وزارة التعليم العالي والبحث العلمي
جامعة ميسان
كلية التربية الاساسية

مجلة ميسان للدراسات الاكاديمية

العلوم الانسانية والاجتماعية والتطبيقية

ISSN (Paper)- 1994-697X
(Online)- 2706-722X

المجلد 23 العدد 49 السنة 2024

مجلة ميسان
للدراسات الاكاديمية

العلوم الانسانية والاجتماعية والتطبيقية

كلية التربية الاساسية - جامعة ميسان - العراق

ISSN (Paper)-1994-697X
(Online)-2706-722X

مجلد (23) العدد (49) اذار (2024)

www.misan-jas.com

journal.m.academy@uomisan.edu.iq

رقم الايداع في دار الكتب والوثائق بغداد 1326 في 2009

Enhancement methods of intrusion detection systems using artificial intelligence methods (TLBO)Algorithm.

Mohammed Saeed Hashim Al-Hammash
mohammed.s.alhammash@uomustansiriyah.edu.iq
Information Technology Center, Mustansiriyah university
https://orcid.org/ 0009-0006-3313-1797
Haitham Maarouf
hmaarouf@mubs.edu.lb
Modern University for Business and Science_ Lebanon

ISSN (Paper) 1994-697X

ISSN (Online) 2706-722X

DOI:
https://doi.org/10.54633/2333-023-049-010

Abstract:

Many methods have been used to build intrusion detection system based on the objective to be achieved in the prescribed manner. Hybrid methods (multiple methods) usually give better results and accuracy. The recent developments and popularization of network & information technologies have necessitated the need for network information security. Human-based smart intrusion detection systems (IDSs) are built with the capability to either warn or intercept network intrusion; this is not possible with the conventional network security systems. However, most information security studies have focused on improvement of the effectiveness of smart network IDSs. This study used TLBO algorithm as a feature selection algorithm to choose the best subset features and SVM classifier to classify the packet if it is intrusion or normal packet, two machine learning datasets used to test the proposed algorithm, the results show that the proposed algorithm perform better than many of the existing work in IDS.

Keywords: IDS, TLBO, SVM, Feature Selection

1.Introduction:

Cyber-attack refers to the deliberate alteration, disruption, destruction, deception, and degrading of computer networks or systems by an intruder over a sustained period [1, 2]. For accurate detection of network attacks, several approaches have been developed with the aim of detecting, preventing, and reducing cyber-attacks-related damages to computer systems/networks [3]. Such approaches include the use of firewalls, intrusion detection systems (ISD) and intrusion prevention systems (IPS). Cyber-attacks can be launched in several ways; however, system and network intrusions remain the commonest forms of cyber-attack [4]. Regarding IDS, they are either hardware or software systems that can monitor and detect network intrusions in real time to prevent network exploitation [5]. The financial implication of cyber-attacks on U.S. organizations has been quantified in many studies [6]. Research also indicated that cyber-attacks were increasing in frequency and complexity growing by 11% between 2012 and 2013 in U.K. small businesses [7] and from 1,334 incidents in 1993 to 137,529 incidents in 2003[8] resulting in a need for improved defense mechanisms [9]. An increased dependence on information across countries, organizations, and the military is a reason given for the need for improved methods of blocking cyber-attacks [10] This dependence on information increases the overall value to the organization. Being that organizations mostly rely on information, it has become necessary to ensure adequate security of

such information for the safety of the society and the economy [11]. Both network and system intrusions are greatly increasing in complexity and frequency at the same time. Before now, network attacks are targeted at just a single system component, but today, network attacks involve the use a several techniques to target numerous segments of the network [12]. For instance, the viruses that are available today are more complicated when compared to those of the early years of computing; this has increased the challenges related to the detection and elimination of such viruses from computer systems [13].

2. Problem statement:

There are several types of latency that can affect intrusion detection systems. Network latency is defined as the delay from the initial transmission of the packet header at the source to the reception of the end of the packet at the destination [18]. This encompasses the very first part of a transmission until the very end of the transmission and indicates how long it takes to transmit information between two locations and is a measure of that delay. In contrast, latency period, or detection latency as used in this research, is a measure of the time between when an attack initially starts and the detection of the event by the system. A third latency factor is the period of time between the start of the attack and the start of corrective action by system administrators. All three of these latency factors can increase in duration as network traffic increases and the workload approaches a saturation point[19]. Detection latency is seldom used for measuring IDS systems but is an important measure to study. A goal used in past research and IDS development was a 100% detection rate, but a need is being recognized for reducing the amount of time that it takes to detect intruders with a goal of limiting the amount of potential damage done. Detection latency is a critical measure and should be developed as a key metric[20].

government/military, and medical/health care [6]. Those attacks and breaches could make a huge impact on the customers or on the organization as a whole; however, Pfleeger in his book [134] stated that dealing with harm is sought in several ways as follows:

Prevent it, by preventing attackers from doing what they intended to do or closing,vulnerabilities. Deter it, by mitigating attacks, Deflect it, by deluding attackers to attack other attractive targets, Detect it as it happens or sometimes after the fact , Recover from their effects.

Each one of these tasks has its own procedures in order to design and implement a system with the above capabilities. This research is focused primarily on how to detect these breaches and attacks.

Our proposed method is to reduce the latency time by reducing the processed data to detect the intrusion. the proposed study to minimize the classification time for intrusion detection, however, the objectives are To present a Teaching learning-based optimization technique (TLBO) for intrusion detection subset feature selection.

3. Methodology proposed:

The necessity for network data security has arisen as a result of current advances and popularization of information technology. Human-based smart intrusion detection systems (IDSs) are capable of warning or intercepting network intrusions, which is not achievable with traditional network security systems.

Recently, the teaching–learning-based optimization algorithm (TLBO) was developed as a new metaheuristic and applied to numerous intractable optimization tasks with a reasonable level of success. TLBO performs better than most of the existing algorithms as it requires fewer tuning parameters during execution when compared with the other existing algorithms. Therefore, this study proposes the merging of TLBO and supervised ML techniques for FSS in binary classification problems (BCPs) for ID. The selection of the least number of features in FSS without compromising accuracy is a MOO problem where the number of features is the first objective while detection accuracy is the second objective. Being that TLBO performs better than other metaheuristics, this study employed ITLBO and a set of supervised ML techniques for optimal feature subset selection.

ISSN (Paper)-1994-697X (Online)-2706-722X

4. Preprocessing proposed work:

This work based on TLBO-SVM method , the evaluations of the results are represented into parts. Firstly, this part represents the proposed of basic TLBO algorithm based on the intrusion-detection system, which it's compared with TLBO. The methods represented with a different number of features that are selected by the methods, for fairer comparison number of features are equalled in each comparison but not the same features for each method (each method choose the optimal subset features based on its performance).

Applied Machine Learning

We used three machine learning approaches to evaluate the ITLBO solutions (LR, SVM, and ELM). SVM is well-known for its efficacy in binary classification; ELM is a newly introduced but promising classifier. LR is a common, fast, and easy-to-implement classifier; SVM is well-known for its effectiveness in binary classification; and ELM is a newly introduced but promising classifier.

Logistic regression: Classification with LR is done by evaluating the probability of an event occurring based on the similarity of data points. It uses a sigmoid function to calculate the probability of an event occurring. If an event's occurrence probability is greater than 0.5, the LR classifies it as "occurred" or "not occurred," depending on the situation.

SVMs: The development of a dividing line between provided data points [19] is used to conduct classification tasks using SVM, with data points closest to this line being labeled as SVs. The maximum margin between the SV and the line of the classes is maximized iteratively to create this line. The assumption behind this concept is that increasing the margin will lower the generalization error.

ELM: With a hidden layer, an input layer, and an output layer, the ELM is created as a feedforward neural network (FFNN). The input layer feeds the training data into the model, which is subsequently weighted and transferred to the hidden layer through a function. Between the hidden and output layers, a comparable modification is performed. The FFNN requires iterative parameter adjustment; whereas, the ELM does not require parameter tuning. As a result, when compared to traditional FFNNs, ELM takes less time to train.

5. Discussion experiments and results:

In this section, the experimental scenario, problem cases, and experiment results are all described. The studies were carried out on two incursion datasets (KDDCUP 99 and CICIDS), which were limited to allow only two classes due to the focus on binary classification (normal and intrusion). Then, 30% of each dataset was randomly selected for training, and 100% of the data were used in the testing phase.

KDDCUP'99: This dataset was first used at the 3rd International Knowledge Discovery and Data Mining Tools Competition to build a network intrusion system. The DARPA Intrusion Detection Evaluation Program was established by the MIT Lincoln Lab in 1998 as a simulated environment for obtaining raw TCP/IP dump data for a local area network (LAN) (Gao, Li, Zhang, Lin, & Ma, 2019). It was created to examine the effectiveness of different intrusion detection systems. The KDDCUP '99 dataset contest used a variant of the DARPA'98 dataset (Abdulhammed et al, 2019). The DARPA98 dataset is made up of compressed raw tcpdump data from seven weeks of network traffic; it's around 4 GB in size and can be broken down into about 5, 000,000 connection records, each of which is about 100 bytes long. The collection contains roughly 2, 000,000 connection records from two weeks of testing. The KDD training dataset contains approximately 4,898,431 single connection vectors with 41 characteristics, each classified as normal or attack (with a specific type of attack). The assault types in the dataset were classified into four groups.

i. Probing attack: The attacker tries to gather network information just to get beyond the security restrictions on the network.

ii. Denial of service (DOS): The attacker makes the system too busy to process valid requests, denying genuine network access.

www.misan-jas.com ISSN (Paper)-1994-697X (Online)-2706-722X

iii. User to root (U2R): The attacker acquires network access by logging in as a legitimate user and then exploiting security flaws in specific systems to gain root access.
iv. Remote to local (R2L): The invader takes advantage of system flaws by sending packets over a network to acquire local access as a legitimate user.

Table 1. KDD dataset

Attack classes	22 types of attacks	No. of instances
Normal		972781
DoS	smurt, neptune, pod, teardrop, back, land,	3883370
R2L	phf, ftp-write, imap, multihop, warezclient, warezmaster, spy, guess password	1126
U2R	perl, loadmodule, buffer-overflow, rootkit	52
Probing	portsweep, ipsweep, satan, nmap	41102
Total		4,898,431

The CICIDS 2017 dataset is comprised of both normal and the most recent form of attacks; it mimics real-world data (PCAPs). This dataset also contains the network traffic analysis results collected using a CICFlowMeter; the flows are labeled based on date, source and destination IPs, source and destination ports, protocol, and attack. Anonymity, attack diversity, complete capture, available protocols, feature set, complete network configuration, total interaction, complete traffic, labelling, metadata, and heterogeneity were all present in the dataset (Vinayakumar et al., 2019). This dataset has around 3,057,503 rows, which are organized into 8 files and each row contains 79 features. Each row in the CICIDS 2017 is labeled either as benign or as one of the 14 types of attack. Table 6 presents a summary of the attack types distribution and the benign rows in the CICIDS 2017 dataset.

Table 2. Types of attacks

Attack class	14 types of attacks	No. of instances
Benign (normal)		2,499,857
DOS	DDoS, slowloris, Heratbleed, Hulk, GoldenEye, Slowhttptest	380699
PortScan	Portscan	158930
Bot	Bot	1966
Brute-Force	FTP-Patator, SSH-Patator	13835
Web Attack	Web attack XSS, web attack SQL injection, web attack brute force	2180
Infiltration	Infiltration	36
Total		3,057,503

Experiments were carried out on a machine equipped with an Intel Core i7-4810 processor running at 2.80 GHz and 8 GB of main memory. MATLAB 2017a was used to finish the classification portion of the methods. The population size and the number of generations are two crucial criteria to determine before running TLBO. A higher value for these parameters assures great result accuracy at the cost of increased computation time. It takes a long time to investigate a new person.

ISSN (Paper)-1994-697X (Online)-2706-722X

The parameter settings for the simulation model are shown in Table 3.

Table 3. Parameter settings

Parameter	Value
Population size	20
Number of generations	40
Crossover type	Half-uniform
Mutation type	Bit-flip
Size of training data	30% (Random)
Size of testing data	100% (Full dataset)

Tables 4 and 5 present the accuracy results for both datasets, respectively. The accuracy result of the KDDCUP99 dataset is presented in Table 8.

Table 4. Accuracy result of the KDDCUP99 dataset

Classifier	**TLBO**	
LR	No. of features	Accuracy
	3	0.995
Total time	12.2512	
SVM	3	0.995
	6	1.00
Total time	2382.3301	
ELM	3	0.97
	4	0.99
	5	0.995
	8	1.00
Total time	4.0717	

As seen in Table 4, The number of features, accuracy, and execution time for each ML were calculated. The red figures represent TLBO's best results. TLBO regularly outperforms other algorithms utilizing the three ML techniques, and it also outperforms other algorithms employing the LR and SVM ML techniques in terms of time accuracy. With ELM, however, TLBO has a faster execution time. Table 9 displays the results of the CICIDS2017 dataset.

On the CICIDS 2017 dataset, TLBO consistently shows good accuracy using the three ML techniques. matrix for the best accuracy from the application of ITLBO to the KDDCUP99 dataset (Table 6). Table 7 shows the confusion matrix for the ICIDS dataset.

Table 5. Accuracy result of the CICIDS 2017 dataset

Classifier	**TLBO**	
LR	No. of features	Accuracy
	14	0.94
	15	0.965
	27	0.97
Total time	33.06	
SVM	24	0.84
	26	0.92
Total time	4161.3924	
ELM	13	0.86
	15	0.885
	16	0.905
	19	0.91
	20	0.92
Total time	3.4071	

ISSN (Paper)-1994-697X (Online)-2706-722X

Table 6. KDDCUP99 confusion matrix

	Predicted intrusion	Predicted normal
Actual intrusion	3,923,506	2,144
Actual normal	83	972,698

Table 7. CICIDS dataset confusion matrix

	Predicted normal	Predicted intrusion
Actual normal	2,606,223	25,402
Actual intrusion	58,719	367,159

We can calculate other evaluation metrics based on the confusion matrix for the ITLBO algorithm, with the first one being the detection rate given in Eq. (1).

Detection Rate =TP/(TP+FP) (1)

Another metric is the error rate, which is calculated based on Eq. (2).

Error Rate =(FP+FN)/(TP+FP+FN+TN) (3)

The other metrics used are given in Eqs. (3) to (8).

False Positive Rate (FPR) = FP/ (FP+TN) (4)

False Negative Rate (FNR) = FN / (TP+FN) (5)

True Postive Rate(TPR) =Recall = TP / (TP+FN) (6)

Precision = TP / (TP+FP) (7)

F- Measure = 2*((Precision*Recall) / (Precision + Recall)) (8)

Table 8 shows the result of Eqs. (3) to (8).

Table 8. Results of Eqs. (3) to (8)

	KDDCUP99	CICIDS2017
Detection rate	0.9995	0.99
Error rate	0.0045	0.027
FPR	0.000085	0.13
FNR	0.00054	0.0096
TPR	0.999	0.99
Precision	0.9995	0.99
F-measure	0.998	0.99

Another statistical test (T-test) was applied to show the superiority of ITLBO over TLBO. Table 9 showed the p-values & T-values; the small values manifests the significance of the ITLBO.

Table 9. T-test results

	KDDCUP99	CICIDS2017
P-value	0.0156	0.0068
T-value	3.174	4.044

ISSN (Paper)-1994-697X (Online)-2706-722X

6. Conclusion

This study adopted a system engineering approach to apply machine learning techniques to network intrusion detection problem. The existing relevant literature was reviewed, and methods were explored towards combining the classification decisions of a diverse set of "learners" in a way that will offer good performances. This study introduced 3 new methods to ML which were benchmarked against an established intrusion detection dataset (NSL-KDD).

Three methods were deployed in this work; firstly, the TLBO-SVM was used to work as subset feature selection in intrusion detection system. The results of the ITLBO-SVM were compared to that of SVM algorithm to evaluate the proposed model. The results showed TLBO-SVM to have a higher accuracy compared to SVM because of the reduction have been done on the data and remove irrelevant features. Secondly, to reduce the impact of randomized parameters select in SVM algorithm. The results of this method were compared to TLBO-SVM based on several standard evaluation criteria.

Further studies are recommended in the following aspects: finding a modern method of parameters selection, developing novel methods of classification combination, addition of learning & experimentation with data pre-processing, incremental and adversarial learning. Furthermore, studies are required on techniques for on-line processing, hybrid techniques of combining partial batching with online techniques and unsupervised learning could help in achieving practical IDSs that are applicable in a real-world network operating scenario. Studies should also focus on testing the proposed methods with other intrusion-detection data sets.

8.References:

[1] Mohammed, M. A., Salih, Z. H., Ţăpuş, N., & Hasan, R. A. K. (2016, September). Security and accountability for sharing the data stored in the cloud. In 2016 15th RoEduNet Conference: Networking in Education and Research (pp. 1-5). IEEE.

[2] ebar, H., Dacier, M., & Wespi, A. (2000). A revised taxonomy for intrusion-detection systems. In Annales des télécommunications (Vol. 55, No. 7-8, pp. 361-378).

[3] Gautam, S. K., & Om, H. (2016). Computational neural network regression model for Host based Intrusion Detection System. Perspectives in Science, 8, 93-95.

[4] Manzoor, I., & Kumar, N. (2017). A feature reduced intrusion detection system using ANN classifier. Expert Systems with Applications, 88, 249-257.

[5]Ashfaq, R. A. R., He, Y. L., & Chen, D. G. (2017). Toward an efficient fuzziness based instance selection methodology for intrusion detection system. International Journal of Machine Learning and Cybernetics, 8, 1767-1776.

[6]Chiba, Z., Abghour, N., Moussaid, K., El Omri, A., & Rida, M. (2018). A novel architecture combined with optimal parameters for back propagation neural networks applied to anomaly network intrusion detection. Computers & Security, 75, 36-58.

[7]Wang, H., Gu, J., & Wang, S. (2017). An effective intrusion detection framework based on SVM with feature augmentation. Knowledge-Based Systems, 136, 130-139.

[8]Pozi, M. S. M., Sulaiman, M. N., Mustapha, N., & Perumal, T. (2016). Improving anomalous rare attack detection rate for intrusion detection system using support vector machine and genetic programming. Neural Processing Letters, 44, 279-290.

[9]Tao, P., Sun, Z., & Sun, Z. (2018). An improved intrusion detection algorithm based on GA and SVM. Ieee Access, 6, 13624-13631.

[10]Raman, M. G., Somu, N., Kirthivasan, K., Liscano, R., & Sriram, V. S. (2017). An efficient intrusion detection system based on hypergraph-Genetic algorithm for parameter optimization and feature selection in support vector machine. Knowledge-Based Systems, 134, 1-12.

ISSN (Paper)-1994-697X (Online)-2706-722X

[11]Shang, W., Zeng, P., Wan, M., Li, L., & An, P. (2016). Intrusion detection algorithm based on OCSVM in industrial control system. Security and Communication Networks, 9(10), 1040-1049.

[12]Chen, Y., Li, Y., Cheng, X. Q., & Guo, L. (2006, November). Building efficient intrusion detection model based on principal component analysis and c4. 5. In 2006 International Conference on Communication Technology (pp. 1-4). IEEE.

[13] Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications, 41(4), 1690-1700.

[14]Moon, D., Pan, S. B., & Kim, I. (2016). Host-based intrusion detection system for secure human-centric computing. The Journal of Supercomputing, 72, 2520-2536.

[15]Papamartzivanos, D., Mármol, F. G., & Kambourakis, G. (2018). Dendron: Genetic trees driven rule induction for network intrusion detection systems. Future Generation Computer Systems, 79, 558-574.

[16]Rajeswari, L. P., & Kannan, A. (2008, January). An intrusion detection system based on multiple level hybrid classifier using enhanced C4. 5. In 2008 International Conference on Signal Processing, Communications and Networking (pp. 75-79). IEEE.

[17]Nia, F. Y., & Khalili, M. (2015, November). An efficient modeling algorithm for intrusion detection systems using C5. 0 and Bayesian Network structures. In 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI) (pp. 1117-1123). IEEE.

[18]Koucham, O., Rachidi, T., & Assem, N. (2015, November). Host intrusion detection using system call argument-based clustering combined with Bayesian classification. In 2015 SAI Intelligent Systems Conference (IntelliSys) (pp. 1010-1016). IEEE.

[19]Altwaijry, H. (2013). Bayesian based intrusion detection system. In IAENG Transactions on Engineering Technologies: Special Edition of the World Congress on Engineering and Computer Science 2011 (pp. 29-44). Springer Netherlands.

[20]Xiao, L., Chen, Y., & Chang, C. K. (2014, July). Bayesian model averaging of Bayesian network classifiers for intrusion detection. In 2014 IEEE 38th International Computer Software and Applications Conference Workshops.

ISSN (Paper)-1994-697X (Online)-2706-722X