Assessment of the effectiveness of immune neural network

K Zhigalov,1,2 E O Pokhomchikova3 and M V Sygotina4

1 V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia
2 Kh. Ibragimov Complex Institute of the Russian Academy of Sciences (CI RAS), Russia
3 Irkutsk National Research Technical University, 83, Lermontov str., Irkutsk, 664074, Russia
4 Bratsk State University, 40, Makarenko str., Bratsk, 665709, Russia

E-mail: kshakalov@mail.ru

Abstract. Nowadays the information technology industry is facing a boom in the research of neural networks and related technologies. Networks are used almost everywhere. However, the question of their effectiveness is omitted. Although there is a growing need to move from quantity to quality. The paper considers a variant of the assessment of the effectiveness of the immune neural network with an example.

1. Introduction
During the development of information security system based on human immune system (working on the model of it), we faced the need to conduct research on the effectiveness of its work. Since this problem can be posed by other researchers, we will try to describe it as much as possible in this paper.

2. Mathematical base of assessment
For the experiment, we took the immune network we were developing, which should respond to all events that are not present in the information system. Decision-making on the presence of an “alien” information process \( P_{al}(t) \) in the computer system of the workplace can be made upon the fact that the frequency \( f \) exceeds a certain threshold value \( f_t \):

\[
P(t) = \begin{cases} 
P_c(t), & \text{if } f \leq f_t, \\
P_{al}(t), & \text{if } f > f_t. 
\end{cases}
\]  (1)

Moreover, in accordance with the system identification method, any process \( P(t) \), which differs from all processes \( P_n(t), n=1, M \), occurring at a given time, will be considered “alien”.

The statistical probability \( Ver \) of the presence of an “alien” information process in a computer can be reflected by a value inverse to the number of coincidences of new information processes with the hosts of the artificial immune network (2):

\[
Ver\left(P_{al}(t)\right) = 1 - Ver(P(t)), \text{where}
\]
Thus, the formula for the calculation of the number of general features can be modified as follows (6):

\[ Y = \sum_{i=0}^{n} F_{i}^{uniq}(q), \]

where \( F_{i}^{uniq}(q) \) – the number of hosts in a given artificial immune network, selected in accordance with the affinity criterion \( q \). The number of network hosts is the greater, the higher the \( q \) index and it depends on the number of operations performed at the workplace at a given time.

The monitoring system must function in real time. Therefore the speed of the comparison of the processes of the information environment with the hosts of the artificial immune network is a key factor in the response. The reduction of \( q \) parameter of the immune network affinity will reduce the number of hosts involved in the comparison, but will reduce the quality of the system's response to malicious processes. The increase in the parameter \( q \) will lead to the increase in the hosts of the system and therefore it will increase the time spent on processing incoming events.

To speed up the procedure of event processing, the comparison has exponential nature, from the most general feature to the most unique, which is described above when compiling a process card. For the analysis of the most general features, \( q \) parameter is not used, and therefore it does not affect the number of created hosts. Therefore, the time spent on the comparison of general features will grow linearly depending on the number of these features (4):

\[ T = \sum_{i=0}^{n} F_{i}^{gen} \cdot k, \]

where \( T \) - comparison operation for general features of information processes, \( F_{i}^{gen} \) – general features, \( k \) – the number of general features of a given type, \( i=0, n \) – feature reference.

In general, the parameter \( T \) can be represented as a constant. Then the assessment of the number of comparisons for parameters unique to information processes can be represented as (5):

\[ O = T + k \cdot \sum_{i=0}^{n} F_{i}^{uniq}, \]

where \( O \) – comparison operation in general, \( F_{i}^{uniq} \) – features unique for each host, \( k \) – number of unique features of a host, \( i=0, n \) – feature reference.

Unique features will not arise without general ones, since the comparison begins with them. In addition, each general feature depends on the previous one. On the other hand, when the degree of affinity is zero, the number of features for comparison will be equal to the number of general features. Thus, the formula for the calculation of the number of general features can be modified as follows (6):

\[
O = \begin{cases} 
\sum_{i=1}^{n} F_{i}^{gen}(F_{i-1}^{gen}) \cdot k_{i} + \hat{k} \cdot \sum_{l=0}^{m} F_{l}^{uniq}(F_{n}^{gen}), l = 0, m, i = 1, n \\
F_{i}^{gen}(F_{i-1}^{gen}) = 0, if F_{i-1}^{gen} = 0; \\
F_{i}^{gen}(F_{i-1}^{gen}) = 1, if \exists F_{i}^{gen} \cap F_{i-1}^{gen} \neq 0 \\
F_{i}^{uniq}(F_{n}^{gen}) = 0, if F_{n}^{gen} = 0 \\
F_{i}^{uniq}(F_{n}^{gen}) = 1, if F_{n}^{gen} = 1 
\end{cases},
\]

where \( \hat{k} \) – the number of unique features.
Consequently, the number of testing will grow depending on the growth of indicators $n$ and $m$, i.e. on the number of unique hosts in the system. For $m=0$, i.e. the absence of unique features of hosts, the number of checks will be limited to $n \times \sum k$. In the case when $m \neq 0$, the number of checks is limited to $n \times \sum k + m$.

The graph of the function of the maximum number of testing will grow linearly depending on the growth in the number of general and unique features. According to the table for the generation of information about the process, the reference to general features $i$ will change from 0 (if the name of a user who launched the process was not known earlier) to 3 ($i=0,3$). The reference to unique features $l$ will always be equal to 3, since if the process exists and found among general features, it will always have activity described by unique features.

The minimum number of testing will correspond to the number of general features of the first characteristic. The least commonly used general feature of a corporate computer network of user's workstation is the name of a user who started the process. In the pre-collected data, the number of users is 4. Accordingly, the minimum number of testing will be compared with 4 hosts containing different usernames. If the username does not exist, the system marks this process as alien and does not perform verification with further host fields. It was mentioned above that it was the field with the name of the user who started the process that the author chose as the first sign of the general characteristics of running processes.

### 3. Realization of experiment

In order to conduct the experiment, the data collected for the primary analysis were loaded into the prototype of the program for the formation of an artificial immune network. After processing data consisting of 4447 lines of the csv-file collected by the Process Monitor program, each of which described a process passing with the user's workstation, 634 hosts were formed in the system with an affinity coefficient of $q = 0.8$. Changing the affinity coefficient, we can observe the change in the number of hosts processing the data belonging to the system:

![Figure 1. Graph of dependence of system hosts on the affinity coefficient](image-url)
For the assessment of the effectiveness of the algorithm to detect an artificial immune network, the formed network was provided with data collected on the same workstation at a slightly different time. The collected data consisted of 1238 lines, each of which characterized the process taking place in the system. Since the data for detection was collected at a different time of a user's work, the control of the time when the process started was cancelled in the detection algorithm. However, it was planned to use this parameter during the current functioning of the network of hosts of the artificial immune network, since it allowed tracing the patterns of user activity depending on time.

The presented graph (Figure 1) shows that with the increase in the affinity coefficient, the number of formed hosts increases. This graph also indicates that processes continue to occur on user's workstation, corresponding to the processes reflected in the data collected in previous periods and used to build the network, which is logical. Similar experiments have been carried out in other industries [1-3].

![Figure 1](image1.png)

**Figure 1.** Graph of dependence of detection of anomalies of user processes depending on the affinity coefficient

This graph (Figure 2) shows that with the increase in the affinity coefficient, the number of processes detected as “alien” ones increases. This graph also indicates that processes continue to occur on user's workstation, corresponding to the processes reflected in the data collected in previous periods and used to build the network, which is logical.

In order to analyze the work of the immune network with abnormal data, the authors changed the data used for the initial training of the system. A total of 21 lines were changed to describe the file. The authors of the research did not deliberately change parameters such as the name of the process or the name of a user who started the process, but only touched upon those features that were unique for each process. The graph (Figure 3) of the dependence of the changes on the coefficient of affinity is presented below.

![Figure 2](image2.png)

**Figure 2.** Graph of dependence of detection of anomalies of user processes depending on the affinity coefficient
Due to the significantly larger number of processes occurring at the workstation of the employees of an organization than those presented for training the system, it is necessary to maintain a balance between the required accuracy of the correspondence of the processes performed at the workstation and the size of the detecting network. In addition, it is necessary to take into account that a user of a workstation can periodically perform actions that he has not performed before, which do not pose any threat to information security. To eliminate the redundancy of the system's response to “alien” information processes, as described earlier, the process weight coefficient was introduced which solves two problems at once.

First, the processes that are too rarely performed by a user are removed from the set of hosts of the immune network, which means that the triggering of these detectors will be quite rare and can be neglected. Secondly, the process weight coefficient is assigned to an “alien” process, and the system responds to it only when the weight gains a certain value. In tests that measure the response of the system to alien processes, the weight of the process is not taken into account, due to the very modest set of input data.

This system, already at the prototype stage, shows its ability to react to user actions, revealing the data that is different from the original. Having created an artificial immune network, trained by data collected over a sufficient period of time to observe a user of a workplace, a snapshot of the system will be obtained, reflecting the activities of an employee at his workplace.

4. Conclusion
The analysis of these experiments will reveal the effectiveness of the immune network or any other neural network, and hence the reasonability of further refinement of the system interfaces for its further industrial use. This will allow developers to more efficiently choose certain algorithms for the operation of neural networks for specific tasks.
References

[1] Klyuev R V, Bosikov I I, Revazov V Ch, Krysanov K S and Dzeranov B V 2020 A comprehensive analysis of the properties of reservoir rocks to determine the productive oil horizons *IOP Conf. Ser.: Mater. Sci. Eng.* 952 012017. DOI: 10.1088/1757-899X/952/1/012017

[2] Kortiev L I, Klyuev R V, Totchiev F G, Vaneev S D and Babunska N I 2020 Temporary and operational transport roads in construction of transfer linear structures and their scientific and technical support *IOP Conf. Ser.: Mater. Sci. Eng.* 952 012023. DOI: 10.1088/1757-899X/952/1/012023

[3] Klyuev R V, Bosikov I I, Tavasiev V Kh and Gobeev M A 2020 Improving the energy efficiency of transport equipment in ore mining *IOP Conf. Ser.: Mater. Sci. Eng.* 918 012123. DOI: 10.1088/1757-899X/918/1/012124