DOI: https://doi.org/10.53052/9788367652001.10

Yanina SHESTAK[1], Anna TORCHYLO[2], Serhii DAKOV[3]

Supervisor: Serhii TOLIUPA[4]

NEURAL NETWORK ALGORITHMS FOR DATA CENTERS CYBERSECURITY

Streszczenie: W artykule opisano główne podejścia stosowane w organizacji strategii ochrony centrów danych przed cyberatakami. Zaproponowano aparat matematyczny do oceny skuteczności systemu bezpieczeństwa informacji usług sieciowych na poziomie ilościowym. Przedstawiono metodykę optymalizacji algorytmów sieci neuronowych wykrywających próbki kodu cyberataku, odpowiednio ekstrema funkcji celu dokładności klasyfikacji i całkowitego czasu analizy danych maszynowych.

Słowa kluczowe: centrum danych, analiza maszynowa, algorytmy sieci neuronowych, głębokie uczenie maszynowe, cyberatak, zarządzanie zdarzeniami bezpieczeństwa informacji, funkcja celu

ALGORYTMY SIECI NEURONOWYCH DLA CYBERBEZPIECZEŃSTWA CENTRÓW DANYCH

Summary: The paper describes the main approaches used in the organization of the data center protection strategy against cyberattacks. A mathematical apparatus for evaluating the effectiveness of the network service information security system on a quantitative level is proposed. The methodology of optimization of neural network algorithms that detect cyberattack code samples, respectively, extrema of objective functions of classification accuracy and total time of machine data analysis is presented.

Keywords: data center, machine analysis, neural network algorithms, deep machine learning, cyberattack, information security event management, objective function

[1] Assistant of Taras Shevchenko National University of Kyiv, Faculty of Informational Technology, The Department of Cyber Security and Information Protection, yaninashestak@gmail.com

[2] Student of Taras Shevchenko national university of Kyiv, Faculty of informational technology, The Department of Cyber Security and Information Protection, speciality Cybersecurity, atorouss@gmail.com

[3] Assistant of Taras Shevchenko National University of Kyiv, Faculty of Informational Technology, The Department of Cyber Security and Information Protection, serhii.dakov@knu.ua

[4] Doctor of Engineering Science, Professor, Taras Shevchenko National University of Kyiv, Faculty of Informational Technology, The Department of Cyber Security and Information Protection, tolupa@i.ua

1. Setting the task of managing data center information security events

The organization of data centers based on cloud resources with a distributed information system scheme significantly expands the toolkit for modifying, combining, dividing and scaling the service infrastructure, but increases the requirements for the level of data center protection from external cyberattacks. The development of a holistic methodology in this field, which can be effectively adapted to the given task, necessitates the construction of a Security Information and Event Management (SIEM).

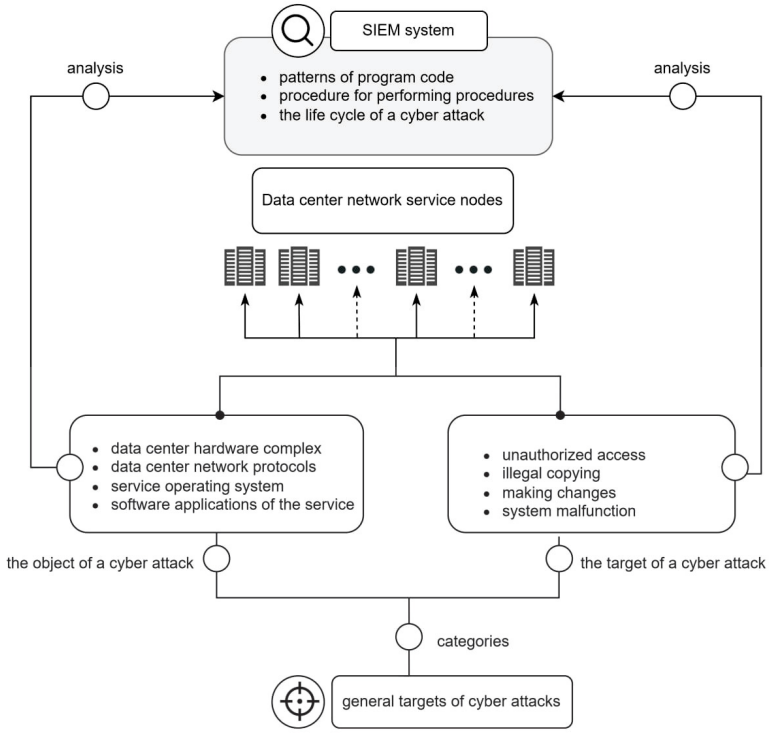

Figure 1. The basic scheme of the organization of the information security event management system of the data center

At the same time, it should be noted that building a SIEM system is a complex task, which includes the determination of the following key components (Fig. 1):

- The object of the impression of a cyberattack, which can include hardware platform resources, network protocols, operating system (OS), software applications and blocks of data to be stored within the datacenter network resource.
- The purpose with which a cyberattack is carried out. This category includes illegal copying of data sets or making changes through unauthorized access, as well as disruption of the stable operation of the information network of the data center due to the impact on the operation of the hardware and software complex platforms.

- Patterns of software code and peculiarities of the execution of procedures (the behavior according to which the life cycle is determined) of the algorithms based on which a cyberattack is carried out.

The identification of features of software code patterns, behavior and the life cycle of a cyberattack is implemented the most effectively with machine analysis based on neural network algorithms [1-7]. Typical neural network architectures used in this area include:

- deep belief networks (DBN), which within the framework of the task are considered as generative graph models [1];
- recurrent neural networks (RNN), which work efficiently with sequences of events [2];
- convolutional neural networks (CNN), used to highlight typical code patterns, and the choice of CNN provides an opportunity to reduce the load on the computing resource of the general complex of machine analysis [3];
- recursive neural networks (RvNN), which are based on the recursive application of one set of weights to a structured dataset [4];
- generative-adversarial neural networks (GAN), as a combination of two neural networks, where the generative model generates samples of program code, and the discriminative model performs their classification and selection of relevant cyberattack patterns [5];
- neural network autoencoder architecture, which can be extended to a multilevel autoencoder type architecture for extracting high-level features [6-7].

Preliminary analysis points to the advantages of using Deep Learning Artificial Neural Network Architecture (DL-ANN) in machine analysis, but for an accurate assessment, it is necessary to build appropriate mathematical models, offer quantitative indicators of the effectiveness of neural network algorithms, determine the extrema of the objective functions and correlate the obtained results with statistical data.

2. Construction and adaptation of neural network algorithms according to the task of detecting a cyberattack on data center infrastructure

The construction of an mathematical model for conducting a machine analysis procedure with the aim of detecting a cyberattack on the data center infrastructure based on program code and behavior includes the possibility of calculating such target indicators as classification accuracy, the load on the hardware platform resource, as well as the time of processing an incoming request according to the flow of incoming data and work tasks in real time. Neural network algorithms can be based on the standard ANN architecture (SL –ANN - Shallow Learning Artificial Neural Networks) and deep learning algorithms based on DL-ANN.

According to the specified performance criteria of the machine analysis SIEM system, deep learning neural network algorithms have the following advantages and disadvantages (Fig. 2):

- High accuracy of machine analysis, which includes the possibility of extracting high-level features, which provides an opportunity to effectively extract patterns of software code and behavior after the attacker makes changes, provided that the general principles of cyberattack are preserved.

- The possibility of effective work with large volumes of data, which significantly expands the functionality of the SIEM system.
- Increasing the load on the computing resource and memory resource (random memory and information storage) of the complex of machine analysis of the execution of the software code, which may be unacceptable according to the limitations of the hardware complex.
- An increase in the processing time of input data, which may be unacceptable according to the task of working in real time mode and the time of learning neural network algorithms on the training sample.

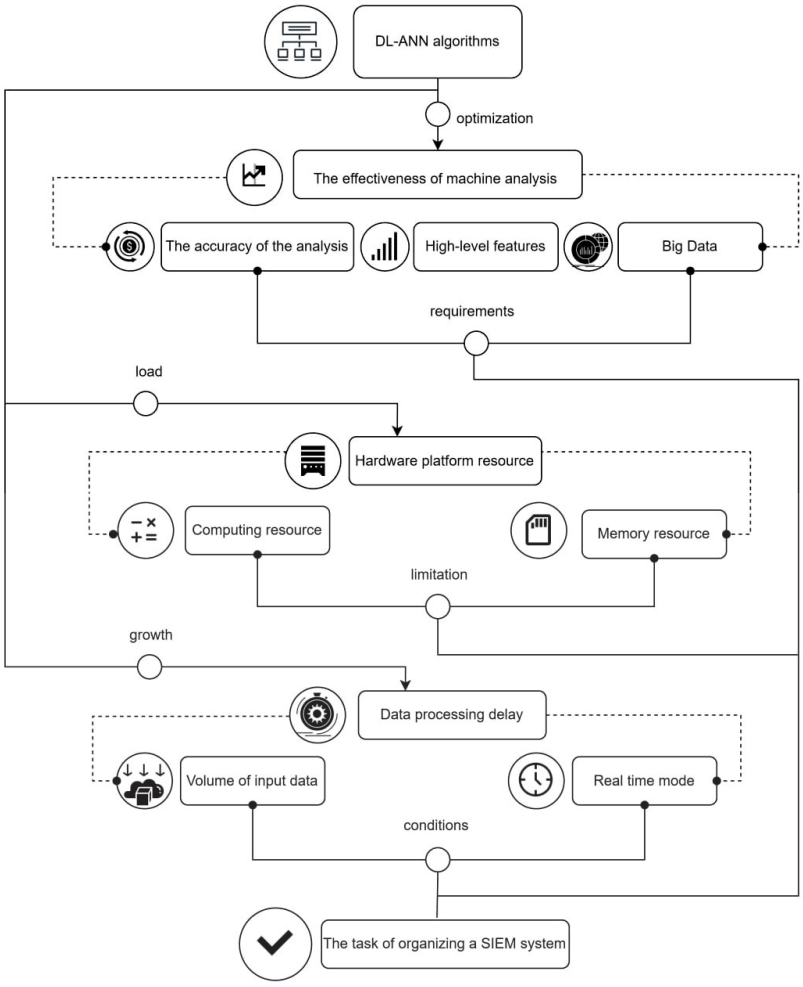

Figure 2. Scheme of adaptation of neural network algorithms of deep learning in the organization of the SIEM system of the data center

In order to determine the quantitative indicators of the accuracy of the identification and classification of the software code of the cyberattack, according to the statistical

results of the study in relation to the total number of objects of analysis N_Σ, the following notations are introduced:

- N_{TP} — number of true positive (TP) results of machine analysis;
- N_{TN} — number of true negative (TN) results of machine analysis;
- N_{FP} — number of false positive (FP) results of machine analysis;
- N_{FN} — number of false negative (FN) results of machine analysis.

Based on them, the following objective functions of the accuracy of the classification of the program code pattern can be obtained as F_{AL}(AL - Accuracy Level) and F_{PL}(PL - Precision Level) :

$$F_{AL} = (N_{TP} + N_{TN})/N_\Sigma, \tag{1}$$

$$F_{PL} = N_{TP}/(N_{TP} + N_{FP}). \tag{2}$$

To expand the set of objective functions of the accuracy of machine analysis when building a complete and universal evaluation system, it is necessary to introduce such indicators as the total sum of classification errors as N_F and the total sum of correct classification results as N_T, based on which the coefficients of false and correct detection results are entered (κ_F and κ_T), which provides an opportunity to calculate the completeness index $F_{RE}$and F1 - the classification accuracy index F_{F1}:

$$\begin{bmatrix} \kappa_F = N_F/N_\Sigma \\ \kappa_T = N_T/N_\Sigma \end{bmatrix}, \tag{3}$$

$$F_{RE} = N_{TP}/(N_{TP} + N_{FN}), \tag{4}$$

$$F_{F1} = 2F_{RE} \cdot F_{PL}/(F_{RE} + F_{PL}). \tag{5}$$

3. Conclusion

According to the analysis carried out, the task of optimizing the complex of machine analysis based on neural network algorithms for the detection and classification of cyberattacks can be reduced to the task of finding the global extremum of the target function. At the same time, in accordance with the specific task of the organization of the SIEM system, one of the accuracy functions of the set F_{AL}, F_{PL}, $F_{RE}$and $F_{F1}$is considered as a target, and for the others, together with the indicator of the number of false classification results κ_F, permissible limits are introduced:

- Permissible limits of the number of false classification results $\kappa_F \in [0; \kappa_F^{max}]$, where the value $\kappa_F^{max}$is chosen in accordance with the requirements determined at the level of the organization of the SIEM -system .
- Permissible limits classification accuracy functions $F_{AL} \in \left[F_{AL}^{min}; 1\right]$, where the value $F_{AL}^{min}$is chosen according to the requirements defined at the level of the SIEM -system organization.
- Permissible limits classification accuracy functions $F_{PL} \in \left[F_{PL}^{min}; 1\right]$, where the value $F_{PL}^{min}$is chosen according to the requirements defined at the level of the SIEM -system organization.
- Permissible limits classification accuracy functions$F_{RE} \in \left[F_{RE}^{min}; 1\right]$, where the value $F_{RE}^{min}$is chosen according to the requirements defined at the level of the SIEM -system organization.

- Permissible limits of F1 - the classification accuracy indicator$F_{F1} \in \left[F_{F1}^{min}; 1\right]$, where the value $F_{F1}^{min}$is chosen in accordance with the requirements determined at the level of the SIEM -system organization.

The arguments of the corresponding functions and accuracy indicators, in turn, will be:

- a set $\{x_i\}$of neural network architecture parameters, such as the number of hidden layers, the number of neurons in the input, output, and hidden layers, the number of shift neurons, and the available connections between neurons;
- a set $\{x_j\}$defining the selection of the activation function and parameters of the activation function;
- recruitment$\{x_k\}$ parameters determining the learning process of the neural network;
- a set of $\{x_m\}$ parameters that determine the peculiarities of the preparation of the training sample of the neural network algorithm.

Thus, at the general level, the optimization problem is solved by finding the global extremum of one of the objective functions $Y_0\left(\{x_i\}, \{x_j\}, \{x_k\}, \{x_m\}\right)$within the limits for $Y_1 \in \left[Y_1^{min}\right]$, $Y_2 \in \left[Y_2^{min}\right]$and $Y_3 \in \left[Y_3^{min}\right]$other objective functions, as well as the indicator $\kappa_F \in [0; \kappa_F^{max}]$.

The presented technique allows to generalize the currently relevant approaches to optimizing machine analysis for detecting cyberattacks on data center infrastructure components, and can be used to solve a wide range of tasks related to the construction and configuration of a SIEM system.

REFERENCES

1. SARKER I.H.: Deep cybersecurity: A comprehensive overview from neural network and Deep Learning Perspective. SN Computer Science 2(2021)3. 35-46.
2. MA X., ZHANG X., DONG C. & CHEN X.: A survey on Secure Outsourced Deep Learning. Cyber Security Meets Machine Learning (2021), 129–163.
3. ABU AL-HAIJA: An efficient and robust multi-object recognition and tracking algorithm using mask region based convolution neural network (R-CNN). International Journal of Innovative Technology and Exploring Engineering, 8(2019)9, 607–613.
4. JIANG P., WU H., XIN C. DeepPOSE: Detecting GPS spoofing attack via deep recurrent neural network. Digital Communications and Networks. (2021), 6-19.
5. MAO X., LI Q. Generative Adversarial Networks (GANS). Generative adversarial networks for image generation, (2020) 1–7.
6. SONG Y., HYUN S., CHEONG Y.G. (2021). Analysis of autoencoders for network intrusion detection. Sensors, 21(2021)13, 4294-4302.
7. YU Y., LONG J., CAI Z.: Network intrusion detection through stacking dilated convolutional autoencoders. Security and Communication Networks (2017), 1-10.