Development of an intelligent decision-making system to support scientific and industrial formations VPN connections

Artur Zaenchkovski and Alexey Lazarev

1 Branch of the National Research University ‘Moscow Power Engineering Institute’ in Smolensk, Russia

Abstract. The development of information systems to ensure the safe coordination of information flows in scientific and industrial clusters makes it possible to automate a number of tasks aimed at increasing the cooperative interaction productivity. The use of existing traffic encapsulation solutions or the new client-server algorithms development for network interaction affects the decision-making component for managing the TCP/IP structure, authorization of subjects, and support for correct load distribution. At the moment, most VPN servers do not have the specified functionality, which does not allow integrating solutions into existing scientific and industrial clusters. As the main solution, a flexible decision support system is proposed that takes into account all aspects of the virtual tunnel software component.

The proposed solution is based on the use of complex methods for assessing the software modules state to make decisions on changing the operation of functional modules. The development result of the proposed system and the conducted functional testing made it possible to automate the operation of VPN tunnels when working with a complex network interaction structure.

1 Introduction

Today, the national innovation system formation is a very important task and an integral part of the country’s policy. The innovation system is designed to ensure the efforts unification of all state governing bodies, the scientific and technical sphere various organizations, as well as the business economy sector in order to accelerate the application of certain scientific achievements and various technologies for the strategic priorities implementation of the state [1]. An example of such interaction in a broad sense can be scientific and industrial clusters.

It is possible to consider the model in Figure 1 as a basic local approach to virtual information coordination between participants in the innovation process. This model assumes the specific network of the system definition and the remote communication provision of the innovation process subjects with the distribution of their unique roles in the regional innovation environment [2].

In this case, the B2B systems and portals presented in this model are used to interact with government agencies in order to manage the innovation infrastructure. Business intelligence

* Corresponding author: anonymous.prodject@gmail.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
portals provide marketing and analytical support for innovative developments, including e-commerce systems. The virtual operator coordinates the overall interaction of the virtual information infrastructure in the region for each allocated technology.

Fig. 1. Information exchange model between participants of the integrated innovation process in the region

The considered information processes in Figure 1 represent multiple two-way communications. As a consequence, modern features of changes in the environment and the subjects themselves, as well as the speed of these transformations in complex socio-economic systems, determine the need for individual approaches to the verifying the interacting subject’s authenticity stages.

On the other hand, the secure interaction process of scientific and industrial clusters subjects is a complicated data transmission scheme, in which a number of vulnerabilities can be identified that are susceptible to exploitation of CVE vulnerabilities (Figure 2) [3].
According to the data presented in Figure 2, the client's authorization points, shutdown points, as well as when executing update requests are subject to exploitation of vulnerabilities. As the main solution to the presented problem, it is proposed to develop an intelligent decision-making system for managing active VPN connections, taking into account the correct operability verification of the modules in the presented diagram [4].

2 Materials and methods

2.1 Development of a decision-making system based on a tree structure

Taking into account the fact that the proposed traffic encapsulation system assumes the use of multifactor authentication modules, two-way IP addresses verification, DNS caching services, as well as intelligent algorithms for predicting traffic consumption, it is proposed to use a decision-making system based on a tree structure [5, 6]. This approach involves the use of nonparametric models that use sets of logical rules to predict the result [7]. The cyclic process initiated in this context carries out a conditional division of the space from the features to the stopping point set by certain criteria, called the cost function [8]. These functions allow to clarify the correct choice fact when dividing the tree into certain sections, optimizing the choice of such a section. The first function is entropy, which determines randomness at data points – as can be seen from the presented equation 1, the success of the result depends on the tendency of the value to zero [9].

$$E = \sum_{i=0}^{n} - p_i \cdot \log_2(p_i)$$  \hspace{1cm} (1)

An alternative approach to optimizing cuts is the Gini index [10]. The approach of this index is isolated by calculating the incorrect value probability in the case of a certain section with an interval of [0;0.5], as presented in equation 2.

$$G = 1 - \sum_{i=1}^{n} (p_i)^2$$  \hspace{1cm} (2)
According to the presented figure, in the structures under consideration there are two pairs of leaf nodes denoting critical points (True / False), as well as many internal nodes following from the root ones (making decisions on token and traffic). As the basic data used for classification and decision-making, it is proposed to use data from authorization keys, IP pool, as well as data on traffic speed changes, as shown in Figure 4. The representation of this figure displays the combined pandas library dataframe [11]. The TokenAuth / Authorization columns display the current status of the keys being authorized, IPSrv / IPClient - two-way DHCP addressing, and the TrafficCurrent / TrafficAllowed columns - the current and allowable traffic load. Based on the values of columns three, six, eight, it is possible to determine the correct functioning of the modules.

The algorithmic implementation of the trained decision-making model is reduced to passing through all nodes without the specific data presence by using the Gini index and regression applicable to the dependent values of the input dataset \( y \) according to equation 3 [12, 13]. Equation 4 demonstrates the next step of calculating the root-mean-square error (MSE) for each value \( y^*_z \) corresponding to the regression approach [14].

\[
y^*_z = \frac{1}{N_z} \sum_{i=1}^{N_z} y_i, \quad (3)
\]

\[
MSE(D_z) = \frac{1}{N_z} \sum_{i=1}^{N_z} [y_i - y^*_z], \quad (4)
\]

where \( z \) is node of a particular node, \( y \) is value of the dependent variable, \( N \) is dimension of the dataset, \( D \) is input matrix of values.

The software model learning method implementation is carried out at the expense of the scikit-learn library, which includes many functional capabilities for forecasting and...
classifying data, as, for example, in the listing of Figure 5. As can see from the presented listing, in the process of work, columns are selected from the starting dataset indicating the current and allowable traffic consumption, followed by the conversion of False as 0, True as 1. Then the data frame is divided into training and test samples, as a result of which the two vectors prediction is carried out due to the predict function of the `DecisionTreeClassifier`, `DecisionTreeRegressor` modules [15].

**Fig. 5.** Models for building decision trees through classification and regression

The result of the two approaches joint use allows both to perform regression when predicting changes in traffic, and to identify the authenticated token validity through classification.

2.2 Development of an alternative fuzzy comparison module

An important component in the implemented decision-making system is the automated receipt processes of information about the operability of each module, for example, during authentication or fixing changes in traffic. Taking into account the fact that the result of the previous module allows both predicted values and classified values for each parameter in the dataframe, it is possible to automate the decision-making process by fuzzy character-by-character comparison of two predicted sequences. This is possible by using an algorithm for calculating the intersymbol distance in two output vectors ($W_1$, $W_2$), as presented in equation 5.

$$X(j, k) = \min \{[X(j, k - 1) + 1], [X(j - 1, k)], [X(j - 1, k - 1) + m(W_{1j}, W_{2k})] \}, \quad (5)$$

were $X$ is the distance between character; $k$ is the insertion operation; $j$ is the deletion operation.

The calculation result according to the presented formula allows to determine the distance of each $i$-th character in two lines. The software implementation of this function can be performed using the `fuzz` library, which includes the partial character-by-character comparison `fuzz.partial_ratio` function. The output result of this function allows to determine the percentage match of two sequences, for example, matches by authorized characters during TOTP authentication.

On the other hand, the most accurate result in determining the state of indicators is a fuzzy comparison process based on a rules set using the `skfuzzy` library [17]. Let's consider the indicator variability of the permissible speed of traffic consumption – this indicator depends on the current input traffic speed [range –0;100] and the users on the selected channel number.
At the same time, the authorization indicator from the output of the Fuzz module should be interpreted identically by levels [acceptable; unacceptable]. As a result, the set of rules for two functions can be represented by separate variables, as shown in Figure 6.

Fig. 6. A set of fuzzy rules for two parameters

This figure representation displays a set of rules, where three rules are defined for user traffic with the levels «minor», «average», «significant», and for authorization — three rules with the output of «available», «unavailable». The software component uses the control module, which interacts with the library API. Thus, a character-by-character fuzzy comparison module use in combination with a rule-based fuzzy inference system allows automated decisions to be made on the management of individual functional modules, taking into account established security standards and available resources.

3 Results and discussions

The main result of the development of this system is the process of logical data output for traffic consumption indicators and user authentication in accordance with the specified output data, shown in Figure 7.

Fig. 7. The logical output process of authentication indicators and traffic consumption

According to the results of the testing, can say that the permissible user traffic speed was 27.002 at a speed of 65 and the number of users on the network was 38. The testing authentication levels result showed that with a match value of 83, the success rate of authorization is 79.664.

```python
1 rtraf1 = ctrl.Rule(speed['good'] | users['poor'], consumption['minor'])
2 rtraf2 = ctrl.Rule(speed['average'], consumption['average'])
3 rtraf3 = ctrl.Rule(speed['poor'] | users['average'], consumption['significant'])
4 rauth1 = ctrl.Rule(auth['good'], authenticate['available'])
5 rauth2 = ctrl.Rule(auth['normal'], authenticate['available'])
6 rauth3 = ctrl.Rule(auth['bad'], authenticate['unavailable'])
```
4 Conclusion

Considering the model of information exchange between the participants of the complex innovation process in the regional structure, as well as a vulnerabilities number in the client-server software for traffic encapsulation, it was concluded that there are multiple vulnerabilities in various places of VPN connection initialization. As the main solution to this problem, a multi-module intelligent decision-making system was developed, which is based on the combining the means approach of building a tree structure and decision-making based on fuzzy logic. The developed decision-making models based on classification and regression allowed optimizing the approach to working with multiformat data. An important functional feature of this solution is a module for combining fuzzy character-by-character comparison tools and fuzzy logic apparatus based on a set of rules. This approach made it possible to automate the decision-making process when working with traffic consumption and authentication indicators in accordance with the specified reference values.

4 Acknowledgements

The research was supported by RSF (project No. 22-21-00487).

References

1. E. Kirillova, T. Kakatunova, V. Khalin, Lecture Notes in Networks and Systems, 474, 53-64 (2022) doi:10.1007/978-3-031-05778-6


