Calculation of expected output signal of neural network model for detecting of cyber-attack on network resources

B Aitchanov[1], I Bapiev[1]*, I Terejkowski[2], L Terejkowska[3], V Pogorelov[2]

[1]*Kazakh National Research Technical University after K.I. Satpaev, Satpaev Street 22a, 050013, Almaty, Republic of Kazakhstan*

[2]*National Technikal University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Polytechnichna str. 14-a, 03056, Kyiv, Ukraine*

[3]*Kyiv national university of construction and architecture, Povitroflotsky Avenue 31, Kyiv-037, 03680 Ukraine*

**Corresponding author's e-mail: bapiev@mail.ru*

Abstract

The article is dedicated to development of neural network models for detecting of cyber-attack on network resources. Procedure of detecting of expected output signal is developed for such neural networks. The key feature of the procedure is consideration of similarity of cyber-attack etalon samples and similarity of etalon safe states in expected output signal. Similarity of samples is calculated by expert estimation. Numerous conducted experiments have revealed that application of developed procedure for training of neural network model that is aimed at detection of network cyber-attacks allows to decrease training iterations for approximately 20%. This achievement, in its turn, allows enhancing efficiency of creation of neural network model.

Keywords: Neural network, training samples, neural network model, cyber-attacks detection.

During last few years, the important direction of enhancement of protection level of network informational systems resources (ISR) is development and implementation of efficient detection neural network tools (NNT) for detection of cyber-attacks on these resources [1-3, 6]. Although analysis of sources [3, 6-8] indicates at a rather significant scientific-practical start therein, this very analysis also indicates at insufficient efficiency of training of neural network models (NNM) which are the part of indicated NNTs. Due to this problem, the NNT build-up time increases, and cyber-attack detection accuracy decreases. It should be noted that those NNMs are reviewed, for training of which the "teacher inclusive" training model is applied. Precisely these models are a base of majority of tested NNT applied in the field of information protection [2, 7]. Such NNMs base on classic models of multi-layer perceptron, radial bottom-level function or their known modifications (Elman network, Jordan network, deep neural networks). Results [3, 4] also let us to affirm that training efficiency of modern NNMs may be enhanced via improval of quality of training examples through displaying similarity of samples of discernible states of protection in expected output signal. At the same time, [3, 6, 7] the mentioned works containing thorough analysis of modern NNTs used in means of protection of information indicate absence of formalized definition procedure for the output signal which will take into consideration the similarity of discernible states. The independent analysis of available literary sources has also revealed absence of information on scientific researches in this field.

Thus, the objective of this research is the development of formalized definition procedure for the expected output signal for neural network model of detection of cyber-attacks onto network resources of informational systems.

By analogy with [5], the following formula is taken as a basis of the developed procedure:

$$Y_{\Phi} = f\left(d_{\Phi}\right), \tag{1}$$

where Y_{Φ} is the expected NNM output signal for training samples corresponding to cyber-attacks type Φ, d_{Φ} is the multitude of similarity between components Φ.

As it is foreseen to detect cyber-attacks based on analysis of corresponding features of ISR, it is presumed that similarity of these features was displayed to the extent cyber-attacks are alike to each other. It also must be taken into consideration that two basic NNM structural decisions are possible during cyber-attacks detection. In the first case, the output signal is being realized through one processing element in the output layer:

$$N_y = 1, \tag{2}$$

where N_y is a number of processing elements in the output layer.

It the second case, the number of processing elements in the output layer is equal to a number of discernible types of cyberattacks on network ISR:

$$N_y = K_s\ , \tag{3}$$

where K_s is the number of discernible protection states.

That's why, according to possible NNM structural decisions, it is necessary to consider two variants of issue of output signal [2, 3].

Let us detail the first variant. Normally, for the case of

NNM with one processing element its signal may be in the range of:

$$y \in]\, a..b\, [\ , \tag{4}$$

where *y* is the NNM output signal, while *a* and *b* are any real numbers.

During use of sigmoid function of output layer processing element activation typical for NNM based on multi-layer perceptron, the output signal is in the range from 0 to 1:

$$y \in]\, 0..1\, [\ . \tag{5}$$

During detection of any possible cyber-attacks, and to any of possible safe states of the network ISR, a certain range of output signal values is applied accordingly. Further on, discernible cyber-attack types and discernible ISR safe states will be hereinafter referred to as safe states for brevity.

Without losing discourse generality, it may be presumed that range values for different safe states are different too. Besides, the output signal for training samples corresponding to safety states samples will be at the middle of the indicated range. Due to even quantization of the range of possible values *y*, the expected output signal for the sample of each *i*-value for safe state is being calculated as follows:

$$y_{s_i} = \frac{|b-a|}{K_s} i - 0{,}5 \times \frac{|b-a|}{K_s} = \frac{i-0{,}5}{K_s} |b-a|\,, \tag{6}$$

where *i* is the number of the safe state.

In case of use of sigmoid function of activation of processing element in the output layer and due to even quantizing of the range of possible values *y*, the expected output signal for the sample of each *i*-value for safe state is being calculated the following way:

$$y_{s_i} = \frac{1}{K_s} i - \frac{0{,}5}{K_s} = \frac{i-0{,}5}{K_s}\,. \tag{7}$$

Similarity of safe states in formulas (6, 7) is possible to be taken into consideration due to the fact that similar safe states should have close numbers.

Let us detail the second variant. In the training example for the sample of the *i*-value state, the output signal of corresponding *i*-processing element is equal to 1. At this, the numeration order of output processing elements may be optional. Altogether, there is a necessity of definition of expected output signal for all other output processing elements not corresponding the mentioned sample. It should be noted that processing elements corresponding to safe states close to the sample must have close values of the output signal. Thus, based on results [3], it may be noted that the level of the output signal of corresponding processing element in the training example for the loadmodule type cyber-attack sample should differ less from the level of the output signal of corresponding processing element of rootkit type cyber-attack than from the one of corresponding processing element of "search storm" type cyber-attack. In eteram:

$$\left|y_{[а]} - y_{[о]}\right| < \left|y_{[а]} - y_{[б]}\right|, \tag{8}$$

where $y_{[а]}, y_{[о]}, y_{[б]}$ are output signals of processing elements corresponding to loadmodule, rootkit, "search storm" types of cyber-attacks.

Basically, definition of the expected NNM output signal for the second case is the more complicated variant of the first case which is confined to calculation of numerary estimation of similarity of safe states. At this, known analytical methods of such calculation [4] are peculiar by their complexity and poor reliability which makes their efficient use for detection of cyber-attacks onto network ISR more difficult. At the same time, analysis and detection of cyber-attacks against network ISR are objectives which are rather effectively reached by experts in the field of information protection [3, 5]. That's why it looks necessary to determine the numerary estimation of similarity degree of cyber-attacks parameters and ones for safe states based on expert data.

Being based on results [5], it is presumed to use statistical processing methods for expert data. In this case, quantitative data received from experts is being processed for estimation of collective opinion of an expert team, for estimation of concurrence of expert opinions and experts' competence. For estimations definition, statistical methods of selective and interval assessment are applied. At this it is recommended that the number of experts made at least 10.

Let us consider the process of assessment of similarity grade of safe states. Presumably, as a result of inquiry of an expert team consisting of *m* participants, the following data has been retrieved:

$$\begin{array}{lllll} x_{1,1}, & \ldots & x_{n,1} & \ldots & x_{N,1}\ , \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{1,m} & \ldots & x_{n,m} & \ldots & x_{N,m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{1,M} & \ldots & x_{n,M} & \ldots & x_{N,M} \end{array} \tag{9}$$

where $x_{n,m}$ is the assessment grade of the object *n* (safe state) by the expert *m*, *N* is the number of objects (safe states), *M* is the number of experts.

The average collective value of the *n* safe state is being calculated by the formula:

$$x_n = \frac{1}{M} \times \sum_{m=1}^{M} x_{n,m}\,, \tag{10}$$

where $x_{n,m}$ – is the assessment grade of the *n* safe state by the expert *m*, *n* =1…*N*.

Dispersion of the average collective value is being calculated as follows:

$$\sigma^2 = \frac{1}{M-1} \times \sum_{m=1}^{M} \left(x_{n,m} - x_n\right)^2\,. \tag{11}$$

For determination of statistical value of obtained results it is necessary to indicate the confidence range which is matched by the estimated value with the preset confidence level *P*.

Presetting error probability P_n (significance value), the one can determine the interval matched by the assessed value with probability $(1 - P_n)$:

$$I_{x_n} = \left(x_n - \varepsilon_{pn}, x_n + \varepsilon_{pn}\right). \tag{12}$$

The value ε_{pn} determines ranges of the confidence range and is being calculated as follows:

$$\varepsilon_{pn} = t_p \times \frac{\sigma_n}{\sqrt{M}}, \tag{13}$$

where t_p is the ratio depending on the preset confidence level P.

It is considered to be that the assessed value has a normal layout with the center x_i dispersion σ. The rate t_p has Student's t-distribution with (N-1) the degree of freedom and is being assessed with the help of the table, the fragment of which for some confidence level values P is given in the Tab. 1.

Dimension of agreement of expert opinions is being determined with the help of variation rate γ_n which is being calculated using the formula:

$$\gamma_n = \frac{\sigma_n}{x_n}. \tag{14}$$

TABLE 1 Coefficient value t_p

P	0,8	0,85	0,9	0,95
t_p	1,282	1,439	1,643	1,960

Calculated with the formula (14), the variation coefficient γ_n defines the relative value of the expert estimation variation range considering the average value of collective estimation x_n. Due to complete concurrence of experts' opinions when all are $x_{n,m}=x_n$, the variability rate is $\gamma_n=0$.

It is regarded that concurrence of experts' opinions is satisfactory if all are $\gamma_n<0,3$ and minimal required if all are $\gamma_n<0,2$. Otherwise, taking results into consideration, the expert estimation procedure should be repeated.

The estimation of experts' competence may be done via two rates: fair competence rate and expert's relative self-esteem rate. the fair rate is defined via filling-in the special table containing questions on all factors influencing the expert's competence. The expert's relative self-esteem rate is defined through expert's self-estimation of knowledge in the range of a certain scale, e. g. from 0 to 1.

Competence rates allow to correct expert group's collective opinion estimations. At this, the average collective object estimation will be as follows:

$$x_n = \frac{1}{M} \times \sum_{m=1}^{M} \lambda_m x_{n,m}, \tag{15}$$

where λ_m is the competence rate of expert m.

In case the information on expert's competence is missing, it is recommended to use the value of the rate $\lambda_m = 1$.

It should be noted that resulting formulas (4-15) make the basis of mathematical support of separate operations of the definition procedure for expected output signal of the neural network model of network resources cyber-attack detection. At this, the order of completion of these operations corresponds to numbers of indicated formulas.

Experimental research has been conducted for verification of retrieved theoretical results, based on which the NNM has been built-up and habituated aiming at detection of two types of cyber-attacks and one safe state. The main hypothesis of the experiment is that the use of the developed procedure allows to reduce the number of training iterations needed for achievement of the preset training error.

As a source of data for the NNM, a KDD-99 database has been used. It contains values of 41 parameter for network connections corresponding to 22 types of cymer assaults and one safe state. There were four types of type R2L cyber-attacks etalon during the experiment. These were aimed at providing an access to a computer sideways a remote device to the non-registered user. Types of detectable cyber-attacks: buffer_overflow, perl, loadmodule and rootkit. The example of the record describing parameters of network connection for cyber-attack type perl: 25, tcp, telnet, SF, 269, 2333, 0, 0, 0, 0, 0, 1, 0, 1, 0, 2, 2, 1, 0, 0, 0, 0, 1, 1, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 69, 2, 0.03, 0.06, 0.01, 0.00, 0.00, 0.00, 0.00, 0.00, perl.

Discerning of safe connection is foreseen too. The KDD-99 record example for such a connection: 0, tcp, http, SF, 181, 5450, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 8, 8, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 9, 9, 1.00, 0.00, 0.11, 0.00, 0.00, 0.00, 0.00, 0.00, normal.

The NNM build-up is realized based on methodology developed in [1, 4, 5]. As a basic NNM type, a two-layer perceptron with input $N_x = 41$ and $N_y = 1$ output processing element is used. The number of source parameters has been selected based on KDD-99 record structure, and the number of output parameters has been motivated by simplification of the model structure. Selection of the number of training examples $P = 1000$ has been based on use of formula (16) justified in [6]:

$$P_{\min} > 20 \times N_x, \tag{16}$$

where P_{min} is the minimal number of training examples.

There is the same number of examples for each discernible safe state foreseen at formation of training samples.

The number of hidden processing elements $N_s = 405$ is calculated using the expression (17) also defined in [1, 6]:

$$N_s = Round\left(\frac{2\sqrt{P \times N_X}}{N_Y}\right), \tag{17}$$

where $Round(X)$ is the operation of definition of the nearest real number from argument X.

TABLE 2 Values of expected output signal

Safe state	**Values of expected output signal**	
	Experiment No. 1	*Experiment No. 2*
Safe connection	0.1	0.1.
cyberattack type buffer_overflow	0.3	0.9
cyberattack type loadmodule	0.5	0.3
cyberattack type perl	0.7	0.7
cyberattack type rootkit	0.9	0.5

There were two series of numeric experiments conducted for definition of a number of training NNM iterations for achievement of faultless discerning of training examples. In the first experiment, the expected output signal has been defined via formula (7) using the supposition that safe states are rated alphabetically. In the second experiment, the expected output signal has been defined via the offered procedure. Retrieved values are presented in Table 2.

During experiments it has been established that at use of offered procedure of expected output parameter rate definition, the number of training iterations needed for faultless memorizing of all training examples by the NNM has reduced for approximately 20% confirming the accepted hypothesis. Also, in the first approximation it can be regarded that operational efficiency of creation of NNM will increase for approximately 20% due to reduction of quantity of training iterations.

Conclusions

The procedure of calculation of expected output signal of neural network model for definition of cyber-attacks on network resources has been developed. It helps to enhance operational efficiency of creation of mentioned models via taking into consideration similarity of etalon cyber-attack samples and etalon samples of safe states.

Conducted numerous experiments have revealed that application of the developed procedure for training of neural network model aimed at detection of network cyber-attacks of type buffer_overflow, perl, loadmodule and rootkit allows to enhance operational efficiency of creation of such a model for approximately 20%.

Perspectives of further researches in this field need to be aimed at improval of the mentioned procedure based on development and implementation of fair coefficients of similarity for etalon samples.

References

[1] Bapiev I M, Akhmetov B S, Korchenko A G, Tereykovsky I A 2016 *The use of a neural network with radial basis functions to detect script viruses* Actual problems of cyber security and data protection software: Kyiv p. 20-3

[2] Bapiev I M, Korchenko A G, Tereykovsky I A 2016 *Development of criteria for evaluating the effectiveness of neural network recognition means cyber-attacks on network resources information systems* Global and regional problems of informatizition in society and nature using: Kyiv p. 80-2

[3] Korchenko A, Tereykovsky I, Karpinski N, Tynymbaev S 2016 *Neural network models, methods and tools to evaluate the security settings of Internet-oriented information systems* Nash Format: Kyiv p. 275

[4] Rudenko O G, Bodyansky E V 2006 *Shtuchni neyronni merezhi* SMIT: Kharkiv p. 404

[5] Tereykovska L O 2016 *Neural models and methods of recognition of phonemes in bare-owl signal in distance learning* Kyiv p. 312

[6] Tereykovskiy I 2007 *Neural networks in means of information protection* PolihrafKonsaltinh: Kyiv p. 209

[7] Tereykovskiy I 2013 *Neural network recognition methodology of Internet-based malware* Information Security **19**(1) p. 24-8

[8] Tereykovskiy I 2011 *Perceptron bilayer structure optimization, designed to identify anomalous values of operating parameters of computer network* Managing the development of complex systems: Kyiv p. 128-31