КОМП'ЮТЕРНІ НАУКИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UDC 632:654
DOI https://doi.org/10.32782/tnv-tech.2025.2.1

USING NEURAL NETWORKS IN NETWORK SECURITY PREDICTION

Antonenko A. V. *– PhD in Technical Sciences, Associate Professor,*
Associate Professor at Department of Standardization and Certification
of Agricultural Products
National University of Life and Environmental Sciences of Ukraine
ORCID ID: 0000-0001-9397-1209

Solobaiev S. H. *– Postgraduate Student at the Department of Computer Engineering*
State University of Information and Communication Technologies
ORCID ID: 0009-0008-6298-4777

Vostrikov S. O. *– Postgraduate Student at the Department of Computer Engineering*
State University of Information and Communication Technologies
ORCID ID: 0009-0008-8425-8872

Tkachenko O. V. *– Postgraduate Student at the Department of Computer Engineering*
State University of Information and Communication Technologies
ORCID ID: 0009-0009-6972-5388

Khodosov A. O. *– Master at the Department of Computer Engineering*
State University of Information and Communication Technologies
ORCID ID: 0009-0006-8759-5730

Ostapenko O. S. *– Master at the Department of Computer Engineering*
State University of Information and Communication Technologies
ORCID ID: 0009-0002-7488-2088

The article analyzes four algorithms: SVM, fuzzy clustering, K-Means and Apriori. We have described in detail the four stages of ensuring the security of network users and controlling their access. Research on a specially created reliable model for predicting network security. Intrusion pattern detection based on neural networks was developed, capable of identifying anomalies and attacks associated with abuse. This model performs three types of classification tasks: distinguishing between an attack and a normal state, as well as between ideal types of attacks or a normal state. In addition, the model demonstrates the classification accuracy, execution speed and amount of memory used. The main models include achieving high accuracy, reducing processing time and minimizing memory consumption. The proposed model based on neural networks successfully meets these goals. In the modern world, networks are becoming increasingly complex, closely interconnected and widely applicable. Network traffic volumes are growing almost exponentially, making networks more vulnerable to attacks by attackers who want to disrupt their functioning. Such vulnerabilities threaten economic losses and the leakage of confidential information. Therefore, there is an urgent need to improve methods for detecting vulnerabilities and improving the quality of network security prediction. A network security prediction model has been developed, aimed at reducing memory consumption and increasing the speed and accuracy of detecting various types of attacks. The results showed that the model is characterized by low memory consumption, fast attack detection time, and high accuracy. The methods used to create this model are characterized by simple implementation. They are also cost-effective, since the use of neural networks does not require additional costs. The model simplifies calculations, which makes it an effective solution for predicting network security. Thus, neural networks are a recommended tool for developing such models. In the future, it is planned to improve the models for more accurate and faster intrusion detection.

Key words: *SVM algorithm, K-Means algorithm, Apriori algorithm, fuzzy clustering algorithm, neural network.*

Антоненко А. В., Солобаєв С. Г., Востріков С. О., Ткаченко О. В., Ходосов А. О., Остапенко О. С. Використання нейронних мереж у прогнозуванні безпеки мережі

У статті проаналізовано чотири алгоритми: SVM, нечітку кластеризацію, K-Means та Apriori. Ми детально описали чотири етапи забезпечення безпеки користувачів мережі та контролю їхнього доступу. Дослідження присвячене створенню надійної моделі для прогнозування мережевої безпеки. Було розроблено модель виявлення вторгнень на основі нейронних мереж, яка здатна ідентифікувати аномалії та атаки, пов'язані зі зловживаннями. Ця модель виконує три види класифікаційних завдань: розрізнення між атакою та нормальним станом, а також між різними типами атак або нормальним станом. Крім того, модель демонструє показники точності класифікації, швидкості виконання та обсягу використаної пам'яті. Основні цілі моделі включають досягнення високої точності, скорочення часу обробки та мінімізацію споживання пам'яті. Запропонована модель на базі нейронних мереж успішно відповідає цим цілям. У сучасному світі мережі стають дедалі складнішими, тісно пов'язаними між собою та широко застосовними. Обсяги мережевого трафіку зростають майже експоненціально, що робить мережі більш уразливими до атак зловмисників, які прагнуть порушити їх функціонування. Такі вразливості загрожують економічним збиткам і витоку конфіденційної інформації. Тому існує нагальна потреба в удосконаленні методів виявлення вразливостей і підвищення якості прогнозування мережевої безпеки. Розроблена модель прогнозування безпеки мережі спрямована на зниження споживання пам'яті та покращення швидкості й точності виявлення різних типів атак. Результати показали, що модель характеризується низьким споживанням пам'яті, швидким часом виявлення атак і високою точністю. Методи, застосовані для створення цієї моделі, вирізняються простотою реалізації. Вони також є економічно вигідними, оскільки використання нейронних мереж не потребує додаткових витрат. Модель спрощує обчислення, що робить її ефективним рішенням для прогнозування мережевої безпеки. Таким чином, нейронні мережі є рекомендованим інструментом для розробки подібних моделей. У майбутньому планується вдосконалити моделі для ще більш точного та швидкого виявлення вторгнень.

Ключові слова: *алгоритм SVM, алгоритм K-Means, алгоритм Apriori, алгоритм нечіткої кластеризації, нейронна мережа.*

Introduction. In today's world, networks are becoming increasingly complex, interconnected, and widely used. Today, network traffic is growing almost exponentially. Networks are also becoming more vulnerable to attacks from hackers or anyone with malicious intent to disrupt network systems. Vulnerable networks are at risk of hitting

the economy and destroying confidential information. Thus, there is a need to improve network vulnerability detection mechanisms and improve network security prediction. The network security prediction model also aims to reduce memory consumption, as well as improve the detection of various types of attacks in terms of time and accuracy.

The aim of the study. The purpose of this article is to describe the creation of files used to detect anomalous attacks. Describe the details of the number of attacks or common cases for the anomaly or misuse-based attack detection process. Describe a method for removing unnecessary or unhelpful junk data to obtain the optimal amount of data for classifications. Use a neural network to detect various attacks. Distinguish between the different classification processes that occur in anomaly detection attacks, misuse detection attacks, and individual attack types. Detect anomalies using the classification of the occurrence of an attack or common case.

Analysis of recent research and publications. Attack detection models were considered. All models have a common goal – to detect vulnerabilities in the network more accurately, efficiently and faster. To achieve this goal, various algorithms were proposed [1]. An intrusion detection method based entirely on the K-means algorithm was proposed [2]. A hybrid intrusion detection algorithm based on the K-means and a selection tree was proposed [3]. Data feature monitoring was used to pre-process the 41-features in the statistical set [4].

Presentation of the main research material. Let us consider four algorithms, namely SVM algorithm, fuzzy clustering algorithm, K-Means clustering algorithm and Apriori algorithm. Next, we will detail 4 different steps of network user security and their access control. SVM algorithm is used to solve classification problems. Based on the basic construction of statistical principle, a kernel function is added to the calculation process to map the low-dimensional problem into the high-dimensional space and obtain a high-dimensional solution space. This means that using SVM algorithm will unlock hidden patterns in a large amount of data to reveal information. After loading the information, the system can identify the time series or the trend of the data and make accurate inferences [5]. The method of automatically obtaining the most optimal Gaussian parameters was also used to obtain the best hyposphere [6]. An improved version was developed in which the K-means algorithm was changed to a combination with Apriori to achieve the correct detection value of Root to Learn and User to Root according to the KDDCUP99 database information set to 98 % and 79 % [7]. The idea of using a set of dimensionality rules mixed with a single-class SVM was proposed [8].

The fuzzy clustering algorithm is as follows:

- Defining the similarity function.
- Establishing the appropriate fuzzy similarity matrix according to the similarity function.
- Computing the fuzzy relation and using the flat method. Also includes looking ahead when finding transitive closure.
- Classifying according to extreme thresholds and obtaining a specific dynamic clustering effect.
- The degrees are grouped together into a set of series.
- A pattern analysis algorithm is used to detect an attack with a possible attack sequence.
- Establishing the appropriate fuzzy similarity matrix, according to the similarity function.
- Computing the fuzzy relation and using the flat method. Also includes looking ahead when finding transitive closure.

- Classifying according to extreme thresholds and obtaining a specific dynamic clustering effect.
- The steps are grouped together into a set of candidate series.
- A pattern analysis algorithm is used to detect an attack mode with a possible sequence.

A new approach to fuzzy rule generation has been proposed, in which the clusters in the training pattern are set according to the fuzzy C-method clustering technique, according to the characteristics of each pattern and cluster [9].

The K-means clustering algorithm assumes that the required clustering values are known, but in fact, in security analysis, the values of k are usually unknown. And the choice of the initial clustering center of the K-features of the rule set is important. The K-feature miniseries is used to separate the normal data set and the attack data set into clusters of the same size separately, and the center of each cluster is used as the cluster index. There is a method for selecting representative instances from each cluster. The representativeness of an element is related to both density and distance. Higher representativeness increases the probability that an element will be selected as representative. After selection, each representative element is assigned a weight. This step not only reduces the size of the original data, but also preserves the maximum amount of information [1].

The Apriori algorithm is used to analyze the internal associations of fact security rules because it has high quality significance. The problem of this algorithm is the frequent scanning of the transaction database and the excessive set of additional expectation parameters [10, 11]. The values of minimum support and minimum confidence have a huge impact on the detection results. The Apriori algorithm was first proposed in [10].

It is necessary to ensure that sensitive data does not leak to those with malicious intent. Because of this, there are specific access control objectives for different users. There are 4 such basic user roles. They are data providers, data collectors, data miners, and decision makers [12]. Furthermore, it is extremely important to ensure that sensitive facts do not leak to those with malicious intent. Because of this, there are access control objectives for the extraordinary roles of network security consumers. There are 4 such basic user roles. They are information providers, information collectors, information miners, and decision makers [12]. For data providers, the goal of access control is to effectively control the amount of sensitive data that is disclosed to others. To achieve this goal, one can use protection tools to restrict others' access to their information, promote data in an auction to obtain sufficient compensation for the loss of privacy, or falsify information to hide one's true identity. For data collectors, the goal of access control is to launch useful facts for fact miners without revealing the identities of record providers and sensitive statistics about them. To achieve this goal, it is necessary to develop proper privacy models to quantify the possible loss of access control during exceptional attacks, and to apply strategies for anonymizing statistics [12, 14, 15]. For data miners, the goal of privacy preservation is to obtain correct record mining results while keeping sensitive statistics undisclosed neither within the record mining method nor in the mining results. To achieve this goal, a proper approach can be chosen to regulate information before executing positive mining algorithms. In addition, stable computing protocols can be used to ensure the security of private data and confidential statistics contained in the trained model. For decision makers, the goal of access control is to make the correct conclusion by approximating the reliability of the analysis results of the facts they receive. To achieve this goal, one can use provenance methods to suggest the returned records of the received facts or to create a classifier [12].

Two systems were created, one for anomaly-based attack detection and the other for abuse-based attack detection. These systems had approximately 4,500 records. The input data was divided into a data set (75 %) for training the neural network and a test data set (25 %) for the trained neural network. The first goal of the method was to simplify the facts to be processed. Simplification involves discarding features that are less useful. The advantage is that getting rid of features reduces the size of the data processed to improve the performance of the neural network. The disadvantage may be that if key attributes are accidentally removed, the accuracy of intrusion detection will decrease. All constant features were removed, and features that convey the largest percentage of variance were removed additionally. The "R" scripts were also checked for satisfactory performance based on the variance within samples. Attributes that do not contribute even 1 percent of the total variation in the fact set were left untouched. To deduce the intrusion detection for attacks that are mostly based on anomalies and attacks that are mostly based on misuse, two files were created: an anomaly dataset and a misuse dataset. In the anomaly dataset, the class or prediction variable was either normal, which represented an everyday occurrence, or an attack. The misuse dataset had a category variable of "Normal" or "Attack Name" that represented a specific type of attack, such as Smurf, NMap, Rootkit, etc.

Data cleaning was achieved for files consisting of dataset anomaly and dataset misuse. Using Weka to obtain a selection of dataset anomaly attributes and dataset misuse attributes meant much smaller attributes which contributed to the NN speedup. The package "neuralnet" is available in R and is open source. It has been used for IDS and neural network based analysis. The package agreement provided the capabilities for both neural network creation and classification.

In this work, more than 4,500 attack cases were considered. 10 attack types were selected including Neptune, NMap, PortSweep, Satan, Smurf, BufferOverflow, FTPWrite, GuessPassword, Back and Rootkit attacks. Anomaly-based intrusion detection was implemented in the attack detection process. Abuse-based intrusion detection was implemented for attack detection to offer confusion matrix, classification accuracy, implementation time and resource consumption. A classification was created among 10 attacks and a normal case. For some attacks, the system performed abuse detection.

Anomaly Detection Attack The results set given in Table 1 contains the details of the accuracy in detecting anomalous attacks, execution time, and memory consumption [13]. There is a classification of the occurrence of an attack or a normal case in this process. The values given in Table 1 indicate the number of records. The values of the coordinates (Attack, Attack) and the values of the coordinates (Attack, Normal) are 389 and 6, respectively. The value of the coordinates (Attack, Attack) is higher than the coordinates (Attack, Normal), as shown in Table 1. This means that an attack is present. The classification accuracy is high at 99.57 percent. The minimum execution time is 3.9979 seconds. The memory usage is minimal at 2191.311 Kbit, as shown in Table 2.

Misuse Detection Attack The set of results shown in Table 3 contains the details of the misuse attack detection accuracy, execution time and memory consumption [13].

Table 1

Detection of anomalous attacks

	Axis2	
Axis1	Anomaly attack	Normal
Anomaly Attack	389	6
Normal	3	763

Table 2

Results of detected attacks

Precision	99,57 %
Execution time	3.9979 c
Memory usage	2,189.311 Kbs

There is a classification between 10 attack types and normal cases. In Table 3, as in the case of anomaly detection, there are 2 axes, i.e. axis 1 and axis 2. The values mentioned in Table 3 indicate the number of records. The values of (Back, Back), (Buffer Overflow, Buffer Overflow), (Guess Password, Guess Password), (Neptune, Neptune), (Nap, NMap), (Port Sweep, Port Sweep), (Satan, Satan) and (Smurf, Smurf) coordinates are the highest among all the row values (values 67, 4, 12, 57, 75, 748, 63, 59 and 58 respectively). This indicates easier classification and higher probability of attack. The coordinates (FTP Write, FTP Write), (Rootkit, Rootkit) are significantly low (values 1 and 0 respectively). The coordinates (FTP Write, Normal) and (Rootkit, Normal) are 0 and 1 respectively, which is not the highest value of the row. Therefore, the normal case as shown in Table 3 is absent. As shown in Table 4, the classification accuracy is high and is 98.1 %. The execution time of 48.9282 seconds is higher than the previous case due to the larger amount of information, but still low for the information processed. The memory usage is 2,988.14 KB, increased due to the larger amount of information, but not very high.

Table 3

Abuse Detection Attack – Record Details

Axis1 / Axis2	**Return**	**Bufer Overflow**	**FTP**	**Guess Pass.**	**Neptune**	**NMap**	**Normal**	**Port Sweep**	**Rootkit**	**Satan**	**Smurf**
Return	67	0	0	0	0	0	0	0	0	0	0
Bufer Overflow	0	4	0	0	0	1	1	0	0	0	0
FTP	0	0	1	1	0	0	0	0	0	0	0
Guess Password	1	0	0	12	0	1	0	0	0	0	0
Neptune	0	0	0	0	57	0	0	1	0	0	3
NMap	0	0	0	0	0	75	0	0	0	0	0
Normal	0	0	0	0	0	0	748	0	1	0	0
Port Sweep	0	0	0	0	0	0	0	63	0	0	1
Rootkit	0	0	1	0	3	0	1	3	0	0	1
Satan	0	0	0	0	0	3	0	1	1	59	1
Smurf	0	0	0	0	0	0	0	0	0	0	58

Table 4

Misuse detection attack results

Precision	98.10 %
Execution time	48.9282 c
Memory usage	2988.14 Kbs

Conclusions. In the network security prediction model, the memory consumption was low, the time spent on detecting attacks was also low. The accuracy of detecting attacks was also high. The above methods used to design the model are also easy to design. In addition, the calculations are simplified by using this model. Therefore, the use of neural network is also an effective way to develop a network security prediction model. Thus, the use of neural networks is recommended for developing any type of network security prediction model. Future challenges are to develop models that will detect any intrusions even more accurately and faster.

BIBLIOGRAPHY:

1. Qiuhua, W., Xiaoqin, O., & Jiacheng, Z. (2019). A classification algorithm based on data clustering and data reduction for intrusion detection system over big data. KSII Transactions on Internet and Information Systems, 13(8), 3714–3732.

2. Jianliang, M., Haikun, S., & Ling, B. (2009). The application on intrusion detection based on K-means cluster algorithm. In 2009 International Forum on Information Technology and Applications (pp. 150–152). Chengdu, China.

3. Aung, Y.Y., & Myat, M.M. (2018). Hybrid intrusion detection system using K-Means and classification and regression trees algorithms. In 2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 195–199). Kunming, China.

4. Ravale, U., Marathe, N., & Padiya, P. (2015). Feature selection based hybrid anomaly intrusion detection system using K-Means and RBF kernel function. Procedia Computer Science, 45, 428–435.

5. Xiaoyi, H. (2020). Network security situation prediction based on grey relational analysis and support vector machine algorithm. International Journal of Network Security, 22(2), 177–182.

6. Xiao, Y., Wang, H., & Xu, W. (2015). Parameter selection of Gaussian kernel for one-class SVM. IEEE Transactions on Cybernetics, 45(5), 927–939.

7. Song, C., & Ma, K. (2009). Design of intrusion detection system based on data mining algorithm. In 2009 International Conference on Signal Processing Systems (pp. 370–373). Singapore.

8. Rochim, A.F., Aziz, M.A., & Fauzi, A. (2019). Design log management system of computer network devices infrastructures based on ELK stack. In 2019 International Conference on Electrical Engineering and Computer Science (ICECOS) (pp. 338–342). Batam Island, Indonesia.

9. Jin, C., Ye, Z., Wang, C., Yan, L., & Wang, R. (2018). A network intrusion detection method based on hybrid rice optimization algorithm improved fuzzy C-means. In 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (pp. 47–52). Lviv, Ukraine.

10. Han, J.W., & Kamber, M. P. (2011). Data mining: Concepts and techniques (3rd ed.). Elsevier Science.

11. Huang, Z. (2019). Research and implementation of intrusion detection based on host log. In Proceedings of the International Conference on Big Data Engineering (pp. 98–106).

12. Lei, X., Chunxiao, J., Jian, W., Jian, Y., & Yong, R. (2014). Information security in big data: Privacy and data mining. IEEE Access, 2, 1149–1176.

13. Pratik, M., Saunil, D., & Ravina, D. (2015). Intelligent security systems – Intrusion detection system. Retrieved from https://github.com/jgera/Network-Intrusion-Detection-System/blob/master/report.pdf

14. Цвик О.С. (2023). Аналіз і особливості програмного забезпечення для контролю трафіку. Вісник Хмельницького національного університету. Серія: Технічні науки, (1).

15. Твердохліб А. О., Коротін Д. С. (2022). Ефективність функціонування комп'ютерних систем при використанні технології блокчейн і баз данних. Таврійський науковий вісник. Серія: Технічні науки, (6).

REFERENCES:

1. Qiuhua, W., Xiaoqin, O., & Jiacheng, Z. (2019). A classification algorithm based on data clustering and data reduction for intrusion detection system over big data. KSII Transactions on Internet and Information Systems, 13(8), 3714–3732.

2. Jianliang, M., Haikun, S., & Ling, B. (2009). The application on intrusion detection based on K-means cluster algorithm. In 2009 International Forum on Information Technology and Applications (pp. 150–152). Chengdu, China.

3. Aung, Y. Y., & Myat, M. M. (2018). Hybrid intrusion detection system using K-Means and classification and regression trees algorithms. In 2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 195–199). Kunming, China.

4. Ravale, U., Marathe, N., & Padiya, P. (2015). Feature selection based hybrid anomaly intrusion detection system using K-Means and RBF kernel function. Procedia Computer Science, 45, 428–435.

5. Xiaoyi, H. (2020). Network security situation prediction based on grey relational analysis and support vector machine algorithm. International Journal of Network Security, 22(2), 177–182.

6. Xiao, Y., Wang, H., & Xu, W. (2015). Parameter selection of Gaussian kernel for one-class SVM. IEEE Transactions on Cybernetics, 45(5), 927–939.

7. Song, C., & Ma, K. (2009). Design of intrusion detection system based on data mining algorithm. In 2009 International Conference on Signal Processing Systems (pp. 370–373). Singapore.

8. Rochim, A. F., Aziz, M. A., & Fauzi, A. (2019). Design log management system of computer network devices infrastructures based on ELK stack. In 2019 International Conference on Electrical Engineering and Computer Science (ICECOS) (pp. 338–342). Batam Island, Indonesia.

9. Jin, C., Ye, Z., Wang, C., Yan, L., & Wang, R. (2018). A network intrusion detection method based on hybrid rice optimization algorithm improved fuzzy C-means. In 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (pp. 47–52). Lviv, Ukraine.

10. Han, J. W., & Kamber, M. P. (2011). Data mining: Concepts and techniques (3rd ed.). Elsevier Science.

11. Huang, Z. (2019). Research and implementation of intrusion detection based on host log. In Proceedings of the International Conference on Big Data Engineering (pp. 98–106).

12. Lei, X., Chunxiao, J., Jian, W., Jian, Y., & Yong, R. (2014). Information security in big data: Privacy and data mining. IEEE Access, 2, 1149–1176.

13. Pratik, M., Saunil, D., & Ravina, D. (2015). Intelligent security systems – Intrusion detection system. Retrieved from https://github.com/jgera/Network-Intrusion-Detection-System/blob/master/report.pdf

14. Tsvyk O. S. (2023). Analiz i osoblyvosti prohramnoho zabezpechennia dlia kontroliu trafiku. Visnyk Khmelnytskoho natsionalnoho universytetu. Ceriia: Tekhnichni nauky, (1).

15. Tverdokhlib A. O., Korotin D. S. Efektyvnist funktsionuvannia kompiuternykh system pry vykorystanni tekhnolohii blokchein i baz dannykh. Tavriiskyi naukovyi visnyk. Seriia: Tekhnichni nauky, 2022, (6).