MEST Journal
DOI 10.12709/mest.05.05.01.10

DESIGN OF ADAPTIVE SYSTEM FOR DETECTION OF CYBER-ATTACKS

Taras Petrenko
Chernihiv National University of Technology, Mathematical Simulation and Cybersecurity Department, Chernihiv, Ukraine

JEL Category: **C02**

Abstract

This paper is devoted to the improvement of the mathematical support for the intelligent detection models of cyber threats. The results of the research present the further development of detection models of cyber threats, as well as of common classes of cyber-attacks in mission critical information systems (MCIS). There was the model of detection of cyber-attacks to MCIS designed, which is based on the application of learning samples in the form of matrices of features for each of the modeled classes. The studies on minimization of the number of training samples, represented by a binary form of discerning features were carried out. There was the program "Cyber-attacks Analyzer" developed, which allows the automatic generation of dimensions of training matrix of cyber-attacks features, without requiring the participation of experts. It is shown that for the object detection within known classes of cyber-attacks the usage of representative sets of 3-4 features long in the training matrices increases the effectiveness of the algorithm, reaching up to 95%.

Keywords: *modeling, training matrices, adaptive system of detection of cyber threats, information systems, information security*

1 INTRODUCTION

Incomplete information about threats to information security and cyber security (CS) of mission critical information systems (MCIS) is twofold. Firstly, it is a partial lack of prior information, even at the level of the object structure to attack information, which has, as a rule, stochastic nature. Secondly, it is the limited ability of observation of the object of recognition and attack threats, which belong to a particular class. In the extreme case, it is previously known only to the total set of IS threats and ways to implement them.

However, in practice, one of the main characteristics of today's threats is that they are not activated for a long time, sometimes for two or three years (Ranjan & Sahoo, 2014)(Lakhno, 2016). The targeted attacks, are particularly aimed to IS of enterprises, infrastructure, energy, transport, etc. IS are usually tailored to the environment in which they will be targeted. Active expansion of information-communication environment of the mission critical information systems (IC MCIS), especially in the segment of mobile, distributed and wireless technologies, is

Address of the author:
Taras Petrenko
pta1983@bk.ru

accompanied by the emergence of new threats to cyber security (CS). It is confirmed by the increasing number of incidents connected to information security, and also, detection of new vulnerabilities in the information systems (IS) and the automated control systems (ACS) (Ahmad, Dubrovskiy, & Flinn, 2005) (Chi, Park, Jung, & Lee, 2001).

The purpose of the study is to design a model for training the adaptive system of detection of cyber-attacks (ASDCA), which is being developed, based on the use of the apparatus of logical functions. The model allows taking into account the hard-to-explain features of threats, attacks, and anomalies in the IC MCIS, and it also reduces the time required for training ASDCA under conditions of the increase in the number of cyber threats.

2 THE MODEL OF LOGICAL PROCEDURES OF DETECTION OF ANOMALIES AND CYBER-ATTACKS

In a general case, the problem of detection of cyber-attacks to MCIS boils down to the following (Omar, Ngadi, & Jebur, 2013) (Tsai, Hsub, Linc, & Lin, 2009). A certain set of objects is explored; in our case, this is NPT − the number of possible targets from the side that attacks MCIS. The objects of this set are described by the features $\{s_{ax1},...,s_{axn}\}$, represented, for example, in a binary form. It is known that the set of NPT is displayed in the form of the combination of disjoint subsets (classes) of cyber threats to MCIS – $(CT_1,...,CT_l)$. Let us assume that there is a finite set of objects $\{ss_{a1},...,ss_{am}\}$ from NPT, whereof we know which classes of anomalies, attacks or threats they belong in (these are precedents, i.e. the objects used for training, – OUT). It is required, based on a set of values of features, specified in the OUT, i.e. the description of a certain object ss_{an} from NPT, to identify this class and to adjust the performance of ASDCA for MCIS, accordingly. It is not known in advance, to which class the object can be attributed to (Zhan, Xu, & Xu, 2013) (Baddar, 2014).

A distinctive feature of the logical procedures examined in the work is the ability to obtain a reliable result when there is no a priori information about the function of the distribution of existing values of features of a threat, cyber-attack or anomaly. Hereinafter we shall refer to such procedures as logical procedures. And there is no need to specify the so-called metrics in the space of object descriptions, characterizing each class. Therefore, for each feature of a cyber-attack, a binary function of the similarity between its values is defined, allowing distinguishing objects and their representations (sub-descriptions).

As the informative fragments, it is advisable to use only those fragments in the ASDCA that reflect typical patterns in the descriptions of the objects used for training (OUT). Therefore, the presence (absence) of such fragments in the categorized object allows determining its belonging to the class. When the logical procedures of detection of cyber-attacks (LPDCA) are applied, we also accept as informative those fragments that are found in the descriptions of the objects of the same class of cyber-attacks but missing from the descriptions of objects from other classes. The fragments used include also a meaningful description of the OUT in terms of designing ASDCA (Chertov, Fahmy, S, & Shroff, 2006) (Zhou, 2009).

The algorithms of the synthesis of workable implementations for LPDCA depend directly on the success of the research of metrical (quantitative) properties of many informative fragments, i.e. the features of a cyber-attack (cyber threat, anomaly, vulnerability). And it is necessary to transform the incoming uncategorized training matrix (OUT) into a categorized one and to design, in a training mode, a clear division of the features space of detection into the classes of detection $CT_m^0 \big| m=\overline{1,M}$, where *M* is the power of the alphabet of classes.

Technically, it appears difficult to implement the following tasks in ASDCA (Harel, 1987)(Gorodetski & Kotenko, 2002); (Lin & Tseng, 2004):

1. to calculate the asymptotic estimate of the number of blind coverings for integer matrix of the object's features;
2. to calculate the asymptotic estimate of accepted and maximum values of conjunctions of Boolean function that can be

applied to the synthesis of schematic-technical solutions of the ASDCA hardware for MCIS.

Let us consider the task of designing LPDCA based on the principle of "nonoccurrence" of sets of acceptable values of the features of cyber-attacks (cyber threats, anomalies, vulnerabilities).

Let us define: TN – a total number of cyber threats to MCIS; B_{s_a} – a set of numbers of cyber threats, implemented by an attacking side for achieving p_a – target of the cyber-attack; NP_{s_a} – an acceptable set of discrete features (of threat, anomaly, cyber-attack, etc.) in the $\{s_{a_1},...,s_{a_{jQ}}\}$ form.

The algorithm for calculating the value (ACV) of the significance of a feature for ASDCA can be presented as follows. Let us define the combination of subsets of $NP_{s_a} = \{s_{aj_1},...,s_{a_Q}\}$, $r_{p_a} \le TN$ in the system of the features of OUT. We assume the subsets defined being the reference for ACV. Their total combination is ΩTN.

Let us assign additional parameters: po_{ss_a} – the significance of the target of an attack (object) ss_{ai}, $i = 1,2,..., NPT$; $po_{NP_{sa}}$ – the significance of the object of the referent set $NP_{s_a} \in \Omega TN$.

Let us calculate for each class of cyber-attacks on MCCS $CT \in \{CT_1,...,CT_l\}$, the value of belonging $E(ss_a, CT)$ of the object ss_a to the class CT, which has the form:

$$E(ss_a, \mathrm{CT}) = \frac{1}{|\mathrm{LW_{CT}}|} \cdot \sum_{ss_{ai} \in CT} \sum_{NP_{s_a} \in \Omega TN} po_{ss_a} \cdot po_{NP_{sa}} \cdot BN), \quad (1)$$

where $|LW_{CT}| = |CT \cap \{ss_{a1},...,ss_{aQ.}\}|$, BN is the similarity of objects ss'_a and ss''_a.

Let us define as MC – combination of all elementary classifiers (EC), which were obtained by the totality of features from $\{s_{ax1},...,s_{axn}\}$, i.e. $MC = (\sigma_{DOP}, NP_{sa})$, where $NP_{sa} \subseteq \{s_{ax1},..., s_{axn}\}$,

$$\sigma_{\mathrm{DOP}} = (\sigma_{\mathrm{DOP_1}},...,\sigma_{\mathrm{DOP_r}}),$$

$$\sigma_{DOP_i} \in NP_{s_{aj}},\ for\ i = 1,2,..., r_{s_a}.$$

Let us suppose that a series Z of measurements of the values of the controlled features in MCIS was performed, and we received the matrix of features:

$$S = \begin{pmatrix} s_{ax_{11}} & s_{ax_{12}} & \dots & s_{ax_{1i}} & \dots & s_{ax_{1n}} \\ s_{ax_{21}} & s_{ax_{22}} & \dots & s_{ax_{2i}} & \dots & s_{ax_{2n}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ s_{ax_{l1}} & s_{ax_{l2}} & \dots & s_{ax_{li}} & \dots & s_{ax_{\mathrm{ln}}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ s_{ax_{Z1}} & s_{ax_{Z2}} & \dots & s_{ax_{Zi}} & \dots & s_{ax_{Zn}} \end{pmatrix}, \quad (2)$$

For example, the matrix of features, available in the ASDCA repository, will look like this

$$S = \begin{pmatrix} 0 & 1 & \dots & 1 & \dots & 1 \\ 1 & 0 & \dots & - & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ - & 1 & \dots & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & - & \dots & 0 \end{pmatrix}. \quad (3)$$

Thus, a set of objects to be tested, belonging to a class, is specified by the binary features {1001…–01}. The dash points to the uncertainty of a feature in OUT.

Then the procedure of detection of the object $ss_a = (ss_{a1},..., ss_{aTN})$, for example, cyber-attack in MCCS, is carried out on the basis of the results of calculation by elementary conjunctions – $\Re$. During the study, the results of which are described in the work (Khan, Awad, & Thuraisingham, 2007) (Al-Jarrah, 2014) (Abraham & Nair, 2014), it was justified that the most

economical was the variant to use the algorithm for calculating the conjunctions for coverage of the class of a corresponding object (cyber threat, vulnerability or attack).

Thus, obtaining LPDCA for the modeled class of objects (cyber threats, or cyber-attacks) is reduced to the following:

1. we set the distinctive function
2. we find disjunctive normal form (DNF) that implements this function
3. we find acceptable (maximal) conjunction $\Re$ that defines the belonging of the object in the class under consideration.

Thus, the algorithm of training ASDCA is in an iterative procedure of finding DNF for the distinctive function of the object of detection by the feature matrix (2) and minimizing the number of features, the columns and rows of the OUT matrix to its limit value, which includes acceptable (maximal) conjunction that defines the belonging of the object in the studied class of anomalies, threats and cyber-attacks.

3 THE PROGRAM OF THE SEARCH OF THE MINIMALLY NEEDED NUMBERS OF FEATURES FOR DETECTION OF CYBER-ATTACKS

In the course of the research, a program was designed for evaluation of the complexity of the search algorithm of the minimally needed number of features for different classes of cyber-attacks "Cyber-attacks Analyzer", Fig. 1–3.

Form 1 sets analyzed classes of attacks, Fig. 1. Form 2 shows the calculation results for training matrices in the form of OUT, taking into account the information content of each of the 3 – 21 features. Form 3 visualizes the results of calculation in the form of histograms, as well as the evaluation of the complexity of the algorithm of forming OUT depending on the class of an attack, Fig. 3.

The modeling allowed drawing the conclusion that the objects belonging to different classes of cyber-attacks are often difficult to separate from each other. A rather large number of features (for certain classes of cyber- attacks, up to 50%) have the information weight almost equaling zero. In the case of using a set of features for the formation of the OUT, it is advisable to reject the requirement of its futility. This is done in order to increase the speed of the algorithm.

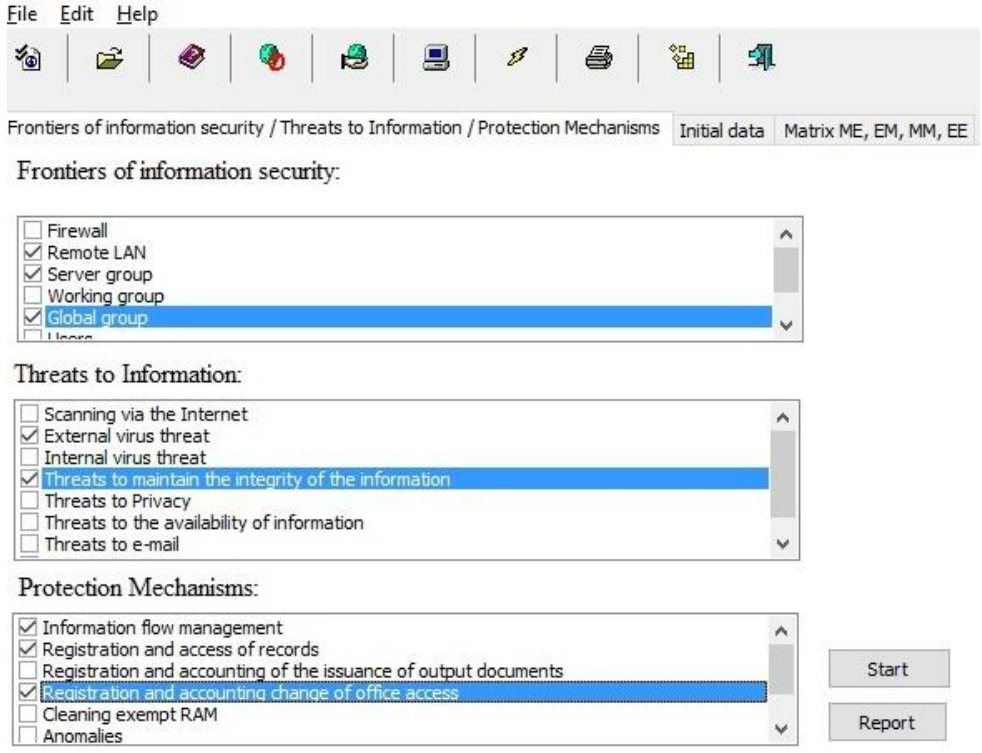

Fig. 1 The interface of the program Cyber-attacks Analyzer; Form 1

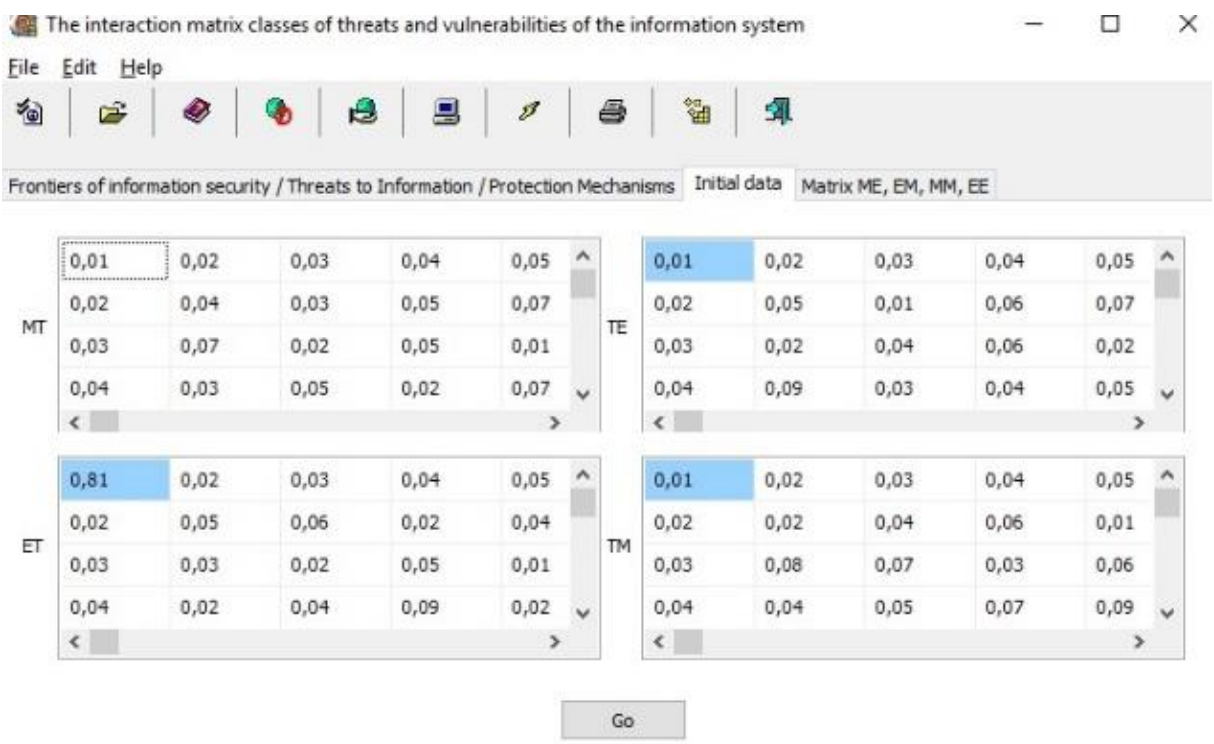

Fig. 2 The interface of the program Cyber-attacks Analyzer; Form 2

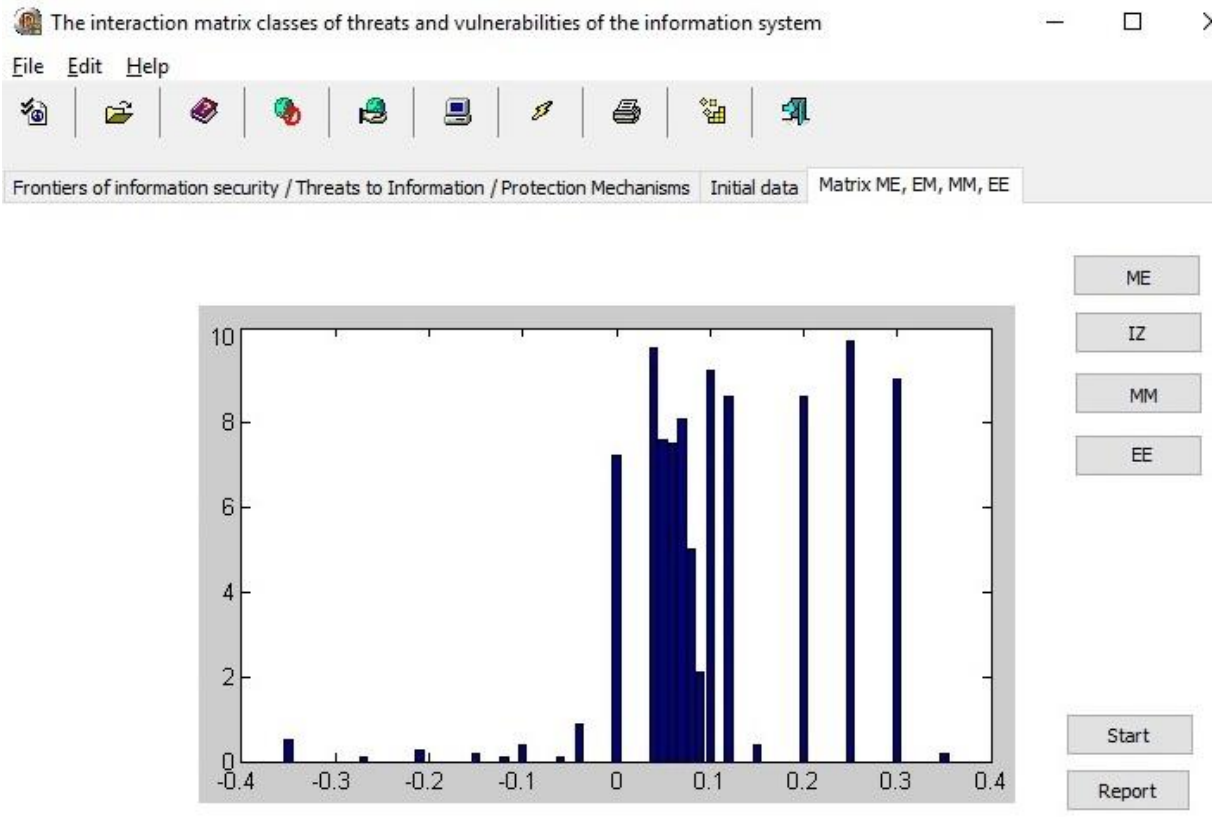

Fig. 3 The interface of the program Cyber-attacks Analyzer; Form 3

For example, in the case of an increase in the number of features from 3 to 6, the average number of checks per object ranged from 150 to 800, respectively. The use of representative sets with the length of 3–4 features in the matrices of OUT made it possible to achieve maximum efficiency of the performance of the algorithm of detection for the majority of the known cyber

attacks. In the situation, where the features of the class of an object (e.g., cyber-attack) were positioned according to the decreasing information content (I), for every object, there was a combination of features with greater information content and then the information content of the group decreased smoothly. Thus, the less meaningful features (PS <60 %) were not included in OUT.

The following feature of the matrix forming the OUT was identified: The information content of the control set formed by the two features, characteristic for different classes of attacks, such as Dos/DDos, U2R, R2L, may describe the object of detection better than each of the features and the EC class separately. And the level of detection of cyber attacks, for which the training matrices of OUT were compiled, ranged from 25% to 30% for 2 features, 85–87% for 3–4 features, 92–98% for 5–9 features, Fig. 4.

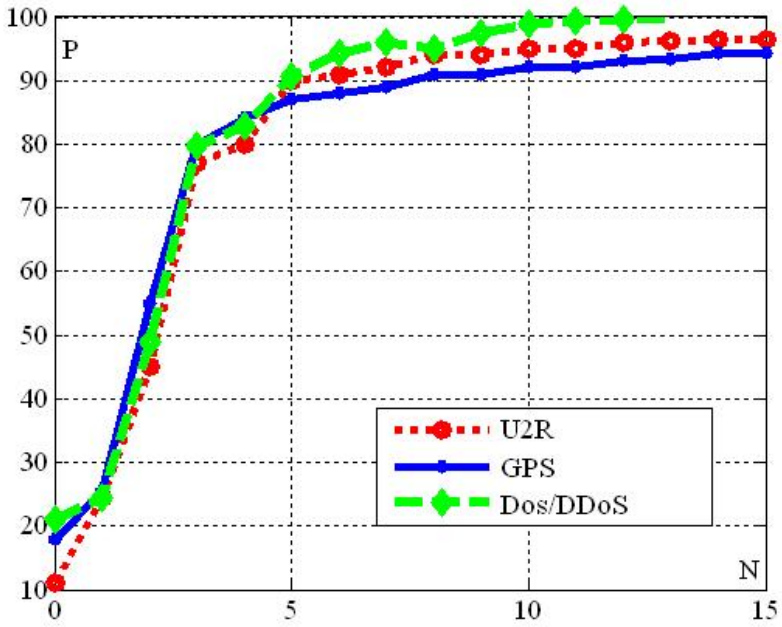

Fig. 4. Visualization of the accuracy of detection (P, %) for attacks of classes U2R, DOS/ DDoS and attacks on satellite systems of GPS, depending on the number of features (N) in the OUT training matrix

Thus the OUT, described by a fragment of 2–3 features, belonging to different classes of objects, described the studied class better than each of the features separately.

It was experimentally found that, compared to the methods of consecutive exhaustive search of features and statistical algorithms of states, the proposed model allows:

- reduction the number of necessary rules of object detection within a class by 2.5–12 times (depending on the class of cyber-attacks);
- reduction of the time of detection of cyber-attacks by 7–9%.

In the test mode of ASDCA training, the rational number of steps of training OUT for the proposed model is amounted to the known classes of objects and for the more sophisticated cyber-attacks.

4 CONCLUSIONS

As a result of the research:

– the model of detection of cyber attacks to mission critical information systems was designed, which is based on the application of training samples in the form of feature matrices for each of the modeled class;

– the studies were carried out on minimizing the number of training samples from the informative features for the ASDCA being developed. It was found that for detection in training matrices of OUT it was sufficient to use representative sets of 3–4 features long. The effectiveness of detection of anomalies and cyber-attacks reached 95%.

5 WORKS CITED

Abraham, S., & Nair, S. (2014). Cyber Security Analytics: A Stochastic Model for Security Quantification Using Absorbing Markov Chains. *Journal of Communications, 9*(12), 899-907.

Ahmad, D., Dubrovskiy, A., & Flinn, X. (2005). *Defense from the hackers of corporate networks.* Moscow: Companies AyTi; DMK - Press.

Al-Jarrah, O. A. (2014). Network Intrusion Detection System using attack behavior classification. *5th International Conference Information and Communication Systems (ICICS), 2014* , (pp. 1–6).

Baddar, S.-H. M. (2014). Anomaly detection in computer networks: a state-of-the-art review. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 5*(4), 29–64.

Chertov, R., Fahmy, S, & Shroff, N. (2006). Emulation versus Simulation: A Case Study of TCP-Targeted Denial of Service Attacks. *Proc. of the 2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities.* Retrieved from https://www.cs.purdue.edu/homes/fahmy/papers/tridentcom.pdf

Chi, S., Park, J., Jung, K., & Lee, J. (2001). *Network Security Modeling and Cyber Attack Simulation Methodology* (Vol. LNCS 2119). Berlin Heidelberg: Springer-Verlag. Retrieved from http://link.springer.com/chapter/10.1007%2F3-540-47719-5_26#page-2

Gorodetski, V., & Kotenko, I. (2002). Attacks against Computer Network: Formal Grammar-Based Framework and Simulation Tool. In *Recent Advances in Intrusion Detection* (Vol. 2516, pp. 219-238). Springer Berlin Heidelberg. doi:10.1007/3-540-36084-0_12

Harel, D. (1987). Statecharts: A Visual Formalism for Complex Systems. *Science of Computer Programming*(8), 231-274.

Khan, L., Awad, M., & Thuraisingham, B. (2007). A new intrusion detection system using support vector machines and hierarchical clustering. *The International Journal on Very Large Data Bases, 16*(4), 507–521. Retrieved from https://www.utdallas.edu/~lkhan/papers/Intrusion%20Detection%20Using%20Clustering%20Approaches.pdf

Lakhno, V. (2016). Creation of the adaptive cyber threat detection system on the basis of fuzzy feature clustering. *Eastern-European Journal of Enterprise Technologies, 2*(9(80)), 18-25. doi:10.15587/1729-4061.2016.66015

Lin, S.-C., & Tseng, S.-S. (2004, Oct). Constructing detection knowledge for DDoS intrusion tolerance. *Expert Systems with Applications, 27*(3), 379–390.

Omar, S., Ngadi, A., & Jebur, H. (2013). Machine learning techniques for anomaly detection: an overview. *International Journal of Computer Applications, 79*(2), 33–41.

Ranjan, R., & Sahoo, G. (2014). A new clustering approach for anomaly intrusion detection. *International Journal of Data Mining Knowledge Management Process (IJDKP), 4*(2), 29–38. doi: 10.5121/ijdkp.2014.4203

Tsai, C.-F., Hsub, Y.-F., Linc, C.-Y., & Lin, W.-Y. (2009). Intrusion detection by machine learning: a review. *Expert Systems with Applications, 36*(10), 11994–12000. doi:10.1016/j.eswa.2009.05.029

Zhan, Z., Xu, M., & Xu, S. (2013). Characterizing Honeypot-Captured Cyber Attacks: Statistical Framework and Case Study. *IEEE Transactions on Information Forensics and Security, 8*(11), 1775-1789. doi:10.1109/TIFS.2013.2279800

Zhou, Y. (2009). Hybrid Model Based on Artificial Immune System and PCA Neural Networks for Intrusion Detection. *Asia-Pacific Conference on Information Processing*, *1*, pp. 21-24. doi:10.1109/APCIP.2009.13

Received for publication: 06.05.2016
Revision received: 11.07.2016
Accepted for publication: 12.12.2016

How to cite this article?

Style – **APA** *Sixth Edition:*

Petrenko, T. (2017, Jan 15). Design of adaptive system for detection of cyber-attacks. (Z. Cekerevac, Ed.) *MEST Journal, 5*(1), 78-85. doi:10.12709/mest.05.05.01.10

Style – **Chicago** *Sixteenth Edition:*

Petrenko, Taras. "Design of adaptive system for detection of cyber-attacks." Edited by Zoran Cekerevac. *MEST Journal* (MESTE) 5, no. 1 (Jan 2017): 78-85. doi:10.12709/mest.05.05.01.10

Style – **GOST** *Name Sort:*

Petrenko Taras Design of adaptive system for detection of cyber-attacks [Journal] // MEST Journal / ed. Cekerevac Zoran. - Belgrade - Toronto : MESTE, Jan 15, 2017. - 1 : Vol. 5. - pp. 78-85.

Style – **Harvard** *Anglia:*

Petrenko, T., 2017. Design of adaptive system for detection of cyber-attacks. *MEST Journal,* 15 Jan, 5(1), pp. 78-85.

Style – **ISO 690** *Numerical Reference:*

Design of adaptive system for detection of cyber-attacks. **Petrenko, Taras.** [ed.] Zoran Cekerevac. 1, Belgrade - Toronto : MESTE, Jan 15, 2017, MEST Journal, Vol. 5, pp. 78-85.