Neural Network Model for Detecting Network Scanning Attacks

Oleg Yuryevich Panischev¹, Artur Tagirovich Makridin², Alexey Sergeevich Katasev³, Amir Muratovich Akhmetvaleev⁴, Dina Vladimirovna Kataseva⁵

¹Researcher, Scientific Laboratory for Near Space Research, Kazan Federal University; Russia. Scopus ID: 8355604900; ORCID: 0000-0001-5490-912X, Kazan Federal University, Russia.
²Research Laboratory Assistant of the Scientific Laboratory for Microwave Design and Radio Telecommunications, Kazan Federal University; tel: +7 999 155-27-54; Scopus ID: no; ORCID: 0000-0001-8192-1917, Kazan Federal University, Russia.
⁴Candidate of Engineering Sciences, Associate Professor, Department of Information Security Systems, Institute of Computer Technologies and Information Security, Kazan National Research Technical University named after A.N. Tupolev “KAI”; Russia. Scopus ID: 57202913457; ORCID: 0000-0003-0384-9539, Kazan National Research Technical University named after A.N. Tupolev, Russia.
⁵Senior Lecturer, Department of Information Security Systems, Institute of Computer Technologies and Information Security, Kazan National Research Technical University named after A.N. Tupolev “KAI”; Russia. Scopus ID: 57193401954; ORCID: 0000-0001-6141-8329, Kazan National Research Technical University named after A.N. Tupolev.

Abstract

This paper discusses the concept and problem of detecting network scanning attacks and describes the targets of network scanning attacks. The main attack methods and approaches to scanning network ports are considered. Intrusion detection systems (IDS) are used to detect network scanning attacks. Based on the method of detecting attacks, such systems are divided into IDS, which detects attacks based on signatures, and IDS, which detects attacks based on anomalies. In practice, it is recommended that these IDS detection methods be used together. It is proposed to use a trained neural network as a tool for detecting network scanning attacks. The implementation of the neural network required to prepare the initial data for training, to determine the parameters of the network, to conduct training, and to evaluate the results of its testing. When developing a neural network model, data from the publicly available set "NSL-KDD" were used. During data processing, entries that were not related to network scanning attacks were removed from the original NSL-KDD set. After processing the initial data, the sample contained 5108 records, 3379 of which characterized normal connections, and 1729 connections were related to network scanning attacks. The Deductor modeling environment was used to build a neural network model. The structure of the constructed neural network was as follows: 11 input neurons, 1 output neuron, and one hidden layer consisting of 23 neurons. The neural network was trained using an error backpropagation algorithm. The quality of the neural network model was assessed using contingency tables with the calculation of the classification accuracy, as well as errors of the first and second kind. The values of these errors turned out to be insignificant. The constructed neural network model revealed most of the connections characterizing network scanning attacks. The neural network assessment confirmed its adequacy and the possibility of effective practical use for detecting network scanning attacks.

Keywords- network scanning attack, information security, data mining, neural network, neural network model.

I. INTRODUCTION

Currently, the issues of protecting information systems from malicious attacks, as well as improving information security policies are becoming important [1, 2]. Information flow is increasing every day. In this flow, it is more and more difficult to ensure the security of information systems and the information processed. The mechanisms for detecting cyberattacks are constantly being improved [3-5]. However, the development of mechanisms for the implementation of threats also does not stand still. Data transmission networks are the most vulnerable to attacks. In this regard, the problem of detecting network attacks is urgent.

Network attacks are malicious actions of cybercriminals, which are performed both by the attackers themselves and by the malware installed on the attacked computer [6]. The main type of network attacks is network scanning [7, 8]. The further success of other network attacks directly depends on the network scanning stage. Therefore, the identification of this particular type of network attacks is especially important. The purpose of these attacks is to identify hosts connected to the network and network services open on them (open TCP / UDP ports) [9].
II. METHODS

There are 4 main methods of scanning network ports depending on the attacker's strategy [10, 11].

1. Horizontal scanning. With this approach, an attacker checks the same port at different IP addresses. This approach is the most common.

2. Vertical scanning. This is an approach in which an attacker scans multiple ports on the same IP address.

3. Distributed vertical scanning. If multiple sources sequentially scan multiple ports at the same IP address, then this approach is called distributed vertical scanning.

4. Distributed horizontal scanning. If multiple sources sequentially scan a specific port at multiple IP addresses, then this approach is called distributed horizontal scanning.

Various utilities can be used to detect network scanning attacks, such as TCP dump, firewalls, or intrusion detection systems (IDSs) [12]. The most advanced of this list are intrusion detection systems. They help solve problems such as intrusion or network attack detection, attack prediction, vulnerability identification, attack source identification, and others.

According to the method of detecting an attack, intrusion detection systems are usually divided into the following categories [13]:

1) Signature-based attack detection;

2) Anomaly-based attack detection.

Thus, a network scanning attack can show certain signs. When an attack occurs, the intrusion detection system will identify it by signature. Also, the intrusion detection system, which knows the normal behaviour of the system, will determine the network scanning attack based on the deviation from this normal state, i.e. recognizes it as an anomaly. In practice, it is recommended to use them together based on the advantages and disadvantages of these attack detection methods.

An efficient way to detect network scanning attacks is the data mining of network traffic based on a neural network approach [14-17]. This approach is based on the creation of a neural network trained with the use of data that include the signs of a network scanning attack. As a result of training, the neural network is able to detect the presence of a network scanning attack. The complexity of this approach lies in finding a sufficient number of network traffic examples which are both normal and defining a network scanning attack to build an adequate neural network model. However, the neural network method attracts the attention of developers of intrusion detection systems with its ability to effectively solve the intrusion detection task by flexible responding to signs of network attacks, and adapt to the current conditions of their implementation.

The construction of a neural network model is preceded by the stage of selection and preparation of data for subsequent mining. In this work, the NSL-KDD dataset was selected to train the neural network to detect network scanning attacks. The data in this set is determined by 41 input parameters of network traffic and 1 output parameter, which determines the availability or unavailability of a network attack. In total, the dataset describes the patterns of 4 categories of network scanning attacks: DoS, U2R, R2L, and Probe.

When processing data from the original set of "NSL-KDD", records that are not related to network scanning attacks were removed. Also, for processing the initial data, the method of calculating Pearson's correlation was used [18], which allows one's to reduce the dimension of the input feature space. As a result of preprocessing the data, the most significant input attributes that are involved in training the neural network were selected. The final training sample contained 11 input attributes and one output (attack), as well as 5108 records, 3379 of which described normal connections, and 1729 - connections related to network scanning attacks.

III. RESULTS AND DISCUSSION

The analytical platform Deductor Studio Academic was chosen as a modelling environment for training and testing a neural network [19]. Deductor software is a user-friendly analytical platform. The analyst's work with it comes down to visual scripting. A script is a sequence of actions that allows a user to obtain useful knowledge from the source data and identify patterns through such operations as data import, data processing, data visualization, and others.

When training a neural network in the Deductor modeling environment, the results of data classification were obtained; they are presented in Table 1 [20].

<table>
<thead>
<tr>
<th>Actually</th>
<th>Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>3351</td>
</tr>
<tr>
<td>True</td>
<td>82</td>
</tr>
<tr>
<td>Total</td>
<td>3433</td>
</tr>
</tbody>
</table>

Table 1. Model Results on Training Data
Based on the data presented in Table 1, we can conclude that the classification accuracy of data from a training sample of 5108 records was 97.85%. At the same time, 3379 records in the training set accounted for normal network connections, and 1729 records corresponded to network attacks.

When testing the neural network in the Deductor modeling environment, the data classification results presented in Table 2 were obtained.

**Table 2. Model Results on Testing Data**

<table>
<thead>
<tr>
<th>Actually</th>
<th>Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>False</td>
</tr>
<tr>
<td>False</td>
<td>1184</td>
</tr>
<tr>
<td>True</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>1232</td>
</tr>
</tbody>
</table>

Based on the data presented in Table 2, we can conclude that the classification accuracy of data from the test sample with a volume of 2066 records was 96.9%. At the same time, 1200 records in the test sample accounted for normal network connections and 866 records corresponded to network attacks.

Based on the results of the neural network model operation on test data, errors of the first and second kind were calculated using the following formulas [21]:

- Error of the first kind \( E_1 = \frac{n_1}{N_1} \times 100\% \), where \( n_1 \) – the number of attacks in the sample mistakenly classified as normal connections, \( N_1 \) – total number of attacks in the sample;

- Error of the second kind \( E_2 = \frac{n_2}{N_2} \times 100\% \), where \( n_2 \) is the number of normal connections in the sample mistakenly classified as attacks, \( N_2 \) – total number of normal connections.

As a result of calculations, the error of the 1st kind was \( E_1 = \frac{n_1}{N_1} \times 100\% = \frac{48}{866} \times 100\% = 5.54\% \), and the error of the 2nd kind was \( E_2 = \frac{n_2}{N_2} \times 100\% = \frac{16}{1200} \times 100\% = 1.33\% \).

The constructed neural network model correctly identified most of the connections characterizing the network scanning attack. The obtained values of errors of the first and second kind are also not critical, and, therefore, the results of neural network modelling are adequate.

**IV. SUMMARY**

Although network scanning attacks are not very harmful by themselves, their main danger is the further implementation by an attacker of other types of attacks on vulnerabilities identified during the network scan. Identifying network scanning attacks is challenging because of the challenge of identifying malicious connections from general network traffic. In this work, a neural network approach was used to detect network scanning attacks. The constructed neural network model successfully coped with the task of identifying and classifying malicious connections. Thus, the model is adequate and suitable for use in intelligent detection systems for network scanning attacks [22]. As a direction for further research, it is proposed to develop other models and evaluate the results obtained on the basis of other methods of data mining [23-28]. In addition, the actual construction and practical use of intelligent decision support systems [29-34] to identify and prevent network attacks on computer systems.

**V. CONCLUSIONS**

Thus, the research has been used to solve the problem of detecting network scanning attacks based on the construction of a neural network model, as well as assessing its effectiveness. The results of the studies have shown the effectiveness of the proposed approach to solving the problem. The constructed neural network showed high accuracy in terms of minimizing errors of the first and second kind. This indicates its effectiveness and practical use for detecting and preventing network scanning attacks.

**ACKNOWLEDGEMENTS**

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

**REFERENCES**


Oleg Yurievich Panishchev is a researcher at the Research Laboratory of Near Space Research, Kazan (Volga Region) Federal University. He graduated from Yelabuga State Pedagogical University with a degree in Physics, Informatics and Computer Science in 2003. Research interests: Time Series Analysis, methods of analysis of non-equilibrium distributed systems.

Artur Tagirovich Makridin is a Research Laboratory Assistant for Microwave Design and Radio Telecommunications, Kazan (Volga Region) Federal University. He graduated from Kazan (Volga Region) Federal University with a degree in radio physics in 2018. Research interests: methods of mathematical modelling of physical processes, automation of scientific experiments.

