, Г.З. Зиятбекова
КВАДРАТ ҚИМАСЫ БАР ӨЗЕКШЕНІҢ ЖЫЛУ ӨТКІЗГІШТІК ТЕҢДЕУІН ҚАРАПАЙЫМ ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕР ЖҮЙЕСІНЕ ҚОЮ АРҚЫЛЫ ШЕШУ.....153

Ж.Ж. Молдашева, Б.Б. Оразбаев, Б.У. Асанова, С.Ш. Искакова, К.Н. Оразбаева
МҰНАЙ ҚҰБЫРЫ АГРЕГАТТАРЫНЫҢ ЖҰМЫС РЕЖИМДЕРІН БАСҚАРУ ҮШІН ЭВРИСТИКАЛЫҚ ТӘСІЛ ҚҰРУ.....,.164

А.Б. Мименбаева, А.С. Аканова
СОЛТҮСТІК ҚАЗАҚСТАН ОБЛЫСЫНЫҢ АУЫЛШАРУАШЫЛЫҒЫ ДАҚЫЛДАРЫНЫҢ КҮЙІН NDVI СЫЗЫҚТЫҚ ТРЕНДТЕРІ АРҚЫЛЫ ЗЕРТТЕУ.....185

М.О. Ногайбаева, Б. Ахметов, Дж.Дж. Расулзаде, Е.А. Максум, С. Рустамов
U-NET КОНВОЛЮЦИЯЛЫҚ НЕЙРОНДЫҚ ЖЕЛІ НЕГІЗІНДЕ ТОПОЛОГИЯЛЫҚ ОҢТАЙЛАНДЫРУДЫҢ ЕСЕПТЕУ ПРОЦЕСІН ЖЕДЕЛДЕТУ.....198

Г.Б. Туребаева, А.К. Сыздықов, А.Р. Тенчурина, Ж.Б. Дошакова
ҚОЛДАНБАЛЫ БАҒДАРЛАМАЛАРДЫ ҚОЛДАНА ОТЫРЫП ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕРДІ ШЕШУДІҢ САНДЫҚ ӘДІСТЕРІ.....214

К.С. Чежимбаева, А.Н. Хайруллина
LORA ҚАБЫЛДАҒЫШ/ТАРАТҰЫШЫНЫҢ ӨНІМДІЛІГІН БАҒАЛАУ.....228

А.Г. Шаушенова, А.А. Нурпейсова, Ж.С. Муталова, Д.Б. Досалянов, М.Б. Онгарбаева
ҚАШЫҚТЫҚТАН ОҚЫТУДА БІЛІМ АЛУШЫНЫ ИДЕНТИФИКАЦИЯЛАУ ЖӘНЕ БЕЙНЕМОНИТОРИНГТЕУ ШЕТЕЛДІК ЖҮЙЕЛЕРІНІҢ ЕРЕКШЕЛІКТЕРІ.....247

К. Якунин, Р.И. Мухамедиев, М. Елис, Я. Кучин, Н. Юничева, А. Сымагулов, Е. Мухамедиева
КОВИД-19 ПАНДЕМИЯСЫ ТАҚЫРЫП БОЙЫНША ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БАҚ БАСЫЛЫМДАРЫНЫҢ ТАҚЫРЫПТЫҚ КЛАСТЕРЛЕРІН ТАЛДАУ.....260

СОДЕРЖАНИЕ

А.С. Аканова, А.А. Макашев, С.А. Наурызбаева, Н.Н. Оспанова
МОДЕЛИРОВАНИЕ ТЕМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ ДАННЫХ ИЗ ИНТЕРНЕТА.....5

Ж.С. Авкурова, С.А. Гнатюк, Б.К. Абдураимова, Л.М. Кыдыралина
МОДЕЛИ ЭТАЛОНОВ И ОПРЕДЕЛЯЮЩИЕ ПРАВИЛА ДЛЯ СИСТЕМ РАННЕГО ВЫЯВЛЕНИЯ АРТ-АТАК И ИДЕНТИФИКАЦИИ НАРУШИТЕЛЕЙ В КИБЕРПРОСТРАНСТВЕ.....19

М.А. Болатбек, К.Б. Багитова, Ш.Ж. Мусиралиева
СИСТЕМАТИЧЕСКИЙ ОБЗОР ТЕМЫ РЕШЕНИЯ ЗАДАЧ КИБЕРБЕЗОПАСНОСТИ С ПОМОЩЬЮ МЕТОДОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА.....52

А.К. Жумадиллаева, М.Д. Кабибуллин, Б.Б. Оразбаев, К.Н. Оразбаева, Ж.Н. Тулеуов
ОПТИМИЗАЦИЯ РЕЖИМОВ РАБОТЫ РЕАКТОРОВ РИФОРМИНГА УСТАНОВКИ КАТАЛИТИЧЕСКОГО РИФОРМИНГА НА ОСНОВЕ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ.....71

Ж.Д. Изтаев, Г.Т. Джусупбекова, Г.К. Ордабаева
РАЗРАБОТКА ЧАСТНОЙ МОДЕЛИ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ДЛЯ УНИВЕРСИТЕТА.....91

Ж.С. Каженова, Ж.Е. Кенжебаева, А.М. Прудник
МЕХАНИЗМЫ БЕЗОПАСНОСТИ ПРОТОКОЛА MQTT (ТРАНСПОРТ ТЕЛЕМЕТРИИ ОЧЕРЕДИ СООБЩЕНИЙ).....117

А.Ж. Картбаев, Г.С. Ыбытаева, О.Ж. Мамырбаев, К.Ж. Мухсина, Б.Ж. Жумажанов
МЕТОДЫ ФОРМАЛЬНОГО ПРЕДСТАВЛЕНИЯ СУЩНОСТЕЙ В КРИМИНАЛЬНЫХ НОВОСТЯХ ДЛЯ АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ОНТОЛОГИИ ПРЕСТУПЛЕНИЙ.....136

А.Т. Мазакова, К.Б. Бегалиева, Т.Ж. Мазаков, Ш.А. Жомартова, Г.З. Зиятбекова
РЕШЕНИЕ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ СТЕРЖНЯ С КВАДРАТНЫМ СЕЧЕНИЕМ ПРИВИДЕНИЕМ К СИСТЕМЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ.....153

Ж.Ж. Молдашева, Б.Б. Оразбаев, Б.У. Асанова, С.Ш. Искакова, К.Н. Оразбаева
РАЗРАБОТКА ЭВРИСТИЧЕСКОГО МЕТОДА ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ УПРАВЛЕНИЯ РЕЖИМАМИ РАБОТЫ АГРЕГАТОВ НЕФТЕПРОВОДА.....164

А.Б. Мименбаева, А.С. Аканова
ИССЛЕДОВАНИЕ СОСТОЯНИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР СЕВЕРО-КАЗАХСТАНСКОЙ ОБЛАСТИ ПО ЛИНЕЙНЫМ ТРЕНДАМ NDVI.....185

М.О. Ногайбаева, Б. Ахметов, Дж.Дж. Расулзаде, Е.А. Максум, С. Рустамов
УСКОРЕНИЕ ВЫЧИСЛИТЕЛЬНОГО ПРОЦЕССА ТОПОЛОГИЧЕСКОЙ ОПТИМИЗАЦИИ НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ U-NET.....198

Г.Б. Туребаева, А.К. Сыздыков, А.Р. Тенчурина, Ж.Б. Дошаков
ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ИСПОЛЬЗОВАНИЕМ ПРИКЛАДНЫХ ПРОГРАММ.....214

К.С. Чежимбаева, А.Н. Хайруллина
ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПРИЕМОПЕРЕДАТЧИКА LORA.....228

А.Г. Шаушенова, А.А. Нурпейсова, Ж.С. Муталова, Д.Б. Досалянов, М.Б.Онгарбаева
ОСОБЕННОСТИ ЗАРУБЕЖНЫХ СИСТЕМ ВИДЕОМОНИТОРИНГА И ИДЕНТИФИКАЦИИ ОБУЧАЮЩЕГОСЯ В ДИСТАНЦИОННОМ ОБУЧЕНИИ.....247

К. Якунин, Р.И. Мухамедиев, М. Елис, Я. Кучин, А. Сымагулов, Н. Юничева, Е.Мухамедиева
АНАЛИЗ ТЕМАТИЧЕСКИХ КЛАСТЕРОВ ПУБЛИКАЦИЙ СМИ РЕСПУБЛИКИ КАЗАХСТАН ПО ТЕМЕ ПАНДЕМИИ COVID-19.....260

CONTENTS

A.S. Akanova, A.A. Makashev, C.A. Наурызбаева, N.N. Ospanova
MODELING OF THEMATIC DATA EXTRACTION FROM THE INTERNET.....5

Zh. Avkurova, S. Gnatyuk, *B.* Abduraimova, L. Kydyralina
MODELS OF STANDARDS AND GOVERNING RULES FOR THE SYSTEMS OF EARLY DETECTION OF APT-ATTACKS AND IDENTIFICATION OF VIOLATORS IN CYBERSPACE.....19

M. Bolatbek, K. Bagitova, Sh. Musiralieva
A SYSTEMATIC REVIEW ON CYBERSECURITY ISSUES USING NATURAL LANGUAGE PROCESSING TECHNIQUES.....52

A. Zhumadillayeva, M. Kabibullin, B. Orazbayev, K. Orazbayeva, Zh. Tuleuov
OPTIMIZATION OF THE OPERATING MODES OF THE REFORMING REACTORS OF THE CATALYTIC REFORMING UNIT BASED ON COMPUTER MODELING.....71

Zh.D. Iztayev, G.T. Dzhusupbekova, G.K. Ordabaeva
DEVELOPMENT OF A PRIVATE MODEL OF INFORMATION SECURITY THREATS FOR THE UNIVERSITY.....91

Zh.S. Kazhenova, Zh.E. Kenzhebayeva, A.M. Prudnik
SECURITY MECHANISMS OF PROTOCOL MQTT (MESSAGE QUEUING TELEMETRY TRANSPORT).....117

A.Zh. Kartbayev, G.S. Ybytayeva, O.Zh. Mamyrbayev, K.Zh. Mukhsina, B.Zh. Zhumazhanov
METHODS FOR FORMAL REPRESENTATION OF ENTITIES IN CRIME NEWS FOR AUTOMATIC CRIME ONTOLOGY CONSTRUCTION.....136

A.T. Mazakova, K.B. Begaliyeva, T.Zh. Mazakov, Sh.A. Jomartova, G.Z. Ziyatbekova
SOLUTION OF THE THERMAL CONDUCTIVITY EQUATION OF A ROD WITH A SQUARE SECTION BY CASTING TO A SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS.....153

Zh. Moldasheva, B. Orazbayev, B. Assanova, Sh. Iskakova, K. Orazbayeva
OPTIMIZATION OF OPERATION MODES OF REFORMING REACTORS OF A CATALYTIC REFORMING UNIT ON THE BASIS OF COMPUTER MODELING.....164

A.B. Mimenbayeva, A.C. Akanova
RESEARCH OF THE STATE OF AGRICULTURAL CROPS NORTH KAZAKHSTAN REGION ACCORDING TO LINEAR NDVI TRENDS.....185

M. Nogaibayeva, B. Akhmetov, J. Rasulzade, Y. Maksum, S. Rustamov
ACCELERATION OF THE COMPUTATIONAL PROCESS OF TOPOLOGICAL OPTIMIZATION BASED ON THE CONVOLUTIONAL NEURAL NETWORK U-NET.....198

G. Turebaeva, A. Syzdykov, A. Tenchurina, J. Doshakov
NUMERICAL METHODS FOR SOLVING DIFFERENTIAL EQUATIONS USING APPLICATION PROGRAMS.....214

K.S. Chezimbayeva, A.N. Khairullina
EVALUATION OF LORA TRANSCEIVER PERFORMANCE.....228

A.G. Shaushenova, A.A. Nurpeisova, Z.S. Mutalova, D.B. Dosalyanov, M.B. Ongarbaeva
FEATURES OF FOREIGN SYSTEMS OF VIDEO MONITORING AND IDENTIFICATION OF STUDENTS IN DISTANCE LEARNING.....247

K. Yakunin, R.I. Mukhamediev, M. Elis, Ya. Kuchin, N. Yunicheva, A. Symagulov, E. Mukhamedieva
ANALYSIS OF THEMATIC CLUSTERS OF KAZAKHSTAN MEDIA PUBLICATIONS ON THE TOPIC OF THE COVID-19 PANDEMIC.....260

Publication Ethics and Publication Malpracticein
the journals of the National Academy of Sciences of the Republic of Kazakhstan

For information on Ethics in publishing and Ethicalguidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any otherlanguage, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*
Заместитель директор отдела издания научных журналов НАН РК *Р. Жәлиқызы*
Редакторы: *М.С. Ахметова, Д.С. Аленов*
Верстка на компьютере *Г.Д.Жадыранова*

Подписано в печать 15.09.2022.
Формат 60х881/8. Бумага офсетная. Печать – ризограф.
17,5 п.л. Тираж 300. Заказ 3.

Национальная академия наук РК
050010, Алматы, ул. Шевченко, 28, т. 272-13-19