Development of Algorithmic Solutions for Solving the Problem of identifying Network Attacks Based on Adaptive Neuro-Fuzzy Networks ANFIS

Denis Parfenov [a], Lubov Zabrodina [a], Irina Bolodurina [a] and Anton Parfenov [a]

[a] *Orenburg State University, Prospekt Pobedy, 13, Orenburg, 460018, Russia*

Abstract

Currently, the problem of detecting and classifying network attacks is one of the topical problems in ensuring network security. Existing intrusion detection systems, as a rule, do not provide the ability to identify all existing types of attacks, since today there is no universal algorithm for solving this problem. As part of this study, we proposed an approach to searching and detecting network attacks based on checking whether network traffic meets certain flexible rules. The problem of forming a base of fuzzy rules lies in the development of optimal functions and the creation of term sets that allow you to create a system of fuzzy conclusions that do not depend on the subjective assessments of specialists in a particular area. One of the effective methods used to solve this problem is the construction of a neuro-fuzzy network ANFIS. However, for its operation, it is necessary to carry out the preprocessing of the data array. We have proposed a solution that makes it possible to sequentially form data arrays using the C4.5 algorithm and the neuro-fuzzy network ANFIS. The study of a hybrid approach to the formation of adaptive neuro-fuzzy networks ANFIS based on various representations of fuzzy rules made it possible to improve the classification of incoming network traffic both in terms of accuracy and in terms of performance.

Keywords [1]2

Classifying network attacks, network traffic, multiclass fuzzy classification

1. Introduction

Currently, there is a tendency to change traditional approaches to the organization of network architecture. This is primarily due to the annual increase in the number of devices participating in the network data exchange. Recent trends are networks of "smart" mobile devices. The main traffic of smart devices, as a rule, is directed towards interaction with neighboring devices. Moreover, the number of such devices can reach several thousand, which, when the network devices interact with each other, generates significant amounts of data transmitted through the existing communication channels of telecom operators.

In networks with such a large number of devices, it is not technically possible to clearly define the border and areas of responsibility between customers and telecom operators. Nevertheless, telecom operators are responsible for ensuring the uninterrupted functioning of the network, which in turn requires the development of new approaches to ensuring network security [1].

The main tool for such networking is traffic analysis tools. To ensure effective security, various researchers propose approaches based on the placement of security elements, as well as data mining methods, including machine learning, etc. [2, 3].

YRID-2020: International Workshop on Data Mining and Knowledge Engineering, October 15-16, 2020, Stavropol, Russia

EMAIL: parfenovdi@mail.ru (Denis Parfenov); zabrodina97@inbox.ru (Lubov Zabrodina); prmat@mail.osu.ru (Irina Bolodurina); anton_parfenov@bk.ru (Anton Parfenov);
ORCID: 0000-0002-1146-1270 (Denis Parfenov); 0000-0003-2752-7198 (Lubov Zabrodina); 0000-0003-0096-2587 (Irina Bolodurina); 0000-0003-0411-3582 (Anton Parfenov)

One of the most common methods for detecting network attacks is to analyze network traffic data for complete coincidence with the existing database of network attack indicators. Due to the increasing complexity of threats, this method has low efficiency. One way to bypass this detection algorithm is to hide or proxy the attacker's IP address. In this case, in the event of a threat, the means of protection will block the hosts of legitimate users, and they will not be able to access resources. In addition to the above, methods for detecting network attacks, such as statistical methods, expert systems, and neural networks, are currently used. They are used both comprehensively and as separate tools for analyzing network traffic. This approach is more flexible and allows you to simplify the process of updating the security system.

Each of the above solutions is effective in its way, but only for traditional networks. Modern networks of telecom operators are increasingly using flexible topologies, characterized by variable volumes of flowing traffic. In such networks, the identified threat metrics captured in the security systems may not be effective.

Therefore, this work is devoted to the study and development of solutions for network protection tools based on multi-class fuzzy classification of network traffic to identify attacks.

The choice of the proposed approach is primarily because in networks with constantly changing topology there is no way to accurately identify attacks. An even more difficult task is to classify detected attacks according to known threat types. Nevertheless, the amount of data collected by network monitoring systems, in conjunction with well-known metrics of security systems, makes it possible, with a certain degree of probability, to identify harmful traffic.

2. Related works

Currently, the issue of detecting and classifying network attacks is one of the most relevant in ensuring network security. Existing intrusion detection systems usually do not provide the ability to identify all existing types of attacks, since there is no universal algorithm for solving this problem. In this regard, the problem of identifying network attacks is studied by various authors around the world.

For example, article [4] suggests an intrusion detection system built using the Feed Forward Deep Neural Network (FFDNN), which includes a feature extraction unit based on the wrapper method (Wrapper Based Feature Extraction Unit, WFEU) using the Extra Trees (ET) algorithm, which generates an optimal set of features. Experimental studies were conducted on the UNSW-NB15 data set. The authors found that the proposed approach exceeds the classical methods of machine learning and allows us to obtain a solution with high accuracy, both for binary classification and for multi-class classification of network traffic.

In work [5], an intrusion detection system based on the ensemble method (IBk(K-NN), Random Tree, REP Tree, j48graft, Random Forest) was developed to improve the accuracy and reliability of network traffic classification. Also, the filter-based attribute evaluation technique is used to reduce the number of attributes. To evaluate the performance of the proposed method, the NSL-KDD data set is selected. The ensemble method demonstrates an accuracy of 99.72% for binary classification and 99.68% for multi-class classification.

By the authors [6] an approach for multiclass classification of network traffic based on the use of deep convolutional neural networks is proposed. The performance of this approach was studied on the UNSW-NB15 dataset. The results of the experimental study allowed us to conclude that the proposed classifier is more effective than traditional machine learning classifiers, as well as dense neural networks. In addition, classification results based on convolutional neural networks are superior to previously obtained results in [7], including for complex types of attacks such as Analysis, Backdoor, Shellcode, Worms, with an increase in the F-score parameter by about 20%.

As part of the study [8], a two-stage intrusion detection system based on a stacked autoencoder and a softmax classifier are proposed. At the first stage, network traffic is classified as abnormal or secure, and at the second stage, abnormal traffic belongs to a certain class of intrusions. Studies of the proposed approach were conducted on two data sets KDD-CUP 99 and UNSW-NB15. This approach shows a high detection accuracy.

One of the promising areas of research for network security is the use of fuzzy logic methods.

In the study [9] an approach based on the use of fuzzy logic implemented using a genetic algorithm to detect intrusions into a wireless network is proposed. The KDD-CUP 99 data set was used to study the proposed approach. The system developed by the authors controls the distribution of messages about the connection Request message.

An approach based on the application of a genetic algorithm for fuzzy classification is also considered in the article [10]. the system proposed by the authors for detecting anomalies in real network traffic flows allows achieving detection accuracy of 96.53% and false-positive speed of 0.56%. This approach shows better results than CNN, SVM, and ACODS (Ant Colony Optimization for Digital Signature).

By the authors [11] a classifier based on a modified improved fuzzy min-max neural network (EFMN) for rapid identification of network attacks is proposed. The authors conducted a comparative analysis of the performance of the developed classifier with other standard classifiers, such as SVM, RIPPER (or JRIP), PART, ANFIS, and FMN. Based on the research, it was found that the proposed approach achieves a higher rate of detection of low-frequency attacks. Also, it surpasses the approaches considered in terms of indicators such as the frequency of false positives and recall time.

In the article [12] an approach using an adaptive neuro-fuzzy inference system (ANFIS) and a particle swarm optimization (PSO) method for detecting and preventing blackhole attacks in mobile ad-hoc networks (MANET) is proposed. According to the results of experimental studies, this approach has a good detection rate (on average 99.13%) and a low false alarm rate (on average 1.39%).

To ensure the security of MANET networks, a fuzzy system for detecting RREQ message flooding attacks is also proposed [13]. This system is based on the first-order Mamdani-type fuzzy inference system. The system proposed by the authors uses network parameters such as routing costs, bandwidth, and packet loss rate. As part of the experimental study, it was found that any deviation from the normal behavior of nodes is immediately detected by the proposed system.

As part of the study [14] a comparative analysis of «soft computing» methods, such as genetic programming (GP), fuzzy logic, artificial neural network (ANN), and a probabilistic model using clustering methods, for binary classification of network traffic packets. All studies were conducted on the NSL-KDD dataset. Based on the research, it was found that algorithms based on fuzzy logic show better performance. Thus, the FURIA algorithm provides a high detection rate of 99.69% with a low false alarm rate of 0.31% for a time of 78.14 sec, and the FRNN algorithm provides an accuracy of 99.51% and an acceptable false alarm rate of 0.49% for a computational time of 0.33 sec.

Research has shown that existing methods for identifying network attacks based on fuzzy classification allow us to determine the type of attacks with high accuracy. Most works based on fuzzy logic methods classify network traffic as secure or abnormal. However, determining the type of attack is an important aspect of network security. The research conducted on the UNSW-NB15 dataset does not address the problem of intersecting classes of different types of attacks in terms of analyzing similar characteristics. In this paper, we will describe the application of a multi-class fuzzy classification for network traffic and conduct a comparative analysis of the results obtained.

3. Problem statement

Consider a network of telecommunications service providers that provide end-user access to information systems. It is necessary to detect and classify malicious fragments of continuous network traffic. In other words, we will consider the network security problem as a multi-class classification of network traffic for detecting network attacks.

For the experimental study, we will use the UNSW-NB15 data set, which contains data on normal traffic and data from 9 classes of attacks [15]:

- Normal is secure data transactions;
- Fuzzers is an attack that causes a program or network to fail due to generating a large amount of random data that is passed to it for input;
- Analysis is an attack that involves scanning ports, sending spam, and embedding in HTML files;

- Backdoors is a method of bypassing the security mechanisms of the system in order to obtain hidden access to the computer or its data or programs;
- DoS is a denial-of-service attack on a server or network resource that makes it difficult for authorized users to access the computer;
- Exploits is an attack that leads to unexpected behavior of the host or network due to the attacker using known errors, failures, and vulnerabilities in the operating system or program;
- Generic is a technique that allows you to detect traffic encrypted with a block cipher;
- Reconnaissance is intelligence attack, i.e. an attack that collects information about a network to circumvent its security system;
- Shellcode is malware that transfers small parts of code used to exploit software vulnerabilities;
- Worms is attack, which is associated with self-replication of the attacking code.

The data set in question was developed in 2015 by the IXIA Perfect Storm tool in the cybersecurity lab of the Australian cybersecurity center (ACCS). The UNSW-NB15 dataset, unlike the KDD CUP 99 and NSL KDD datasets, fixed their main shortcomings and added information about modern types of attacks. The database used contains 2540044 network traffic records stored in four CSV files. Each record is represented as a set of 49 characteristics (attributes) of a specific data type. The corresponding sets containing 175341 and 82332 records are allocated for training and testing, respectively.

All the features considered in the data set can be divided into 5 main groups [16]:

- flow features;
- basic features;
- content features;
- time features;
- additional generated features.

Let's assume that the network attack model is represented as a time-ordered series of events (states) of a single node with an additive overlay of the attack profile on each element of the network of telecommunications service providers.

Let us consider the problem of constructing a system of neuro-fuzzy classification of network attacks from the point of view of predictive modeling, the solution of which can be obtained using supervised machine learning. Because the set of identified attacks is limited only from a practical point of view and is represented by the most common types of attacks, this task is a multi-class classification. Note that the neuro-fuzzy system allows transforming both continuous and categorical data into term sets, which significantly expands the set of variable characteristics that can be used.

Let us describe the formal mathematical formulation of the problem of classifying network attacks. Let us assume that information about the events taking place in the network is recorded with some rather short time interval. In this case, in addition to data about the device itself and its technical characteristics, information about the actions performed by end-users through the devices under consideration is also recorded.

Let the set X contain information about the states of all network objects $x_i \in X, i = 1,..,m$ with which some records of the event log are compared, i.e. $x_i = \{x_{i1}, x_{i2}, ..., x_{ik}\}$. The task of the multiclass classification of network attacks is to associate many types of attacks with network objects $Y = \{1, ..., K\}$.

Thus, the problem of identifying network attacks is that it is necessary to construct a mapping $f_c(X): X \to Y$ that allows describing the dependence between the recorded characteristics of network traffic and comparing the behavior of network objects with characteristics and choosing the most probable one in the absence of attacks and for a specific type of attack.

This study analyzes the classification of network attacks on the UNSW-NB15 dataset [16], which contains information about traffic with five different types of network attacks and the set has the form {Normal, Fuzzers, Generic, Reconnaissance, Exploits, DoS}. Note that the presented data on network traffic is collected by more than 40 characteristics and has more than 2.5 million records. Also, balanced sets for training and testing are compared to the data when analyzing the accuracy of the resulting classification models.

4. Approaches to identifying attacks based on systems of neuro-fuzzy classification

Within the framework of this work, to identify attacks, it is proposed to use the extraction of fuzzy rules from a decision tree built on the UNSW-NB15 training dataset (175341 unique records). The input features for building a decision tree are the network traffic characteristics extracted from the presented data set at the preprocessing stage. The output feature is the field with the class label of the attacking effect. The C4.5 algorithm is used to construct a decision tree.

```
Algorithm C4.5 (T)
Input: training data set T; attributes S.
Output: decision tree $Tree$.
if T is NULL then
    return failure
if S is NULL then
        return Tree as a single node with most frequent class label  in T
if |S| = = 1 then
        return Tree as single node S
set Tree ={}
for a ∈ S do
    set $Info(a,T)=0$, and $SplitInfo(a,T)=0$
    compute $Entropy(a)$
    for $v \in values(a,T)$ do
        set $T_{a,v}$ as the subset of $T$ wiyh attribute $a=v$
        $Info(a,T)+=\frac{|T_{a,v}|}{|T_a|}Entropy(a_v)$
        $SplitInfo(a,T)+=\frac{|T_{a,v}|}{|T_a|}\log\frac{|T_{a,v}|}{|T_a|}$
    $Gain(a,T)=Entropy(a)-Info(a,T)$
    $GainRatio(a,T)=\frac{Gain(a,T)}{SplitInfo(a,T)}$
set $a_{best}=\arg\max_a\{GainRatio(a,T)\}$
attach $a_{best}$ into $Tree$
for $v \in values(a_{best},T)$ do
     call $C4.5(T_{a,v})$
return $Tree$
```

Neural fuzzy networks are often used for more accurate identification of semi-structured data. Therefore, in addition to the already constructed decision tree, it is proposed to use neuro-fuzzy networks. It is proposed to use the Sugeno-Takagiya algorithm as an algorithm for fuzzy transformations. This method allows one to approximate arbitrary continuous functions dependent on many variables by the sum of functions depending on one variable with a given accuracy. Let us consider the basic ideas of constructing neuro-fuzzy ANFIS networks using the selected algorithm, and also present an approach to the formation of neuro-fuzzy inference.

The Sugeno-Takagi algorithm uses the following fuzzy rule model:

$$R_i\text{:IF } x_i \text{ eq } A_{i1} \text{ and } \ldots \text{and } x_n \text{ eq } A_{in} \text{ then } y=f(X)$$

Note that for each fuzzy Sugeno-Takagi rule, a cut-off level is selected, at which the rule conclusions are calculated. In the framework of this study, a first-order polynomial was used as an output function.

The neuro-fuzzy network ANFIS corresponding to the Sugeno-Takagi inference model is shown in Fig. 1 and has the following structure:

Layer 1. Responsible for matching the continuous input signal values of a specific term-set (fuzzification).

Layer 2. Determines the premises of fuzzy rules taking into account the input values of term sets and is interpreted as the degree of fulfillment of a certain rule.

Layer 3. Calculates the relative frequency of execution of the fuzzy rule (normalization).

Layer 4. Calculates the importance of each fuzzy rule and determines its contribution to the result.

Layer 5. Aggregates the results of fuzzy rules based on the identified importance.

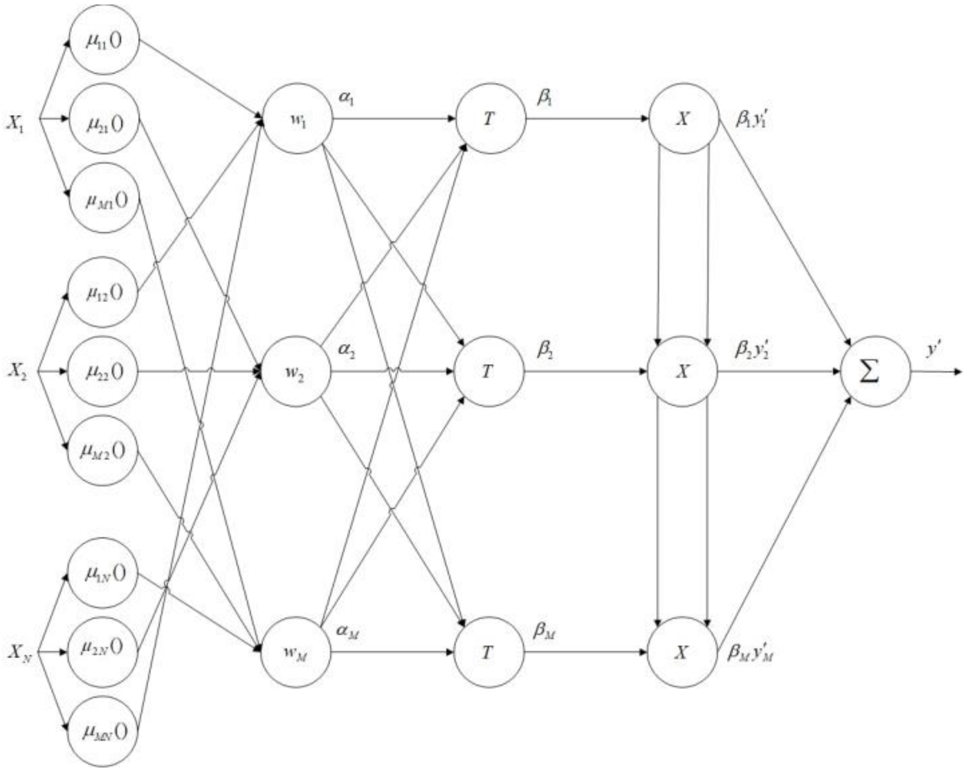

Figure 1: Scheme of neuro-fuzzy network ANFIS using Sugeno-Takagi inference

To test the identification of various types of attacks using neuro-fuzzy classification and fuzzy inference systems, to evaluate the effectiveness, we will conduct an experimental study of the classification of cybersecurity incidents on real network traffic.

5. Simulation results

To carry out a computational experiment to identify attacking influences using the proposed approach, combining the construction of rules using the C4.5 algorithm and the ANFIS neuro-fuzzy classification algorithm, a module for a traffic monitoring system in Python was implemented.

The proposed module was run on a virtual machine running Ubuntu 19.10 LTS Linux. As part of the study, we compared performance against three other machine learning approaches: Naïve Bayes, SVM, and KNN. To implement the proposed research plan, an experimental stand was built, which allows:

1. Use similar parameters of the experiment traffic generator on the equipment, if possible.
2. Use PCAP files with saved original experiment traffic on the hardware.

At the first step of the experimental study, the effectiveness of the constructed fuzzy inference systems was assessed to determine the class of attacking effects; it was estimated based on the analysis of network traffic on the UNSW-NB15 dataset.

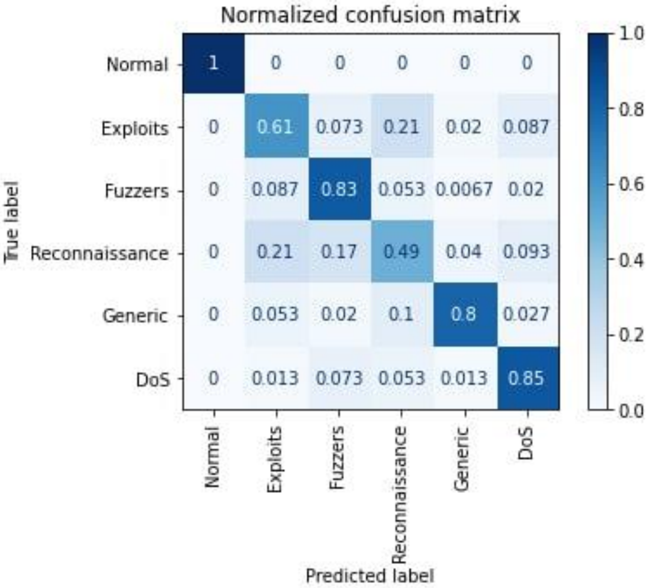

Figure 2: ANFIS error matrix using the Sugeno-Takagi algorithm

The results obtained are also presented as a general assessment of the effectiveness of identifying network attacks using measures of Accuracy, Precision, F-measure, and the number of truly positive classification results:

Table 1
Experimental results

Parameter	Accuracy	Precision	Recall	F-measure
ANFIS	86.15	85.60	86.60	86.40
Naïve Bayes	85.20	84.70	85.65	86.05
SVM	86.10	85.20	84.75	86.25
KNN	84.50	85.35	85.95	86.15
Multiclass Fuzzy Classification	85.00	85.15	85.87	85.97

At the next stage of the experiment, we test the load on the equipment created by each of the traffic analysis modules, measured in the IDS system. We assessed terms of the load on the device in terms of processor and RAM. And also established an import classification method that determines the network before making a decision. The analysis results are presented in Table 2.

Table 2
Experimental results of load equipment

Parameter	ANFIS	Multiclass Fuzzy Classification	Naïve Bayes	SVM	KNN
CPU %	<1%	2%	3%	4%	4%
RAM	<1%	<1%	1%	2%	3%
Delay	<1%	<1%	3%	3%	4%

6. Conclusion

As a result of the study, an analysis of network traffic was carried out for an approach to identifying attacks based on multi-class fuzzy classification. The results obtained showed the possibility of building a sufficiently accurate model to identify certain types of attacks. Also, a study was conducted on the performance of the proposed solution on real traffic. The obtained results of a general assessment of the effectiveness of network attacks using various measures of accuracy, the most optimal neuro-fuzzy classifier ANFIS network. Messages with a message about the information management system and security events. In future studies, it is planned to investigate other neuro-fuzzy classifier algorithms for the ANFIS network.

7. Acknowledgments

The study was carried out with the financial support of the RFBR in the framework of scientific project No. 20-07-01065, as well as a grant from the President of the Russian Federation for state support of leading scientific schools of the Russian Federation (NSh-2502.2020.9) and a grant from the President of the Russian Federation Federations for state support of young Russian scientists - candidates of sciences (MK-860.2019.9).

8. References

[1] I. Bolodurina, D. Parfenov, V. Torchin, L. Legashev "Development and Investigation of Multi-Cloud Platform Network Security Algorithms Based on the Technology of Virtualization Network Functions" 2018 International Scientific and Technical Conference Modern Computer Network Technologies (MoNeTeC) (2018): 1-7.

[2] I. Bolodurina, D. Parfenov "The development and study of the methods and algorithms for the classification of data flows of cloud applications in the network of the virtual data center" International Journal of Computer Networks and Communications 10(2) (2018): 15-22.

[3] A. E. Krasnov, D. N. Nikol'skii, D. S. Repin, V. S. Galyaev, E. A. Zykova "Detecting DDoS Attacks Using the Analysis of Network Traffic as Dynamical System" 2018 International Scientific and Technical Conference Modern Computer Network Technologies (MoNeTeC) (2018): 1-7.

[4] S. M. Kasongo, Y. Sun "A Deep Learning Method With Wrapper Based Feature Extraction For Wireless Intrusion Detection System" Computers & Security 92 (2020): 1-21.

[5] Kunal, M. Dua "Attribute Selection and Ensemble Classifier based Novel Approach to Intrusion Detection System" Procedia Computer Science 167 (2020): 2191-2199.

[6] R. Chapaneri, S. Shah "Detection of Malicious Network Traffic using Convolutional Neural Networks" 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (2020): 1-6.

[7] S. Potluri, S. Ahmed, C. Diedrich "Convolutional neural networks for multi-class intrusion detection system" Lecture Notes in Computer Science 11308 (2018): 225-238.

[8] F. A. Khan, A. Gumaei, A. Derhab, A. Hussain, "TSDL: A Two-Stage Deep Learning Model for Efficient Network Intrusion Detection" IEEE Access 7 (2019): 30373-30385.

[9] S. Sai Satyanarayana Reddy, P. Chatterjee, C. Mamatha, "Intrusion Detection in Wireless Network Using Fuzzy Logic Implemented with Genetic Algorithm" Computing and Network Sustainability. Lecture Notes in Networks and Systems 75 (2019): 425-432.

[10] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão, M. L. Proença "Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic" Expert Systems with Applications 92 (2018): 390-402.

[11] N. Upasani, H. Om "A modified neuro-fuzzy classifier and its parallel implementation on modern GPUs for real time intrusion detection" Applied Soft Computing 82 (2019): 1-16.

[12] H. Moudni, M. Er-rouidi, H. Mouncif, B. E. Hadadi "Black Hole attack Detection using Fuzzy based Intrusion Detection Systems in MANET" Procedia Computer Science 151 (2019): 1176-1181.

[13] B. Nithya, A. Nair, A. S. Sreelakshmi "Detection of RREQ Flooding Attacks in MANETs" Data and Communication Networks. Advances in Intelligent Systems and Computing 847

(2018): 109-121.

[14] J. E. Varghese, B. Muniyal "A Comparative Analysis of Different Soft Computing Techniques for Intrusion Detection System" Security in Computing and Communications. SSCC 2018. Communications in Computer and Information Science 969 (2019): 563-577.

[15] T. T. L. Le "Intrusion detection on the modern database UNSW-NB15 using multilayer neural network" Informatization and communication 1 (2017): 61-66.

[16] N. Moustafa, J. Slay "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)" 2015 Military Communications and Information Systems Conference (MilCIS) (2015): 1-6.