Journal of Theoretical and Applied Information Technology
15th November 2017. Vol.95. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEVELOPING OF THE CYBER SECURITY SYSTEM BASED ON CLUSTERING AND FORMATION OF CONTROL DEVIATION SIGNS

[1] LAKHNO V. A., [2] KRAVCHUK P. U, [3] MALYUKOV V. P., [4] DOMRACHEV V. N., [5] MYRUTENKO L.V., [6]PIVEN O. S.

[1,2] Department of Cyber Security, European University, Ukraine
[3,6] Department of Information Systems and Mathematical Sciences, European University, Ukraine
[4]Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Ukraine
[5]Kyiv Cooperative Institute of Business and Law, Ukraine

E-mail: [1]lva964@gmail.com, [2]p.kr@ukr.net, [3]volod.malyukov@gmail.com, [4]mipt@ukr.net, [5]myrutenko.lara@gmail.com, [6] alekspvn@mail.ua

ABSTRACT

The cyber security (CS) adaptive system is developed. It is based on advanced algorithms of anomalies signs space partitioning and attacks on clusters. A new approach of solving the topical scientific and applied problem of increasing the efficiency of systems of intelligent recognition of cyber attacks and anomalies is proposed. Unlike the existing ones, the present approach allows to take into account the modern statistical and remote parameters of the clustering of the attributes of cyber attacks and provides the opportunity to change the valid tolerance deviations for all the attributes simultaneously, as well as quickly identify new types of complex combined attacks with limited computing resources and variability of conditions. Unlike existing algorithms, the advanced ones enable to take into account the subject area peculiarities, including legal characteristics of cyber crimes of space signs construction. Among PTC Mathcad Prime 4.0, MATLAB (Simulink), using established simulations, the performance of the proposed algorithms are tested in CS systems of various companies.

Keywords: *Clustering Features, Cyber Security, Simulation Experiment, Test Tolerances.*

1. INTRODUCTION

The huge development of IT field, particularly in critical important computer systems (CICS), faces new threats to cyber security (CS).

Modern cyber attack security systems should have been highly adaptable because of increasing of cyber attacks (C-A) in the world. This is the ability to change the algorithm of detecting C-A's immediately. In particular, one can use effective methods for clustering object recognition systems (ORS), such as anomalies or C-A's.

Improving existing algorithms and development of new ones for the clustering of ORS signs in adaptive recognition systems (ARS) of C-A's and anomalies make the present research relevant.

2. LITERATURE REVIEW

The models of constructing clusters in ARS of C-A's in CICS were investigated in many reseaches [1–3]. Preferably, these studies were of high theoretical value.

As indicators (metrics) for the construction of classifiers of ORS were studied the following: traffic parameters [4], unpredictable packet addresses [5], attributes of requests for databases and data warehouses (DB and DW) [6, 7], etc.

However, mentioned reseaches [4–7] do not take into account the possibility of parallel formation of control deviations for signs of anomalies and C-A's [8].

For complex C-A's, information signs may be rather fuzzy [9, 10]. This, in its turn, does not contribute to the construction of effective recognition algorithms.

The effectiveness of recognition can be significantly increased [11, 12] by using cluster analysis methods [13, 14]. However, these reseaches are not brought to the hardware or software implementation. In order to eliminate this disadvantage, we can apply the informational

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

condition of functional productiveness (ICFP) of ARS learning [15, 16].

In [17, 18] it is shown that in the case of the permanent glossary of signs of ORS (anomalies or C-A's), the productivity of ARS might be increased.

As it was shown in [19], K-means algorithm has low computational complexity, it is the main advantage of K-means and it works with a large number of data well. The DBSCAN algorithm works slowly enough with a large amount of data.

Taking into account the prospects of using intelligent detection systems for cyber attacks [11, 18], the article presents an improved clustering algorithm of control attributes of anomalies and cyber attacks for their timely detection in computer systems, which is very perspective.

Therefore, it is important to improve the clustering algorithms and formulate control abnormalities of anomalies and C-A's in CICS in the future taking into account the potential for the use of adaptive intelligence detection systems (AIDS).

3. FORMULATION OF THE RESEARCH PROBLEMS

The purpose of the research is the development of the algorithm for the partitioning of the space of attributes (SA) into clusters in the process of recognition of C-A's in intelligent CS systems due to the simultaneous formation of validation tolerances for different attack classes;

In order to achieve the purpose of the research, the following tasks must be solved:

- to improve algorithms of clustering of signs of the recognition objects in adaptive systems of detecting C-A's;
- to investigate on the simulation models the adequacy of the proposed algorithms.

4. METHODOLGY

The attributes of cyber attacks are found in a large amount of measured information, such as logs, monitoring data, etc. The process of preventing them requires the increasing of information processing speed in cyber attack detection systems.

By combining data into compact clusters, we can analyze the typical representatives of each cluster and decide whether such data is a sign of attack or not. The result is transferred to all representatives of the studied cluster. This approach significantly reduces the amount of information which is necessary to classify an attack successfully.

There are complex forms of clusters in multidimensional space, thus some authors propose to consider different clustering algorithms such as K-means [1, 2] DBSCAN [5], FDBSCAN [8], and others.

The computational complexity of the algorithms used in the binary space of recognition attributes (BSRA) [10, 11, 18] corresponding to the class (classes) depends on the optimal form of a container of the recognition object.

In order to make the construction of the container easier, we can make the following assumption: there is a "pseudo hypersphere" (or a pseudo-spherical container - PSC) [11, 18], which allows to consider the parameters of optimization of PSC in BSRA as a certain vector of the standard. The pinnacle vector will determine the geometric center of the PSC.

In the process of adaptive recognition systems learning, we make an assumption about fuzzy compactness of the implementation of binary learning matrices (BLM) [16, 21, 22], obtained at the stage of splitting of the feature space (FS) into relevant RO classes. Fuzzy partition $RC^{|M|}$ includes the elements that can be attributed to fuzzy RO classes [4, 16].

The rules of ASR learning, according to [2, 4, 23, 26], are built based on the iteration procedure of searching for the $\max$ boundary magnitude of an information condition of functional effectiveness (ICFE):

$$is'_k = \arg\max_{IS_k}\{\max_{IS_{k-1}}\{\ldots\{\max_{IS_1\cap IS_{CE}}\frac{1}{M}\sum_{m-1}^{M}CE_m\}\ldots\}\}, \quad (1)$$

where CE_m is the ICFE of ASR learning to recognize RO that belong to class C_m^0; IS_k is the permissible range of values of the k-th informative attribute of RO; IS_{CE} is the permissible range of ICFE in the course of ASR learning, $m=\overline{1,M}$.

The following constraints are imposed on expression (2):

$$\left(\forall CT_m^o \in RC^{|M|}\right)\left\lfloor CT_m^o \neq \varnothing \right\rfloor; \quad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\left(\exists CT_a^o \in RC^{|M|}\right)\left(\exists CT_b^o \in RC^{|M|}\right)\left[\begin{array}{l} CT_a^o \neq CT_b^o \rightarrow \\ \rightarrow CT_a^o \bigcap CT_b^o \neq \varnothing \end{array}\right]; \quad (3)$$

$$\left(\forall CT_a^o \in RC^{|M|}\right)\left(\forall CT_b^o \in RC^{|M|}\right)\left[\begin{array}{l} CT_a^o \neq CT_b^o \rightarrow \\ \rightarrow B\,CT_a^o \bigcap B\,CT_b^o = \varnothing \end{array}\right], \quad (4)$$

where $B\,CT_a^o$ & $B\,CT_b^o$ are the nuclei of RO classes CT_a^o & CT_b^o, respectively;

$$\bigcup_{CT_m^o \in RC} CT_m^o \subseteq RS_B; a \neq b. \quad (5)$$

Assume: classes CT_a^o are CT_b^o adjacent; the classes have a $\min$ distance between the centers of clusters $cr(ct_a \oplus ct_b)$ among all classes for RO; RO are described by BLM [21–23]. We accepted that ct_a and ct_b are the reference vectors of RO classes, in particular, by the KDD [2-7, 9-11].

The ASR learning procedure is given in the form of predicate expression:

$$\left(\forall CE_a^o \in RC^{|M|}\right)\left(\forall CT_b^o \in RC^{|M|}\right)\left[\begin{array}{l} CT_a^o \neq CT_b^o \rightarrow \\ \rightarrow \left(cr_a' < cr(ct_a \oplus ct_b)\right)\cdot \\ \cdot \left(cr_b' < cr(ct_a \oplus ct_b)\right) \end{array}\right], \quad (6)$$

where cr_a' & cr_b' are the optimal radii of containers C_a^o & C_b^o, respectively.

ASR learning is an iteration procedure [2, 5, 8, 20, 24]:

$$ca^* = \arg\max_{IS_{ca}}\{\max_{IS_{CE} \cap IS_{cr}} \overline{CE}\}, \quad (7)$$

where IS_{ca} is the admissible range of magnitudes of reference deviation ca for RO class $\{CT_m^o\}$; IS_{cr} is the permissible range of RC magnitude cr.

The algorithm of the recognition objects (anomalies and C-A's) classification is functional at the following restrictions:

$$\left(\forall CT_{m,\xi}^o \in RC^{|M|}\right)\left[CT_{m,\xi}^o \neq \varnothing, m = \overline{1,M}\right], \quad (8)$$

$$\left(\forall CT_{m,\xi}^o \in RC^{|M|}\right)\left(\forall CT_{c,\xi}^o \in RC^{|M|}\right)\left[\begin{array}{l} CT_{m,\xi}^o \neq \\ \neq CT_{c,\xi}^o \rightarrow B\,CT_{m,\xi}^o \bigcap B\,CT_{c,\xi}^o = \varnothing \end{array}\right], \quad (9)$$

$$\left(\forall CT_{m,\xi}^o \in RC^{|M|}\right)\left(\forall CT_{c,\xi}^o \in RC^{|O|}\right)\left[\begin{array}{l} CT_{m,\xi}^o \\ \neq CT_{c/\xi}^o \rightarrow \\ \left(cr_{m,\xi}' < cr\left(ct_{m,\xi} \oplus ct_{c,\xi}\right)\right)\wedge \\ \wedge\left(cr_{c,\xi}' < cr\left(ct_{m,\xi} \oplus ct_{c,h}\right)\right) \end{array}\right], \quad (10)$$

$$\bigcup_{CT_{m,\xi}^o \in RC} CT_{m,\xi}^o \subseteq RS, \quad (11)$$

where $B\,CT_{m,\xi}^o$, $B\,CT_{c,\xi}^o$ are the centers of the two nearest (adjacent) clusters $CT_{m,\xi}^o$ & $CT_{c,\xi}^o$, respectively; ξ is the step of increasing the radius of cluster container (RCC); $cr_{m,\xi}'$ & $cr_{c,\xi}'$ are, respectively, formed RCC $CT_{m,\xi}^o$ and $CT_{c,\xi}^o$; $cr(ct_m \oplus ct_c)$ is the inter-center code distance of clusters $CT_{m,\xi}^o$ & $CT_{c,\xi}^o$.

Steps of splitting FS into clusters:

Step 1. Counter of changing (SC) VAD ca_i by features of RO is set as $i := 0$.

Step 2. Calculation of the lower $A_{low_i}[l]$ and the upper $A_{up_i}[l]$ of VAD of RO features for entire FS:

$$A_{low_i}[l] = lm_i - ca\frac{ca_{low_i}}{100}; \quad (12)$$

$$A_{up_i}[l] = lm_i + ca\frac{ca_{low_i}}{100}, \quad (13)$$

where lm_i is the i-th attribute of standard vector-realization of non-classified multi-dimensional matrix (NMLM) $\left\|lm_i^{(j)}\right\|$ [16, 23]; ca_{low_i} is the VAD for RO attributes [2, 16, 21, 23].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Step 3. Formation of BLM $\left\|ct_i^{(j)}\right\|$:

$$ct_i^{(j)} = \begin{cases} 1, if\ A_{low_i}[l] < lm_i^{(j)} < A_{up_i}[l]; \\ 0, else. \end{cases} \quad (14)$$

Step 4. Let $\xi := 0$. Next $\xi := 1$, etc.

Step 5. Splitting of the NMLM into two clusters: $\left\{CT_m^o[\xi] \mid m = \overline{1,2}\right\}$.

Step 5.1. Initial original standard vectors for anomalies or C-A's attributes $\{ct_m\}$ for CT_m^o are calculated:

$$cr(ct_1 \oplus ct^0) \to \min,\ cr(ct_2 \oplus ct^1) \to \min\ \&\ cr(ct_1 \oplus ct_2) \to \max, \quad (15)$$

where ct^0, ct^1 are zero and unity vectors.

Step 5.2. $$cr_m[\xi] := 0,\ n_m := 0, \quad (16)$$

where n_m is the number of realizations of anomalies or C-A's, which belong to CT_m^o.

Step 5.3. RO implementations, belonging to clusters $CT_m^o[\xi]$:

$$ct_i \in CT_1^o[\xi],\ if\ cr(ct_i \oplus ct_1) <= \\ <= cr\ \&\ cr(ct_i \oplus ct_1) < (ct_i \oplus ct_2); \quad (17)$$

$$ct_i \in CT_2^o[\xi],\ if\ cr(ct_i \oplus ct_2) <= \\ <= cr\ \&\ cr(ct_i \oplus ct_2) < (ct_i \oplus ct_1); \quad (18)$$

where $ct_i \mid i = \overline{1,N}$ are the implementations of BLM $\left\|ct_i^{(j)}\right\|$.

Step 5.4. Calculation of ICFE:

$$\overline{CE}^* = (1/M) \cdot \sum_{m=1}^{M} \max_{\{ls\}} CE_c, \quad (19)$$

where CE_c is the value of ICFE of ASR learning for the realization of class of anomalies or C-A's; $\{ls\}$ is the set of steps for ASR learning.

Step 5.5. Rule for defining coordinates:

$$ct_{m,i} = \begin{cases} 1, & if\ \frac{1}{n}\sum_{j=1}^{n} cr_{m,i}^{(j)} > 1/2; \\ 0, else. \end{cases} \quad (20)$$

Step 5.6. Conditions verification:

$$\begin{cases} if \quad N' = \sum_{m=1}^{M} n_m < N\ then \to \\ \to Step\ 5.7\ \&\ Step\ 5.3\ else\ Step\ 5.9. \end{cases} \quad (21)$$

Step 5.7. Conditions verification:

$$\begin{cases} if \quad cr_m[\xi] < cr(ct_1 \oplus ct_2)\ then \to \\ \to Step\ 5.8\ \&\ Step\ 5.3 \\ else\ Step\ 5.9. \end{cases} \quad (22)$$

Step 5.8. $cr_m[\xi] := cr_m[\xi] + 1$.

Step 5.9. Calculation of ICFE (*Step 5.4.*) for conditions:

$$N' = \sum_{m=1}^{M} n_m < N, \quad (23)$$

where N' is the number of RO implementations that belong to RC_ξ & $cr_m[\xi] < cr(ct_1 \oplus ct_2)$.

Step 6. $\xi := \xi + 1$.

Step 7. Splitting of the non-classified multi-dimensional matrix into three clusters: $\left\{CT_m^o[\xi] \mid m = \overline{1,3}\right\}$.

Step 7.1. Calculation of binary learning matrices for cluster CT_3^o:

$$cr(ct_1 \oplus ct_3) \to \min\ \&\ cr(ct_2 \oplus ct_3) \to \min, \quad (24)$$

where ct_1 & ct_2 are the standard realizations of clusters $\left\{CT_m^o \mid m = \overline{1,2}\right\}$.

Step 7.2. $cr_3[\xi] := 0$.

Step 7.3. Conditions verification:

$$ct_i \in CT_3^o\ if\ cr(ct_i \oplus ct_3) <= cr\ \& \\ cr(ct_i \oplus ct_3) <= cr(ct_i \oplus ct_1)\ \& \\ cr(ct_i \oplus ct_3) <= cr(ct_1 \oplus ct_2),$$

where $ct_i \mid i = \overline{1,N}$ are the implementations of $\left\|ct_i^{(j)}\right\|$.

Step 7.4. Calculation of radius of container $\left\{CT_m^o\right\}$:

$$cr_m[\xi] := cr_m[\xi] - 1. \quad (25)$$

Step 7.5. Calculation of ICFE – expression (19).

Step 7.6. Rule for defining coordinates – expression (20).

Step 7.7. Conditions verification:

$$\begin{cases} if\ cr_3[\xi] < cr(ct_1 \oplus ct_3)\ \&\ cr_3[\xi] < \\ < cr(ct_2 \oplus ct_3)\ then \to Step\ 7.8; \\ else\ Step\ 7.9. \end{cases}$$

Step 7.8. $cr_3[\xi] := cr_3[\xi] + 1$.

Step 7.9. Conditions verification:

$$cr_3[\xi] < cr(ct_1 \oplus ct_3)\ \& \\ cr_3[\xi] < cr(ct_2 \oplus ct_3)\ \& \\ \begin{cases} if \quad ca[l] \le 0{,}5 \cdot ca_{low}\ then \to Step\ 2 \\ else\ Step\ 9. \end{cases}$$

Step 8. Conditions verification:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\begin{cases} if \quad \overline{CE}[l] \notin IS_{CE} \; then \to Step\ 9 \\ else\ Step\ 2. \end{cases}$$

Step 9. Calculation of ca^* – expression (7).

Step 10. Calculation of $A_{low_i}^{op}$ and the $A_{up_i}^{op}$:

$$A_{low_i}^{op} = lm_i - ca^{op}\frac{ca_{low_i}}{100}; \quad (26)$$

$$A_{up_i}^{op} = lm_i + ca^{op}\frac{ca_{up_i}}{100}. \quad (27)$$

Step 11. Splitting of the non-classified multi-dimensional matrix into four clusters: $\left\{CT_m^o[\xi] \mid m = \overline{1,4}\right\}$

Step 11.1. Calculation of binary matrix of cluster (BMC) $\left\{CT_4^o\right\}$. Conditions verification:

$$cr\left(ct_1 \oplus ct_4\right) \to \min, \quad (28)$$

$$cr\left(ct_2 \oplus ct_4\right) \to \min \; \& \quad (29)$$

$$cr\left(ct_3 \oplus ct_4\right) \to \min. \quad (30)$$

Step 11.2. $cr_4[\xi] := 0$.

Step 11.3. Rule:

$$ct_i \in CT_4^o, if\ cr\left(ct_i \oplus ct_4\right) <= cr_4[\xi], \quad (31)$$

where $ct_i \mid i = \overline{1, N_4}$ are the implementations of BLM $\left\|ct_i^{(j)}\right\|$.

Step 11.4. Calculation of ICFE – expression (19).

Step 11.5. Rule for defining coordinates – expression (20).

Step 11.6. Conditions verification:

$$\begin{cases} if\ cr_4[\xi] < cr\left(ct_1 \oplus ct_4\right), \\ cr_4[\xi] < cr\left(ct_2 \oplus ct_4\right), \\ cr_4[\xi] < cr\left(ct_3 \oplus ct_4\right) \quad then \to \\ \to Step\ 7.3\ \&\ Step\ 7.8; \\ else \quad Step\ 7.9. \end{cases} \quad (32)$$

Step 11.7. $ct_4 := ct_4 + 1$.

Step 12. Calculation cr^{opt}. At conditions:

$$cr_4[\xi] < cr\left(ct_1 \oplus ct_4\right), \quad (33)$$

$$cr_4[\xi] < cr\left(ct_2 \oplus ct_4\right), \quad (34)$$

$$cr_4[\xi] < cr\left(ct_3 \oplus ct_4\right). \quad (35)$$

Step 13. Stages of algorithm of VAD formation for the attributes of recognition of anomalies or C-A's.

Step 13.1. Repeat step 1.

Step 13.2. Repeat step 2 (Expressions (12) & (13)).

Step 13.3. Repeat step 3 (Expression (14)).

Step 13.4. Repeat step 5.5 (Expression (20)).

Step 14. Restoration of container for CT_m^o.

Step 14.1. $m := 0$ & $m := m + 1$. Also $cr := 0$ & $cr := cr + 1$.

Step 15. Calculation of ICFE – expression (19).

Step 16. Conditions verification:

$$\begin{cases} if\ CE_m \notin IS_{CE} \quad then \to Step\ 14.1 \\ else\ Step\ 17. \end{cases} \quad (36)$$

Step 17. Calculation of ICFE – expression (19).

Step 18. Calculation of global maximal of ICFE:

$$CE_m^*[l] := \underset{\{cr\}}{extrem} CE_m[l, cr]. \quad (37)$$

Step 19. Calculation of optimal RC of RO class CT_m^o.

$$cr_m^*[l] := \arg \underset{\{cr\}}{extrem} CE_m[l, cr]. \quad (38)$$

Step 20. Conditions verification:

$$\begin{cases} if\ m \notin M \quad then \to Step\ 14.1 \\ else\ Step\ 21. \end{cases} \quad (39)$$

Step 21 Calculation.

$$\overline{CE}_{cp} = (1/M) \cdot \sum_{m=1}^{M} \max_{\{ls\}} CE_c. \quad (40)$$

Step 22. Conditions verification:

$$\begin{cases} if\ ca[l] \le ca_{low}/2\ then \to Step\ 13.2 \\ else\ Step\ 23. \end{cases} \quad (41)$$

Step 23. Conditions verification:

$$\begin{cases} if\ \overline{CE} \notin IS_{CE}\ then \to Step\ 24 \\ else\ Step\ 18\ \&\ Step\ 19. \end{cases} \quad (42)$$

Step 24. Calculation of ca^*. Expression (7).

Step 25. End of algorithm operation.

The algorithms (12) – (42) were implemented in the PTC Mathcad Prime 4.0, MATLAB & Simulink [1, 7, 10, 16, 18].

Multidimensional binary learning matrices of anomalies or C-A's classes had from 150 to 200 implementations [8, 20–25]. For the of network C-A's the number of recognition attributes made up 12–41 [7–23], for virus attacks, 3–15 [5, 7] attributes.

5. RESULTS AND DISCUSSION

Fig. 1–5 shows dependences of ICFE learning of simulation model (SM) of ASR [23–25] on radius of container of recognition objects – cr. In Fig. 1–5 the middle section (*Workspace – Orange color*) corresponds to the operation area of the selected

Journal of Theoretical and Applied Information Technology
15th November 2017. Vol.95. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recognition attributes that have the highest informativeness indicator (ICFE) [23–25].

Fig. 6 shows results, obtained in the course of simulation modeling and testing of algorithms of parallel clustering and formation of reference deviations for the recognition attributes, on the example of a unauthorized access to a computer system (class of C-A's). Results of the clustering of attack attributes in the process of testing the improved algorithm and the formation of VAD are shown in blue color. Similar results were also obtained for other classes of anomalies and C-A'S.

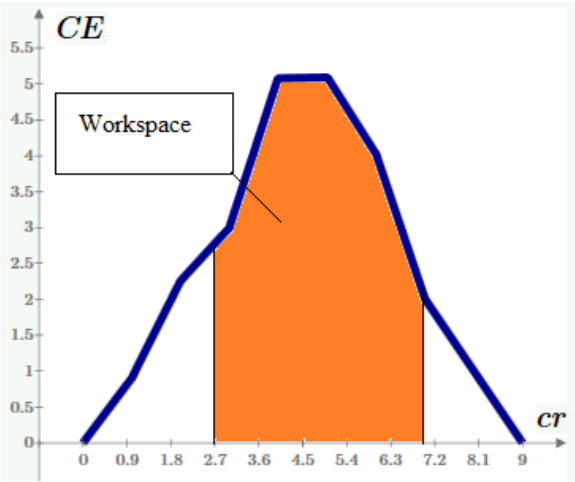


Figure 1: Information condition of functional effectiveness for virus attacks

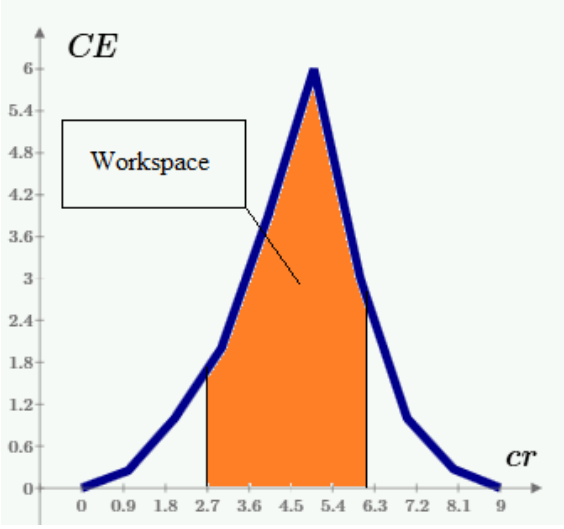


Figure 2: Information condition of functional effectiveness for the DoS/DDoS attacks

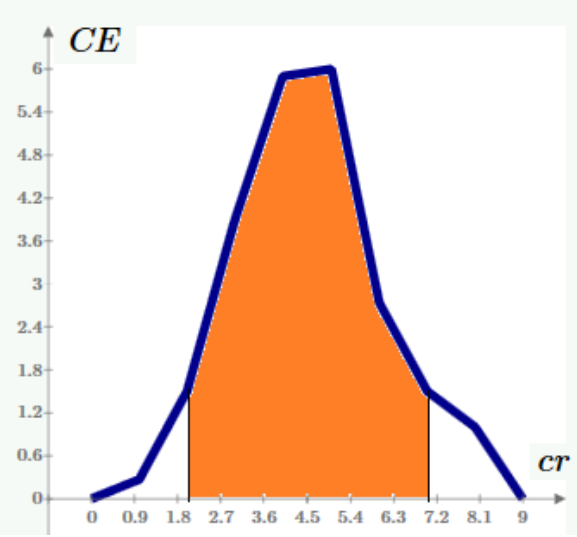


Figure 3: Information condition of functional effectiveness for the Probe attacks

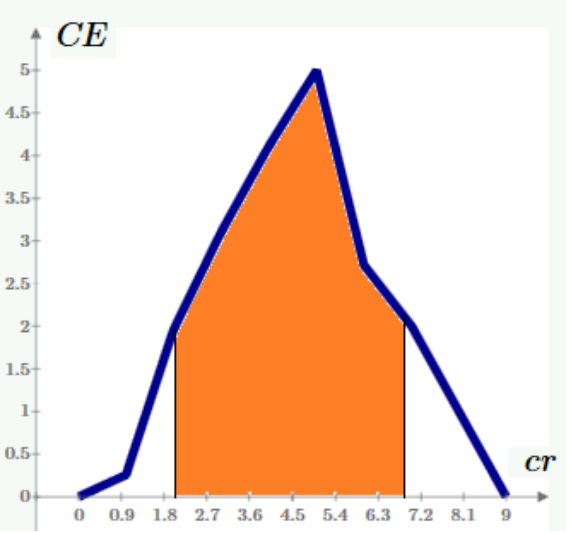
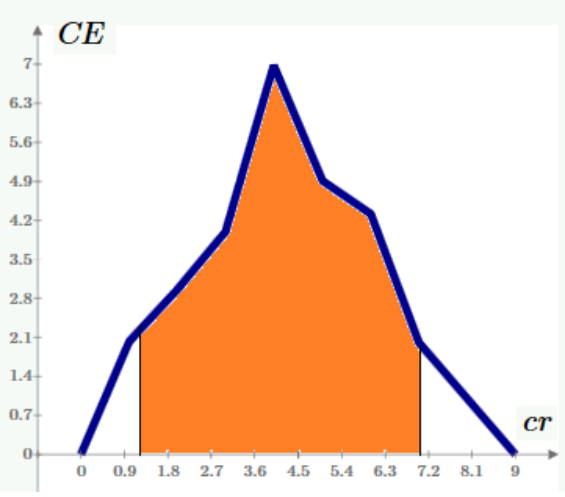


Figure 4: Information condition of functional effectiveness for the R2L attacks



Journal of Theoretical and Applied Information Technology
15th November 2017. Vol.95. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 5: Information condition of functional effectiveness for the U2R attacks

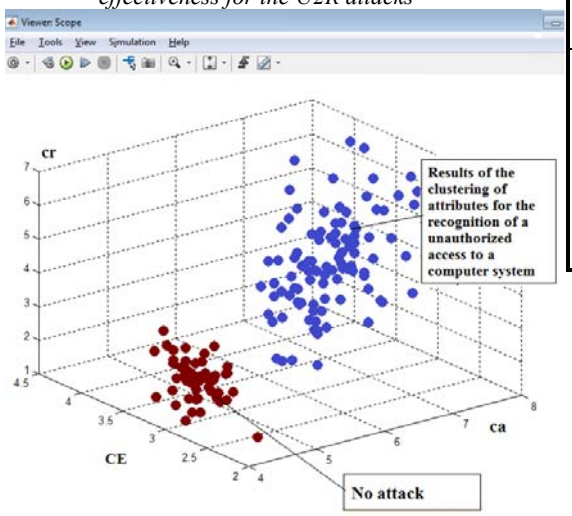


Figure 6: Results of the stages of parallel clustering and formation of VAD for the recognition of attributes

Tables 1 and 2 shows the results of a simulation modelling of values of optimal RC cr for the examined simulation models of ASR learning.

Table 1: Results of the simulation experiment

Parameter	Class of cyber attack (C-A)				
	Virus	DDoS	Probe	R2L	U2R
The averaged **max** value of ICFE of ASR learning is equal to $\overline{CE}$	2,56–2,61	3,19–3,21	3,15–3,17	2,84–2,87	3,27–3,3

The developed model compared with the results obtained for the models presented in works [2, 4, 7, 12, 19, 23, 27–29], provide significantly fewer required signs for classification of complex targeted C-A's.

Prospects of further research lie in improving the signs knowledge base and conducting model research on more objects stored in knowledge and databases of the adaptive expert system (AES).

Table 2: Values of optimal RC cr for the examined simulation models of ASR learning

N	Accepted hypotheses for RO	Values of optimal RC *cr*				
		Virus attacks	DoS/DDoS	Probe	R2L	U2R
Basic hypotheses ($hy_{\gamma i}$)						
1	$hy_{\gamma 1}$: attributes rc_i of RO and IE is within the normal state of CICS	cr_1^{opt} =5–6	cr_1^{opt} =4–5	cr_1^{opt} =3–4	cr_1^{opt} =4–5	cr_1^{opt} =4–5
2	$hy_{\gamma 2}$ – attributes allows drawing a conclusion that IE is lower than the norm	cr_2^{opt} =2–3	cr_2^{opt} =2–3	cr_2^{opt} =1–2	cr_2^{opt} =1–2	cr_2^{opt} =1–2
3	$hy_{\gamma 3}$ allows drawing a conclusion that IE is higher than the norm	cr_3^{opt} =3–4	cr_3^{opt} =3–4	cr_3^{opt} =3–4	cr_3^{opt} =2–3	cr_3^{opt} =2–3
Additional hypotheses for simulation model ($hy_{\gamma i}^{D}$)						
4	$hy_{\gamma 1}^{D}$ – node of CICS demonstrates increased network activity	–	cr_{D1}^{opt} =4	cr_{D1}^{opt} =4	cr_{D1}^{opt} =3	cr_{D1}^{opt} =3
5	$hy_{\gamma 2}^{D}$ – node of CICS demonstrates increased activity during external traffic	–	cr_{D2}^{opt} =3	cr_{D2}^{opt} =3	cr_{D2}^{opt} =3	cr_{D2}^{opt} =2
Note: Indicator IE (characterizes stability of CICS functioning [16, 18, 23–25]) is within the normal state of CICS.						

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6. CONCLUSIONS

The algorithm of the partitioning of the space of attributes into clusters in the course of implementation of the procedure for recognizing anomalies and C-A's in the CICS was improved.

The refined algorithm differs from the existing ones by simultaneous formation of control admittance during the analysis of complex signs of ORS. Thus, at each step of the training, you can change the permissible deviations for all signs simultaneously. The advantage of the modified algorithm is to prevent possible cases of absorption by one ORS class of the basic signs of anomalies and C-A's of the other class. There are predicate expressions for ARS capable of self-learning in the present research. Different researches on the efficiency and adequacy of the developed algorithms on the simulation models in Mathcad Prime 4.0, MATLAB & Simulink were conducted.

It is confirmed that the proposed clustering algorithms of ORS characteristics allow to increase the effectiveness of cybersecurity systems of CICS.

7. ACKNOWLEDGEMENT

The authors acknowledge the financial supported by the Fundamental Research Grant under grant number 0114U005430 received from the Ministry of Education and Science of Ukraine.

REFERENCES:

[1] L. Khan, M. Awad, B. Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering", *The International Journal on Very Large Data Bases*, Vol. 16, Iss. 4, 2007, pp. 507–521.

[2] R. Ranjan, G. Sahoo, "A new clustering approach for anomaly intrusion detection", *International Journal of Data Mining Knowledge Management Process*, Vol. 4, Iss. 2, 2014, pp. 29–38.

[3] J.S. Joseph, "Cybercrime: Legal Standards Governing the Collection of Digital Evidence", *Information Systems Frontiers*, Vol. 6, Iss. 2, 2004, pp. 133–151.

[4] T. Mahmood, U. Afzal, "Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools", *Information Assurance* (NCIA), 2013 2nd National Conference.

[5] S. Dua, X. Du, "Data Mining and Machine Learning in Cybersecurity", *CRC press*, 2016, p. 225.

[6] S. Zhang, D. Caragea, X. Ou, "An Empirical Study on Using the National Vulnerability Database to Predict Software Vulnerabilities", *Database and Expert Systems Applications. DEXA 2011, Lecture Notes in Computer Science*, Vol., 6860, 2011.

[7] K. C. Lee, C. H. Hsieh, L.J. Wei, "Sec-Buzzer: cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation", *Soft Computing*, 2016, pp. 1–14.

[8] A. Buczak, E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection", *IEEE Communications Surveys & Tutorials*, Vol. 18, Iss. 2, 2016, pp. 1153 – 1176.

[9] J. Petit, S. Shladover, "Potential Cyberattacks on Automated Vehicles", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, Iss. 2, pp. 546 – 556.

[10] M. Al. Hadidi, Y. Ibrahim, V. Lakhno, A. Korchenko, A. Tereshchuk, A. Pereverzev, "Intelligent Systems for Monitoring and Recognition of Cyber Attacks on Information and Communication Systems of Transport", *International Review on Computers and Software*, 11(12), 2016, pp. 1167-1177.

[11] A. S. Dovbysh, V. V. Moskalenko, A. S. Rizhova, "Learning Decision Making Support System for Control of Nonstationary Technological Process", *Journal of Automation and Information Sciences*, Vol. 48, Iss. 6, 2016, pp. 39–48.

[12] Ali M. Ameer, G. C. Karmakar, L. S. Dooley, "Review on Fuzzy Clustering Algorithms", *IETECH Journal of Advanced Computations*, Vol. 2, Iss. 3, 2008, pp. 169–181.

[13] Y. Guan, A.A. Ghorbani, N. Belacel, "Y–means: a clustering method for intrusion detection", *Canadian Conference on Electrical and Computer Engineering*, No. 2, 2003, pp. 1083–1086.

[14] M. Halkidi, Y. Batistakis, M. Vazirgiannis, "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, Vol. 17, Iss. 2, 2001, pp. 107–145.

[15] M.M. Gamal, B. Hasan, A.F. Hegazy, "A Security Analysis Framework Powered by an Expert System", *International Journal of Computer Science and Security*, Vol.4, Iss. 6, 2011, pp. 505–527.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[16] A. Petrov, V. Lakhno, A. Korchenko, "Models, Methods and Information Technologies of Protection of Corporate Systems of Transport Based on Intellectual Identification of Threats", *Decision Making in Manufacturing and Services*, Vol. 9, no. 2, 2015, pp. 19-37.

[17] I. Riadi, J.E. Istiyanto, A. Ashari, N. Subanar, "Log Analysis Techniques using Clustering in Network Forensics", *International Journal of Computer Science and Information Security*, Vol. 10, Iss. 7, 2013, pp. 740–749.

[18] V. Lakhno, Y. Tkach, T. Petrenko, S. Zaitsev, V. Bazylevych, "Development of adaptive expert system of information security using a procedure of clustering the attributes of anomalies and cyber attacks", *Eastern-European Journal of Enterprise Technologies*, Vol. 6/9, Iss. 84, 2016, pp. 32–44.

[19] I. Kiss, B. Genge, P. Haller, P. "A clustering-based approach to detect cyber attacks in process control systems", *IEEE 13th International Conference*, 2015.

[20] A.S. Dovbysh, N.N. Budnik, V.V. Moskalenko, "Informatsionno-ekstremalnyy algoritm optimizatsii parametrov giperellipsoidnykh konteynerov klassov raspoznavaniya", *Mezhdunarodnyy nauchno-tekhnicheskiy zhurnal «Problemy upravleniya i informatiki»*, No. 5, 2012, pp. 111–119.

[21] S.M. Lee, D. Se. Kimb, J. H. Lee, J. S. Park, "Detection of DDoS attacks using optimized traffic matrix", *Computers and Mathematics with Applications*, No 63, 2012, pp. 501–510.

[22] P. Gao, M. Wang, J. H. Chow, "Identification of Successive "Unobservable" Cyber Data Attacks in Power Systems Through Matrix Decomposition", *IEEE Transactions on Signal Processing*, 64(21), 2016, pp. 5557 – 5570.

[23] V.A. Lakhno, T.A. Petrenko, M. V. Pirog, "Modelirovanie raboty adaptivnoy sistemy raspoznavaniya kiberatak v usloviyakh neodnorodnykh potokov zaprosov v modulyakh e-business", *Ukrainian Scientific Journal of Information Security*, Vol. 22, Iss. 2, 2016, pp. 135–142.

[24] V.A. Lakhno, P. U. Kravchuk, D.B. Mekhed, H.A. Mohylnyi, V.U. Donchenko, "Development of a support system for managing the cyber protection of an information object", *Journal of Theoretical and Applied Information Technology*, Vol. 95, No 6, 2017, pp. 1263–1272.

[25] B. Akhmetov, V. Lakhno, Y. Boiko, A. Mishchenko, "Designing a decision support system for the weakly formalized problems in the provision of cybersecurity", *Eastern-European Journal of Enterprise Technologies*, Vol. 1/2, Iss. 85, pp. 4–15.

[26] C. Callegari, L. Gazzarrini, S. Giordano, M. Pagano, T. Pepe, "Improving PCA-based anomaly detection by using multiple time scale analysis and Kullback–Leibler divergence", *International Journal of Communication Systems*, Vol. 27, Iss. 10, 2014, pp. 1731–1751.

[27] J. Li, Z. Zhao, & R. Li, "A Machine Learning Based Intrusion Detection System for Software Defined 5G Network", *arXiv preprint arXiv:1708.04571,* 2017.

[28] Y. Dang, B. Wang, R. Brant, Z. Zhang, M. Alqallaf, & Z. Wu. "Anomaly Detection for Data Streams in Large-Scale Distributed Heterogeneous Computing Environments", In *ICMLG2017 5th International Conference on Management Leadership and Governance* (pp. 121). Academic Conferences and publishing limited, 2017.

[29] V. Lahno. "Ensuring of information processes' reliability and security in critical application data processing systems", *MEST Journal*, 2(1), 2014, pp. 71–79.