УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2016 Том 16 № 2 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2016 Vol. 16 No 2 ISSN 2226-1494 http://ntv.ifmo.ru/en

INVESTIGATION OF NEURAL NETWORK ALGORITHM FOR DETECTION OF NETWORK HOST ANOMALIES IN THE AUTOMATED SEARCH FOR XSS VULNERABILITIES AND SQL INJECTIONS

Yu.D. Shabalin[a,b], V. L. Eliseev [a, c]

[a] National Research University "MPEI", Moscow, 111250, Russian Federation
[b] SberBank Technologies, Moscow, 117105, Russian Federation
[c] ISC InfoTeCS, Moscow, 127287, Russian Federation
Corresponding author: Yury.shabalin@gmail.com

Article info
Received 14.11.15, accepted 01.02.16
doi: 10.17586/2226-1494-2016-16-2-318-323
Article in English

For citation: Shabalin Yu.D., Eliseev V. L. Investigation of neural network algorithm for detection of network host anomalies in the automated search for XSS vulnerabilities and SQL injections. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 2, pp. 318–323, doi: 10.17586/2226-1494-2016-16-2-318-323

Abstract

A problem of aberrant behavior detection for network communicating computer is discussed. A novel approach based on dynamic response of computer is introduced. The computer is suggested as a multiple-input multiple-output (MIMO) plant. To characterize dynamic response of the computer on incoming requests a correlation between input data rate and observed output response (outgoing data rate and performance metrics) is used. To distinguish normal and aberrant behavior of the computer one-class neural network classifier is used. General idea of the algorithm is shortly described. Configuration of network testbed for experiments with real attacks and their detection is presented (the automated search for XSS and SQL injections). Real found-XSS and SQL injection attack software was used to model the intrusion scenario. It would be expectable that aberrant behavior of the server will reveal itself by some instantaneous correlation response which will be significantly different from any of normal ones. It is evident that correlation picture of attacks from different malware running, the site homepage overriding on the server (so called defacing), hardware and software failures will differ from correlation picture of normal functioning. Intrusion detection algorithm is investigated to estimate false positive and false negative rates in relation to algorithm parameters. The importance of correlation width value and threshold value selection was emphasized. False positive rate was estimated along the time series of experimental data. Some ideas about enhancement of the algorithm quality and robustness were mentioned.

Keywords

anomalies detection, intrusion detection, neural network, one-class neural network classifier, security, network, Cross-Site Scripting, XSS attack, SQL injection, network attack

УДК 004.492.3

ИССЛЕДОВАНИЕ НЕЙРОСЕТЕВОГО АЛГОРИТМА ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В ПОВЕДЕНИИ СЕТЕВОГО ХОСТА ПРИ АВТОМАТИЗИРОВАННОМ ПОИСКЕ XSS-УЯЗВИМОСТЕЙ И SQL-ИНЪЕКЦИЙ

Ю.Д. Шабалин[a,b], В.Л. Елисеев[a, c]

[a] Национальный исследовательский университет МЭИ, Москва, 111250, Российская Федерация
[b] СберБанк Технологии, Москва, 117105, Российская Федерация
[c] ОАО ИнфоТеКС, Москва, 127287, Российская Федерация
Адрес для переписки: Yury.shabalin@gmail.com

Информация о статье
Поступила в редакцию 14.11.15, принята к печати 01.02.16
doi: 10.17586/2226-1494-2016-16-2-318-323
Язык статьи – английский

Ссылка для цитирования: Шабалин Ю.Д., Елисеев В.Л. Исследование нейросетевого алгоритма для обнаружения аномалий в поведении сетевого хоста при автоматизированном поиске XSS-уязвимостей и SQL-инъекций // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 2. С. 318–323. doi: 10.17586/2226-1494-2016-16-2-318-323

Аннотация

Рассматривается проблема выявления аномального поведения у компьютера, участвующего в обмене данными по сети. Предлагается подход, основанный на анализе динамического отклика компьютера, рассматриваемого как многосвязный объект. В качестве характеристики динамического отклика используется корреляция входных возмущающих сетевых воздействий и выходных наблюдаемых величин, включающих исходящий сетевой трафик и потребление вычислительных ресурсов компьютера. Для распознавания нормального и аномального поведения используется одноклассовый нейросетевой классификатор. В статье представлено краткое описание алгоритма. Представлена схема стенда для проведения экспериментов с реальными атаками на стенд (автоматизированный поиск XSS и внедрение операторов SQL). Очевидно, что корреляционная картина атак от различного вредоносного программного обеспечения, подмены страниц, программных и аппаратных сбоев будет отличаться от нормальной. В заключении алгоритм обнаружения вторжений (аномалий) исследован, сделаны выводы о зависимости ошибок первого и второго рода от параметров алгоритма. Подчеркнута важность значений ширины окна корреляции и выбора порогового значения. Предложены несколько идей о дальнейшем улучшении алгоритма.

Ключевые слова

определение аномалий, обнаружение вторжений, нейронная сеть, одноклассовый нейросетевой классификатор, безопасность, сеть, межсайтовое выполнение сценариев, внедрение операторов SQL, сетевая атака

Introduction

An actual and very important security task of defending network servers from attacks is solved usually by intrusion detection systems (IDS) based on different approaches [1, 2]. One can mention signature based methods, behavioral methods, heuristics and machine learning techniques [3, 4]. Most approaches deal with known attacks and only few ones – with unknown, so called, zero-day vulnerabilities.

Comparison of host-based IDS (antivirus programs as a commonly used case) and network-based ones shows better adoption of host-based approach to unknown attacks detection. The world of network servers is quite different. Intruders can use not only leaks in system software (network stack implementation and the most basic services) but also leaks in application server software. Such software is based on Web technologies mainly and may be implemented by using many different languages, frameworks, libraries, engines and protocols. The number of potential leaks in Internet applications and their availability for hackers significantly exceeds the number of ones in host operating systems and host applications. A perspective way to solve the posed problem is to enrich host-based intrusion detection methods with network specifics. There are many different approaches to perform such task.

The simplest use of neural network to adopt knowledge from data is supervised learning. Straightforward solutions look successful [5], but they face problem of labeling of training data for real network traffic. Unsupervised learning promises classification without preliminary labeling, but preprocessing stage of gathered fixed number of features is suggested as a principle limitation, because new attack or other type of anomaly to be detected properly may need one more feature. Specifics of self-organized maps require additional methods to find out the kind of investigated sample of traffic [6].

One-class classification with neural networks combines simplicity of supervised learning and independence from labeling. There is a brave attempt to apply such approach for network anomaly detection [7, 8] and even to implement IDS on its base. Neural network one-class classification helps to reveal anomalies among typical patterns. This approach leads us to online analysis of multidimensional data [9–11] and it's known to be used not only for network anomaly detection [12]. It postulates that anomaly is just a novelty. Known data patterns combine into normal data set which accumulates indirectly knowledge about normal behavior. In this case all or almost all source data belongs to the one class of normal behavior. A survey [13] mentions support vector machines (SVM) and neural networks as techniques to solve such problem. Advantage of neural networks for one-class classification is proved by successful applications for very complicated high-dimensional cases [4] and suggested as a promising technique for anomaly detection in many surveys [9, 14, 15].

In this work we propose results of the algorithm investigation with real world attack scenarios and quality assessment estimations. Especially XSS and SQL injection vulnerability search were performed as well as SQL injection exploitation to steal the database content from attacked server.

Idea

There is an observation during significantly long time range of server working on processing of incoming requests of all possible types. This period indirectly contains information of normal functioning of the server. Supposing the absence of other disturbances including maintenance actions of server administrator, change of software and hardware configuration, vectors of instantaneous correlation response combine a set which describes all possible behaviors of the server in different conditions of normal functioning.

It would be expectable that aberrant behavior of the server will reveal itself by some instantaneous correlation response which will significantly differ from any of normal ones. It is evident that correlation picture of DoS attacks of different kinds, malware running, the site homepage overriding on the server (so called defacing), hardware and software failures will differ from correlation picture of normal functioning.

For aberrant behavior recognition it's suggested to apply neural network used for one-class classification. It allows to differ known or close to them correlation pictures from unknown – anomaly ones.

More details about the algorithm are provided in [9].

Neural network synthesis

Neural network for one-class classification operates like an auto-associative function with intermediate data compression. Neural network is trained to perform mapping from input vector $R(k)$ to the same one on output $R^*(k)$. As a result of training we will have $R(k) \cong R^*(k)$. An example of such neural network may be multilayer perceptron with narrow hidden layer (Fig. 1).

The number of inputs and outputs of neural network should coincide with dimension of instantaneous correlation response vector. At least one hidden layer in the middle of deep neural network must be narrowed to less dimension than input vector to provide effect of data compression and generalization on auto-associative transformation during learning.

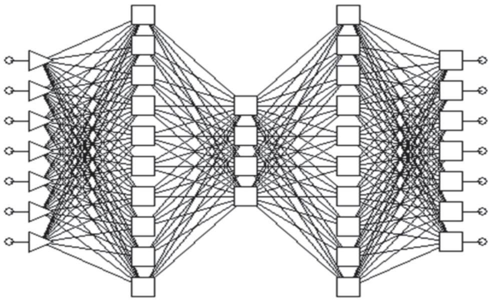

Fig. 1. Example of feedforward neural network for one-class classification problem solution

Training set is composed of correlation response vectors which were gathered up on the period of normal server work. Since the neural network performs generalization and compression of the training data some of training pairs may be not similar to the most of training pairs. If their number is relatively small they will not affect strongly the result of training. In fact this means the neural network will not remember such rare vectors. They might be anomaly samples which were included to training set accidentally.

During supervised training the neural network is tuned to minimize mean squared reconstruction error of vectors from training set (1).

$$\bar{e}_r^2 = \frac{1}{L}\sum_{k=1}^{L}\|R(k) - R^*(k)\|^2. \quad (1)$$

The whole set of instantaneous correlation vectors is subdivided into three subsets for training, validation and final test.

In normal operation mode after training was completed a vector of instantaneous correlation response $R(k)$ is feed to the input of neural network and its auto-association image $R^*(k)$ is obtained on output (2). Reconstruction error characterizes the value of novelty of the correlation image for the neural network:

$$e_r(k) = \|R(k) - R^*(k)\| = \left(\sum_{j=1}^{M}\sum_{i=1}^{N}\left(r_{ij}(k) - r_{ij}^*(k)\right)^2\right)^{1/2}. \quad (2)$$

This error is small for vectors which are close to ones from training set and large for significantly different.

To detect anomaly one has to set threshold value for reconstruction error. The reasonable way to select threshold is to estimate residual mean squared reconstruction error for the test subset because it is independent with training set and algorithm. Exact value of threshold may be calculated by well-known rule of "three sigma".

Stand

In order to fulfill the experimentation process using the algorithm of anomaly detection a stand has been created (Fig. 2). The stand includes the server under anomaly detection system, supportive computer with DBMS, operating as additional network infrastructure and the attacker. Therefore, the observed computer is Windows-based and holds a web-server with functioning vulnerable website. The website uses the database network in order to receive and present necessary information. The attacker computer is operated under Kali Linux. Sqlmap software is used in order to generate anomaly traffic, search and exploit SQL injections, while Acunetincs silent mode is used to scan web-vulnerabilities XSS. This mode imitates real user actions, not just generates traffic. WireShark Network Protocol Analyzer software has been used in order to damp the traffic and receive statistics. The stand has been outlined completely in VMware Workstation virtual environment.

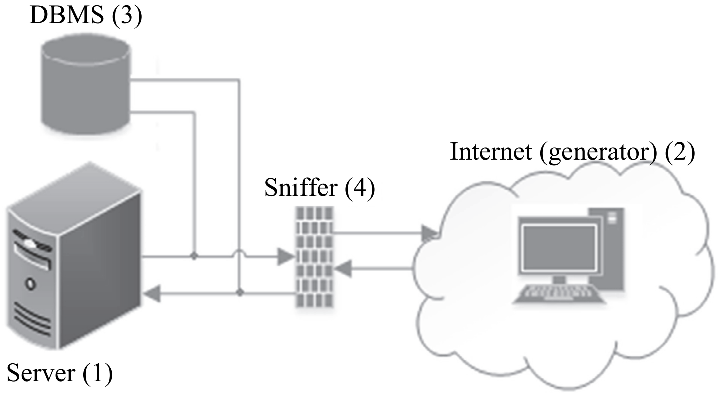

Fig. 2. Stand structure

There are two most popular types of network attacks observed during the experiment: Cross site Scripting (XSS) and SQL injections. Both of them represent the injections from the site of attacking website client, though they acquire some differences. During the XSS vulnerabilities exploitation, the attacker operates only with the website and its content. At XSS automatized retrieval there is the increase of memory space used and CPU usage due to a large number of requests being processed by Web Server. Apart from the communication with the attacked server, SQL injections also interact with the database, which exists in the same network, but is installed on the other server. While interaction the web-server-database-communication traffic arises. Different impact on the system allows conducting more detailed research of the algorithm.

Experiments

While the experiment is in process there are four major parameters observed: incoming and outgoing TCP, UDP and HTTP packets (bytes) and DBMS requests. Then all the data are transposed into the vectors (quantity of one type request per second). At the same time during the experiment the stand statistics is being monitored once a second. Next metrics are observed: processor usage (percent), memory usage (percent), disk input/output (bytes). Therefore, the instantaneous correlation response is derived within 28 parameters.

There are two time series of incoming and outgoing variables formed with time step equal to Δt=1s. The length of each series equals 5200 s. The first one is formed using general web-server workflow on the computer. The second time series, apart from general workflow, includes SQL injections and XSS vulnerabilities search (if found, are exploited and the information from the database is downloaded).

In order to construct a neural network for one-class classifier in MATLAB a feed-forward network with 28 inputs, 28 outputs and 5 hidden layers (with 40, 20, 8, 20 and 23 neurons in consequent layers) has been created (Fig. 3). The neural network is created to be trained in the auto-associative memory mode. During the process of training, it learns how to restore the incoming vector on the outputs. The choice of the network infrastructure is explained by the bottleneck existence with the small amount of neurons. It means that the characteristics of the training data can be generalized. The experiments with the different network structures conclude that there is the straight correlation between the width of the bottleneck and the value of False Positive mistake. The bottleneck optimal width existence may be discussed, because if the bottleneck becomes too narrow the neural network completely fails to function.

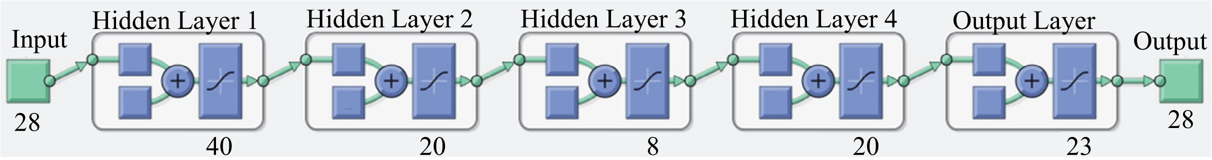

Fig. 3. Neural network structure for one-class classification problem, where «W» is weight matrix and «b» is bias

According to the result of the conducted research of the neural network structures and different training data it has been concluded, that the optimal regression coefficient is the range between 0.88 до 0.98. In cases when the coefficient appears to be below the range, the neural network experiences the incorrect training process and shows almost random numbers in the outputs. When the regression coefficient reaches 1, there is little generalization and the outputs almost repeat the input. The satisfactory optimum between generalization and precision of reconstruction was estimated as regression coefficient value 0.98.

The choice of correlation window size is also important. It is evaluated on the basis of data type and data content. For the incoming heterogeneous data with high deviation in order to decrease the number of false positives, it is needed to increase the window size. For data that are more homogeneous the less-size correlation window is applicable. In this case, since the data are quite homogeneous and have slight deviation, according to the research experiments, it is agreed to use correlation window width of 5.

The graph on Fig. 4 represents the reconstruction errors during attack period. Red lines on timeline mark periods of searching for XSS vulnerabilities and exploiting them, black lines represent SQL injection attacks (search and exploitation). Between attacks the normal web server processing took place simultaneously: usual requests and replies.

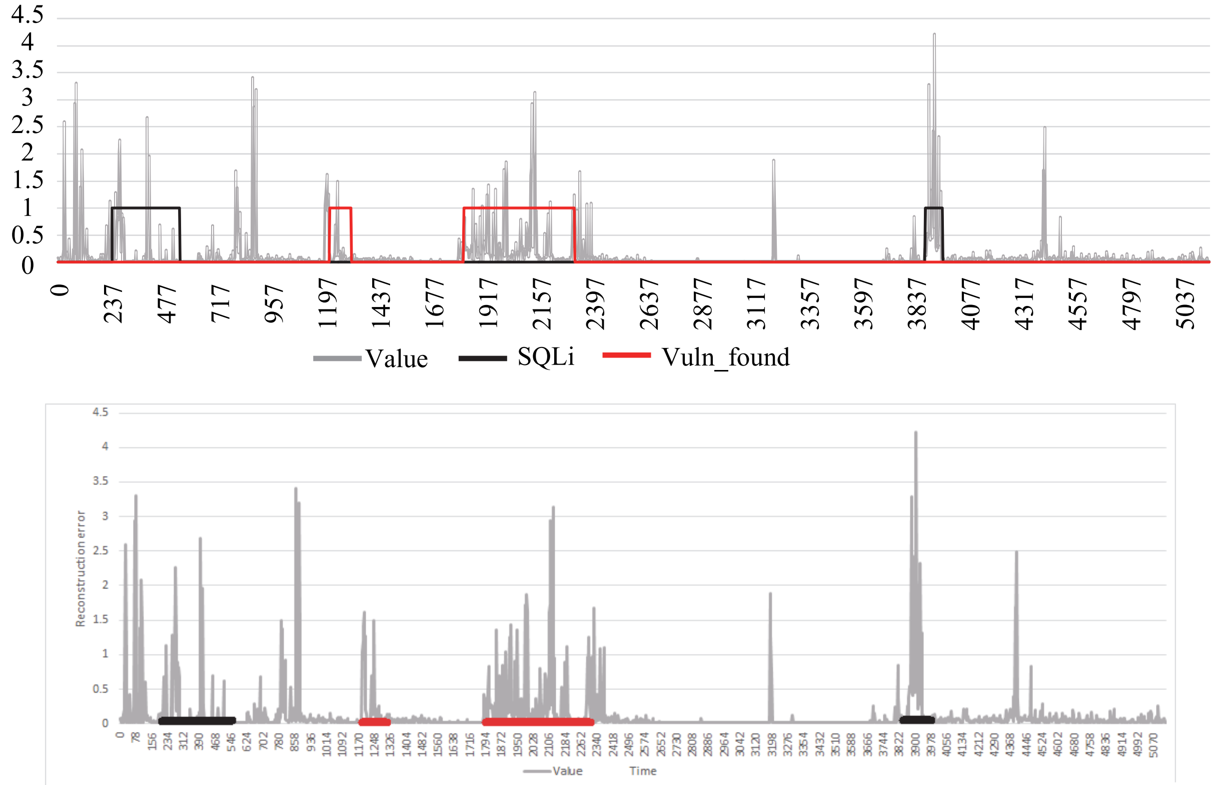

Fig. 4. Reconstruction error graphs with attack period

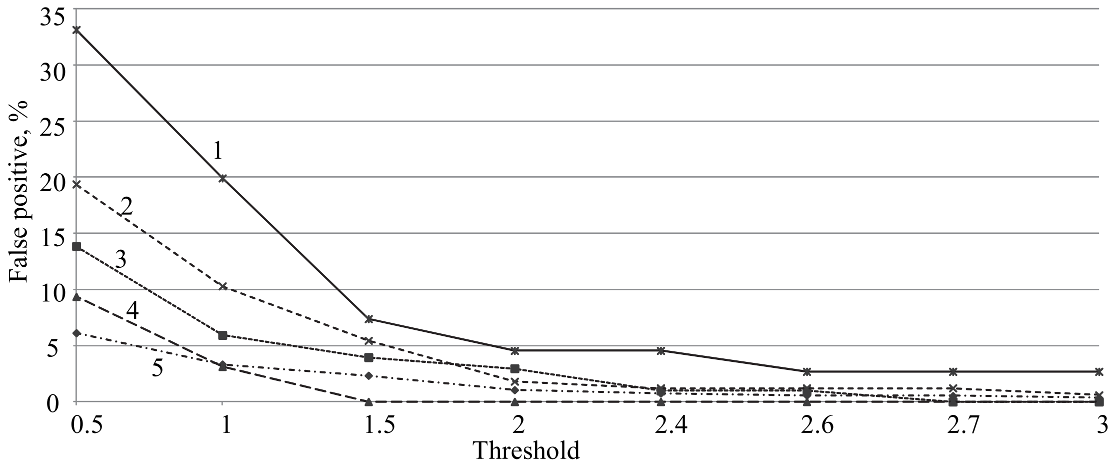

Fig. 5. False positive detection rate between attacks: 1 – Total, 2 – Between Attack 1 and Attack 2, 3 – Between Attack 3 and Attack 4, 4 – Between Attack 2 and Attack 3, 5 – Between Attack 4 for End

The algorithm has successfully detected 4 attacks considering normal operations on the background. If the thresholds are stated correctly the quantity of False Positives significantly decreases. If the threshold is equal to 1.5 the quantity of false positives is 8 points (apart from the first anomaly when database has been launched and extra system load took place). Obviously, the lower the threshold is the more anomaly points and the more false positives are observed (unexpectedly, there appears to be non-proportional increase in numbers. In order to decrease the quantity of false positives it is needed to vary the parameters of neural network and the correlation window width. Fig. 5 presents the relation between false detections and the value of threshold settings. As it is stated there is a point where the dramatic increase of false-positives starts. When choosing the threshold settings equal to 1.5, 8 false-positives are received (in 1400 points studied). This result is relatively good considering there were deviations and normal traffic. Using the signal post-processing techniques or heuristics they can be easily clarified.

Conclusion

New approach to detect anomaly using multidimensional dynamic response of controlled system is introduced and successfully applied on real world types of attack. Some investigations of neural network structure were performed to obtain efficient and robust one-class classifier. The importance of correlation width value and threshold value selection was emphasized. False positive rate was estimated along the time series of experimental data. Some ideas about enhancement of the algorithm quality and robustness were mentioned.

References

1. Kaustav Das. Detecting Patterns of Anomalies. CMU-ML-09-101. Pittsburgh, ProQuest, 2009. 152 p.
2. García-Teodoro P., Díaz-Verdejo J., Maciá-Fernández G., Vázquez E. Anomaly-based network intrusion detection: techniques, systems and challenges // Computers and Security. 2009. V. 28. N 1–2. P. 18–28. doi: 10.1016/j.cose.2008.08.003
3. Hodge V.J., Austin J. A survey of outlier detection methodologies // Artificial Intelligence Review. 2004. V. 22. N 2. P. 85–126. doi: 10.1023/B:AIRE.0000045502. 10941.a9
4. Chandola V., Banerjee A., Kumar V. Anomaly detection: a survey // ACM Computing Surveys. 2009. V. 41. N 3. Art. 15. doi: 10.1145/1541880.1541882
5. Pradhan M., Pradhan S.K., Sahu S.K. Anomaly detection using artificial neural network // International Journal of Engineering Sciences & Emerging Technologies. 2012. V. 2. N 1. P. 29–36.
6. Aneetha A.S., Bose S. The combined approach for anomaly detection using neural networks and clustering techniques // Computer Science & Engineering: An International Journal. 2012. V. 2. N 4. P. 37–64. doi: 10.5121/cseij.2012.2404
7. Клионский Д.М., Большев А.К. Применение искусственных нейронных сетей в задачах обнаружения аномалий в поведении сложных динамических объектов // Нейрокомпьютеры: разработка, применение. 2011. № 11. С. 32–45.
8. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks / In: Advances in Neural Information Processing Systems. 2012. P. 1106–1114.
9. Eliseev V., Shabalin Y. Dynamic response recognition by neural network to detect network host anomaly activity // Proc. 8th Int. Conf. on Security of Information and Networks (SIN'15). 2015. P. 246–249. doi: 10.1145/2799979.2799991
10. Thottan M., Liu G., Ji C. Anomaly detection approaches for communication networks / In: Cormode B.G., Thottan M. Algorithms for Next Generation Networks. London, Springer, 2010. P. 239–261. doi: 10.1007/978-1-84882-765-3_11
11. Dasgupta D., Majumdar N.S. Anomaly detection in multidimensional data using negative selection algorithm // Proc. 2002 Congress of Evolutionary Computation (CEC '02). Honolulu, USA, 2002. V. 2. P. 1039–1044. doi: 10.1109/CEC.2002.1004386
12. Thakur M.R., Sanyal S. A multi-dimensional approach towards intrusion detection system // International Journal of Computer Applications. 2002. V. 48. N 5. P. 34–41. doi: 10.5120/7347-0236
13. Khatkhate A., Ray A., Keller E., Gupta S., Chin S.C. Symbolic time-series analysis for anomaly detection in mechanical systems // IEEE/ASME Transactions on Mechatronics. 2006. V. 11. N 4. P. 439–447. doi: 10.1109/TMECH.2006.878544.
14. Ben-Gal I. Outlier detection / In: Data Mining and Knowledge Discovery Handbook. Springer, 2005. P. 131–146. doi: 10.1007/978-0-387-09823-4_7
15. Bridges S.M., Vaughn R.M. Fuzzy data mining and genetic algorithms applied to intrusion detection // Proc. 23rd National Information Systems Security Conference. Baltimore, USA, 2000. P. 13–31.

Yury D. Shabalin – postgraduate, National Research University "MPEI", Moscow, 111250, Russian Federation; information security specialist, SberBank Technologies, Moscow, 117105, Russian Federation, Yury.shabalin@gmail.com

Vladimir L. Eliseev – PhD, Chief of Research and Department Center, ISC InfoTeCS, Moscow, 127287, Russian Federation; Senior Lecturer, National Research University "MPEI", Moscow, 111250, Russian Federation, Vlad-eliseev@mail.ru

Шабалин Юрий Дмитриевич – аспирант, Национальный исследовательский университет МЭИ, Москва, 111250, Российская Федерация; специалист по ИБ, СберБанк Технологии, Москва, 117105, Российская Федерация, Yury.shabalin@gmail.com

Елисеев Владимир Леонидович – кандидат технических наук, руководитель центра научных исследований и перспективных разработок, ОАО ИнфоТеКС, Москва, 127287, Российская Федерация; старший преподаватель, Национальный исследовательский университет МЭИ, Москва, 111250, Российская Федерация, Vlad-eliseev@mail.ru