Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 2, June 2020, pp.693-702

Addressing big data analytics for classification intrusion detection system

Keyan Abdul-Aziz Mutlaq[1], Hadi Hussein Madhi [2], Hassanain Raheem Kareem [3]
[1] Computer Center, University of Basrah
[2] College of Nursing, Misan University
[3] Mathematics Dep., College of Education, Misan University

Abstract

Currently, with the rapid developments communication technologies, large number of trustworthy online systems and facilities has been introduced. The cybersecurity is quiet on the rise threat from unauthorized; such security threats can be detected by an intrusion detection system. Thus, enhancing the intrusion detection system is main object of numbers of research and developers for monitoring the network security. Addressing challenges of big data in intrusion detection is one issue faced the researchers and developers due to dimensionality reduction in network data. In this paper, hybrid model is proposed to handle the dimensionality reduction in intrusion detection system. The genetic algorithm was applied as preprocessing steps for selecting most significant features from entire big network dataset. The genetic algorithm was applied to generate subset of relevant features from network data set for handling dimensionality reduction. The Support Vector Machine (SVM) algorithm was processed the relevant features for detecting intrusion. The NSL-KDD standard data was considered to test the performance of the hybrid model. Standard evaluation metrics were employed to presents the results of hybrid model. It is concluded that the empirical results of hybrid outperformed the performance of existing systems.

Keywords: Genetic algorithm, SVM, Intrusion detection system, Hybrid model

Corresponding Author:

Keyan Abdul-Aziz Mutlaq
Computer center, University of Basrah
Basrah, 61004, Iraq
E-mail: keyan.alsibahi@uobasrah.edu.iq

1. Introduction

Information and communication technologies are now influencing every aspect of society and the lives of individuals, thereby growing attacks on ICT systems. The communication technology needs concrete, integrated security solutions. The essential part of communication technology security is confidentiality, integrity and availability. In the last decades, the communication technologies have witnessed marvelous development. Enormous advancement in communication technologies have been seen in the begging of 21 century. These technologies obtain interconnected through worldwide network such as internet. However, the Information and communication technologies are used to showing the information over worldwide through web is prone to unauthorized access. In order to save our information form intrusion, it is necessary to secure the data. The security is essential part to grow and developing information and communication technologies for helping societies to improve their lives. The computer networks have different kinds of security for protecting the network data. The Intrusion Detection System (IDS) is one types of computer security use to protect the data from intrusion. But, with increasing develop Information and communication technologies have given more attention to engineering and developers to find robust method avoid the intrusion. The machine learning algorithms are one of solution that is discovered by researchers in field of computer security, for making information and communication technologies more safely. With using machine learning algorithms, the intrusion detection talented to detect attacks automatically and send message warning to users. Throughout, 2017, the planet witnessed some of the Internet era's greatest cyber threats for Uber and Yahoo data breaches [1], is defined as "burst attacks" has grown in scope and intensity. Burst attacks are fast and, like a few minutes, can occur in a small frame. The Cisco Cyber Security Reports show that this form of Distributed Denial of

Service (DDoS) attack was faced by 42 percent of organizations in 2017 [2]. In 2018, numbers of data every day exceeds petabytes, including traces left by internet users when they visit a website, mobile app, and network. Such follows or "log information" are turning out to be tremendous consistently as they are created by one source, however a great deal of sources now and again. Utilizing log information admirably can give a bit of leeway in identifying vindictive connections, in this way protecting the system against future attacks. In any case, since they strike in a brief timeframe, the brief timeframe range that programmers use can disable even great frameworks. The requirement for an ongoing discovery framework that can scale up to the measure of information being expended and act rapidly as far as reaction time can give an edge over these kinds of assaults. Enormous information examination for interruption identification and counteractive action of system security issues has been quickly standing out as it encourages the investigation of huge volumes of mind boggling and divided information with various positions from heterogeneous sources, distinguishes peculiarities and fights digital attacks. Ultra-high-dimensional information models can be made to profile stream information precisely on the web, which predicts and distinguish interruption and assaults in genuine time [3]. Enormous Data advancements like the Hadoop environment and stream preparing can store and break down huge heterogeneous datasets at a rapid, changing security investigation by:(A) assembling huge scale information from different inner and outer sources, for example, helplessness databases; (b) directing top to bottom information examination; (c) leading in-house stream information investigation; and (d) giving a coordinated perspective on security-related data. Because of the requirement for appropriate plan of big data examination apparatuses, framework investigators and planners need a cozy are required to have a private learning of their frameworks [4]. In this research work, we propose an adapting model to classify the cyber-attacks form the real network dataset. The big data set from real network is gathered to test the proposed model. The machine leaning algorithm is used to enhance the intrusion detection system over big data. For overcome the issue of intrusion detection system issues the feature selection methods is proposed to select the significant features from entire the dataset. The hybrid model is proposed to handle the dimensionality reduction in intrusion detection system. The genetic algorithm was applied to generate subset of relevant features from network data set for handling dimensionality reduction. The Support Vector Machine (SVM) algorithm was processed the relevant features for detecting intrusion. The NSL-KDD standard data was considered to test the performance of the hybrid model. The remainder of the paper is composed as pursues. In Section 1 presents introduction. The foundation investigation of this examination is discussed in section2. In section 3 displays the material and methods. The exact outcomes are appeared in section 4. At long last, section 5 gives conclusion of paper and future work.

2. Related work

Every time Intrusion detection has a major concern for researchers and developers and engineering in field of computer network. Scientific papers have always been concerned with detecting intrusion [5-6]. Despite Big Data rising to the surface along with machine learning algorithms, scientists are more interested in finding new solutions to this problem since the emergence of the modern Internet. Several techniques, ranging from data mining approaches to machine learning algorithms, have been used to detect intrusions over a network. Muhammad et al. [7] used Support Vector Machine (SVM) algorithm to detect intrusion by using knowledge discovery database (KDD) cup 99 datasets. The Apache Storm was applied for developing the system. They have used 13,600 packets in a second by using a single machine. The proposed system was obtained 92.60% accuracy on testing data. Mustapha et al. [8] applied four machine learning algorithms, namely Support Vector Machine (SVM), Naïve Bayes, Decision Tree and Random Forest. The UNSW-NB15 dataset was used to test the proposed system. The Apache Spark is implemented for processing the data. Pallaprolu et al. [9] presented the KNN algorithm to detect intrusion detection system; Apache Spark Streaming was used to detect zero-day intrusion. Gupta et al. [10] used feature selection algorithm to select significant feature for improving the classification process. The correlation-based feature selection method was applied over big data select subset features from entire the data. These features were processed by using five machine learning algorithms (Logistic Regression, SVM, Naïve Bayes, and Random Forest). Using big data technologies crowd sourcing, natural language processing, etc. [11-12]. Distributed file systems, cluster file systems, and parallel file systems are the main tools used in big data [13]. Sadly, redundant attributes and records make it a particularly complicated and daunting job to detect intrusion in big data analytics [14]. PCA was used to extract features from high-dimensional dataset attributes, especially redundant attribute datasets. This method was used by experts in the selection of IDS features [15]. Uncontrolled identification of anomalies in unstructured data. Numbers of existing intrusion detection system use machine learning, e.g., Support vector machines [16–18], K-Nearest

Neighbor (KNN) [19], Random Forest (RF) [20-21], etc. nevertheless, these algorithms create numerous false alarms and have a low detection rate for attacks in IDS. Kim et al. [22] presented a hybrid model for classification intrusion detection system by using the C4 decision tree classification algorithm and SVM algorithm. The NSL-KDD dataset was used to evaluate the hybrid system. Panda et al. [23] used Naive Bayes (NB) algorithm for anomaly detection. This algorithm is tested by using KDD Cup dataset and it is observed that the hybrid mode is outperform several existing IDS in terms of the low false alarm rate and low computation time with low cost. Aldhyani et al [27] applied hybrid model support vector machine and J48 with soft clustering to handle the ambiguity in intrusion detection system. Theyazn et al. [28] used machine learning algorithm to reduce dimensionality reduction in IDS. [29–31] used the internet of thing to classification the intrusion detection system. Potluri et al. [32] applied CNN-based detection method. They have tested the proposed model using to standard datasets-KDD and the UNSW-NB 15 datasets. Zhang et al. [33] Applied XGBoost model to extract the features of intrusion detection system. Zhang et al. [34] presented machine leaning algorithms to classify the attacks using KDD99 dataset. [35] Proposed support vector machine detection method based on protocol grouping.

3. Methods and materials

In this section, presents the methodology of hybrid model for intrusion detection system.

3.1. 3.1 NSL-KDD dataset

The NSL-KDD is used to design proposed hybrid mode for intrusion detection. The NSL-KDD data set for intrusion detection and is an updated version of the KDD cup'99 data set due to the KDD cup data set has problem. The NSL-KDD data set is being built to solve McHugh's problem [24]. To run the experiments on the full set without choosing a small portion randomly. There are 4,898,431 entries in the NSL-KDD dataset. The data set of the NSL-KDD is collected as raw network packets and is independent of an operating system or application. Consequently, a tag indicating which category label the record belongs to has been given for each record in this data set. All the labels in this data set should be right. The NSL-KDD includes 37 types of attacks. The simulated attacks fell into one of the four categories in particular: Denial of Service, Probe, and user to Root and Remote to Local (U2R and R2L). In the NSL-KDD data set, the table 1 shows all types of attacks.

Table 1. All types of attacks in NSL-KDD

Attacks in Dataset	Type of attacks
Dos	Back, Land, Neptune, Pod, Smurf, Teardrop, Mailbomb, Processtable, Udpstorm, Apache2, Worm
Probe	Satan, IPsweep, Nmap, Portsweep, Mscan, Saint
R2L	Guess_password, Ftp_write, Imap, Phf,Multihop, WarezmasterXlock, Xsnoop, Snmpgue ss, Snmpgetattack, Httptunnel, Sendmail, Named
U2R	Buffer_ overflow, Loadmodule Rootkit, Perl, Sqlattack, Xterm, Ps

3.2. Preprocessing

Preprocessing is an important stage used to control data sets in a comprehensible format in the real world. Clearly, in specific behavior, the real-world datasets are incomplete and noisy. The preprocessing stage is very critical for analyzing big data for discovering the patterns. Pre-processing methods are necessary to improve the hybrid model for classification big data IDS. In addition to improving the reliability and performance of the machine learning algorithm for addressing the big data, the preprocessing steps is most significant for handling dimensionality reduction. In the present research study, the genetic method is used to obtain significant features from dataset. The detailed description of genetic algorithm method is presented in the subsequent subsections.

3.2.1. Genetic algorithm

Genetic algorithm is one of evolution computing which is a quickly developing in the area of artificial intelligence filed. Genetic algorithm is based on search algorithms which use nature selection and genetics. The

Genetic algorithm uses population chromosomes solution to solve the problem; each chromosome has fixed length [24]. The principal pupations of chromosomes are developed by distribution 1 and 0 arbitrarily, the distribution used average assignment of 0 and 1. In this encoding plan, each chromosome is a bit of strings (0s) whose length is computed by the quantity of key principal in the search space. The bit with 1 is chosen and bit with 0 isn't chosen in this encoding plan. Each chromosome demonstrates an applicant arrangement or a subset of main parts. The population develops by searching for obtain optimal features [25]. The Genetic algorithm has two noteworthy issues of nearby optima and being costly computationally. The flowchart of Genetic algorithm for feature section is appeared in Figure 1. The calculation connected is portrayed. Tables 2, 3 and 4 show the subset features of KDD cup dataset using Genetic algorithm. The Genetic algorithm is automatic selection features so, it is observed that the Genetic algorithm select more features. 23, 17 and 17 most significant features are selected with respective Probe, DOS, U2R and R2L attacks respectively using GA method from NSL-KDD dataset.

Step 1. Initial population creation (*n*chromosomes)
Step 2. Population evaluation (fitness evaluation of each chromosomes)
Step 3 if(Criterion is not met)
{
Selection;
Crossover;
Mutation;
}
Step 4. Return best individuals

Pseudo code of genetic algorithm

Table 2. Most significant features of DOS attack in NSL-KDD using GA method

Features numbers	Features Name
1	duration
2	protocol_type
3	service
4	flag
5	src_bytes
8	wrong_fragment
9	urgent
10	hot
11	num_failed_logins
12	logged_in
13	num_compromised
16	num_root
17	num_file_creations
21	is_host_login
22	_guest_login
23	count
24	srv_count
26	srv_serror_rate
27	rerror_rate
31	srv_diff_host_rate
34	dst_host_same_srv_rate
39	dst_host_srv_serror_rate
41	dst_host_srv_rerror_rate

Table 3. Most significant features of probe attack in NSL-KDD data set using GA method

Features numbers	Features Name
2	protocol_type
3	service
4	flag
5	dst_bytes
6	src_bytes
7	Land
10	Hot
11	num_failed_logins
14	root_shell
15	su_attempted
17	num_file_creations
19	num_access_files'
22	is_guest_login
23	Count
25	serror_rate
29	same_srv_rate
30	diff_srv_rate
31	srv_diff_host_rate
32	dst_host_count
33	dst_host_srv_count
34	dst_host_same_srv_rate
39	dst_host_rerror_rate
40	dst_host_srv_serror_rate

Table 4. Most significant features of U2R and R2L attack in NSL-KDD using GA method

Features numbers	Features Name
1	duration
3	service
5	src_bytes
10	hot
11	num_failed_logins
12	root_shell
13	dst_host_srv_diff_host_rate
14	num_shells
18	is_guest_login
22	rerror_rate
27	srv_rerror_rate
28	same_srv_rate
29	srv_diff_host_rate
31	dst_host_count
32	dst_host_srv_count
33	dst_host_same_srv_rate
34	dst_host_diff_srv_rate
35	dst_host_same_src_port_rate
36	num_compromised
37	dst_host_srv_serror_rate
39	

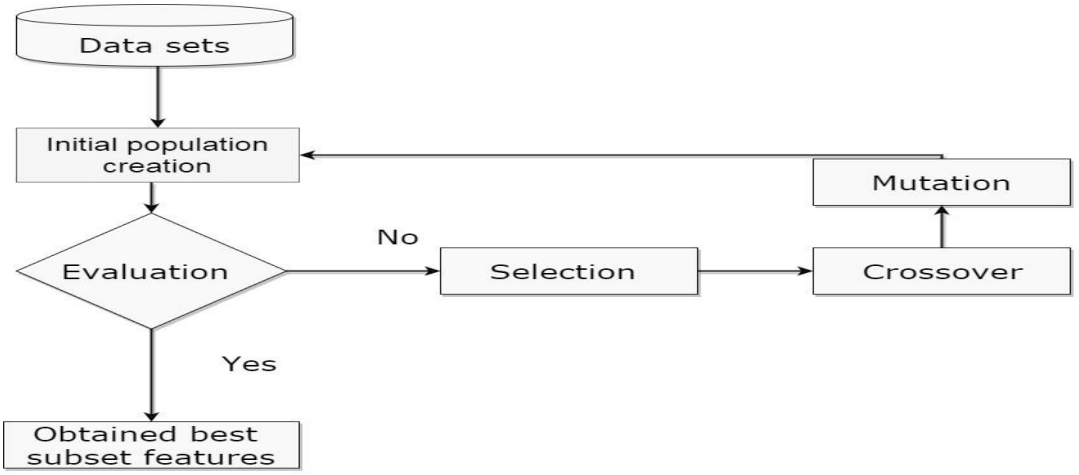

Figure 1. Flowchart of GA for feature selection

3.3 Support vector machine (SVM)

In 1963, Vapnik [26] suggested support for Vector Machine. It is an effective supervised machine learning algorithm used for classification large dataset and provides more detailed performance. The support vector machine was designed for regression and classification issues. In order to SVM use as binary classification with two classes or multiple classes for classification problem. The support vector machine is working to find the optimal hyper plane of the dichotomy that can help maximize the gap that can allow two or more classes the largest separation. Two parallel hyper planes are built to distinguish two classes, the help vector machine tries to figure out how to divide the hyper planes and optimize the distance between these two hyper planes. Hence, the longest length of the hyper plane is called good separation [34]. The SVM obtained lower error when the margin is large. The support vector machine obtained lower error when the margin is large.

Let $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \chi \times \{\pm 1\}$ is set of training dataset, where the χ denote to some nonempty set in the pattern x_1, and the function $f: \chi \to \{\pm 1\}$.

$$(w.x) + b - 0 \quad (1)$$

Where the $w \in \mathbb{R}^N, b \in \mathbb{R}$

In the following formula, show decision of SVM corresponding for obtaining the results.

$$f\ (x) - \text{sgn}\big((w.x) + b\big) \quad (2)$$

For saving the problem for optimization problem, it needs to construct an optimal hyper plane.

$$\text{minimize}_{w,b}\ \frac{1}{2}||w|^2 \quad (3)$$

Subject to

$$y_i\big(w^T.x_i + b\big) \geq 1 \text{ where } i = 1, \ldots., N \quad (4)$$

For reducing the misclassification that happened in Equation 3, putt the margin large as Equation 4.

$$\text{minimize}_{w,b}\ \frac{1}{2}||w|^2 + C\sum_{i=1}^{N}\xi_i \quad (5)$$

3.4 Performance measures

The performance indicators were carried out to evaluate the results of hybrid model. Five evaluation metrics namely accuracy, false positive, precision, true positive and time are used for testing the hybrid model. The equations performance indicators are as follow:

False positive rate (FPR) $= FP/(TN + Fp)\%\ 100$ (6)

True positive rate (TPR) $= TP/(TP + FN)\%100$ (7)

Accuracy $= (TP + TN)/(TP + FP + TN + FN)\%100$ (8)

Precision=$TP/(TP+FP)\%100$ (9)

True negative (TN): Correctly classified of valid records as normal record.
True positive (TP): Correctly classified of attack records as attacks.
False positive (FP): The percentage of incorrect records normal data as attacks.
False negative (FN): The percentage of incorrect records attacks as normal record.

4. Experimental analysis

The hybrid model is proposed to classification big data intrusion detection system. The hybrid model is implemented by using MATLAB R2018a-64 windows 10 Ultimate with the core i7 processor and 8 GB RAM. Five evaluation metrics are applied to examine the hybrid model. In this experiment hybrid model combines of genetic algorithm and support vector machine approach are applied. In this empirical results, 31 major attacks are selected. The size of data 29 MB data, 258961 records of attacks and normal class for all attacks.

Table 5. Results of hybrid model

	False Positive Rate	*True Positive Rate (TPR*	*Accuracy*	*Precision*	*Time (second)*
DOS	0.001	99.33	99.16	99.03	16.23
Probe	0.002	98.02	99	99.03	20.33
U2R and R2L	0.006	94.60	96.15	98.23	18.12

In order to improve analysis and classification intrusion detection system the hybrid model is suggested. The dimensionality reduction is one of biggest challenges in intrusion detection system, handling this dimensionality the genetic algorithm was applied to irreverent features from entire data set. The genetic method is proposed for improving the support vector machine classifier. The genetic algorithm assists to improve the accuracy of support vector machine classifier for building the hybrid model. Selection features are processed by support vector machine algorithm. We have most significant feature for each attack separately. Table 5 shows the empirical results of proposed model. The NSL-KDD dataset the genetic method is obtained subset features DOS= 23, Probe=17, U2R and R2L =17. The dataset has been divided into 70% training data and 30% testing data. From the results analysis, it is noted results of SVM with genetic algorithm are 0.001, 99.33%, 99.03% and 16.23 with respective to False Positive Rate, True Positive Rate (TPR), Accuracy, Precision and Time (second) respectively. Whereas the results of probe attack are False Positive Rate 0.002%, True Positive Rate (TPR) 98.02%, Accuracy 99%, Precision and Time (second) 20.33. The results of U2R and R2L 0.006, 94.60, 96.15, 98.23 and 18.12 according to False Positive Rate, True Positive Rate, Accuracy, Precision and Time (second) correspondingly. Figures 2,3 and 4 illustrate the performance of using NSL-KDD dataset. It is noted that the results of proposed model are satisfactory.

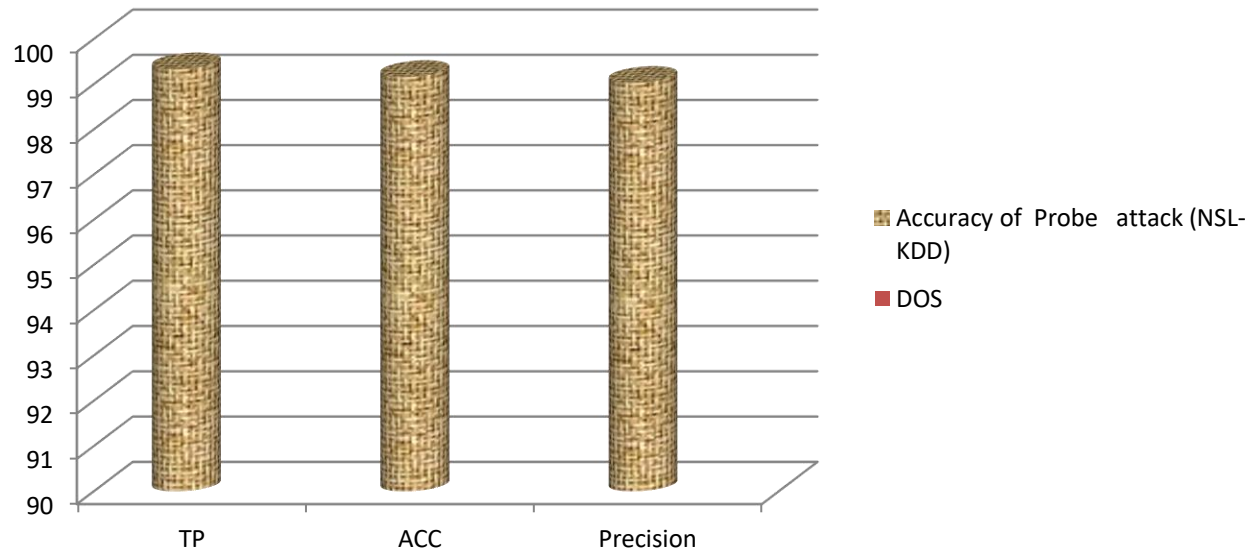

Figure 2. Displays performance hybrid model for DOS attack in NSL-KDD data set

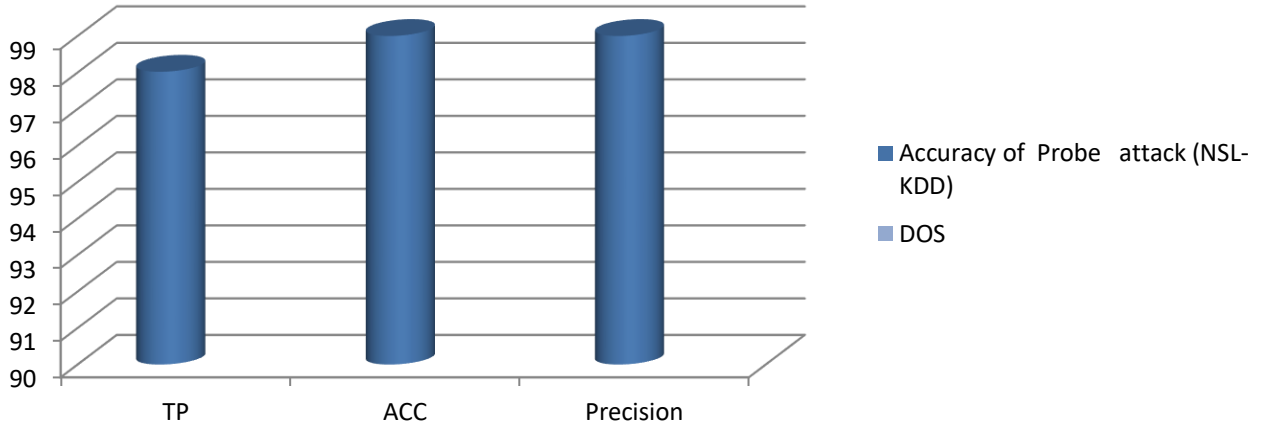

Figure 3. Displays performance hybrid model for Probe attack in NSL-KDD data set

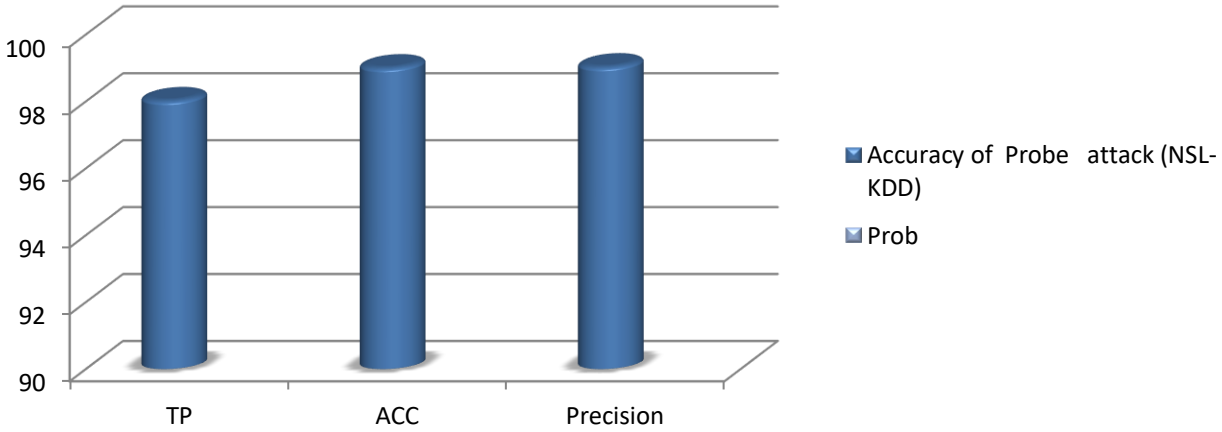

Figure 4. Displays performance hybrid model for U2R and R2L attack in NSL-KDD data set

5. Conclusion

The main object of proposed research is to improve classification big data intrusion detection system. The hybrid model combines of a genetic and SVM algorithms were implemented for classification intrusion from network traffic data. The genetic algorithm is applied to obtain most significant features from network data for handling dimensionality reduction. The genetic method is automatic selection features so, it is observed that the genetic method selects more features. 23, 17 and 17 most significant features are selected with respective Probe, DOS, U2R and R2L attacks respectively using genetic method from the original 41 features in NSL-KDD. The subsets features are processed by support vector machine algorithm, is investigated that the accuracy of proposed hybrid model is increased when genetic algorithm reduce the dimensionality reduction. Finally, it is concluded that the results of proposed model are more satisfactory. In future, the researcher will use deep learning algorithm with using more network datasets.

References

[1] Top-5-cybersecurity-concerns-for-2018. [online]. Available: *https://www.csoonline.com/article/3241766/ cyber-attacks-espionage/top-5-cybersecurity-concerns-for-2018.html.* [accessed: 23 June 2018].

[2] Cisco cybersecurity reports. [online]. Available: *https://www.cisco.com/c/en/us/products/security/security-reports.html#~{}stickynav=2*. [accessed: 9 August 2018].

[3] Santos, L.; Rabadao, C.; Gonçalves, R. "Intrusion detection systems in internet of things: A literature review". *In Proceedings of the IEEE 13th Iberian Conference on Information Systems and Technologies (CISTI)*, Caceres, Spain; pp. 1–7, 13–16 June 2018.

[4] Tsiropoulou, E.E.; Baras, J.S.; Papavassiliou, S.; Qu, G. "On the mitigation of interference imposed by intruders in passive RFID networks". In *International Conference on Decision and Game Theory for Security*; Springer: Cham, Switzerland, pp. 62–80, 2016.

[5] Hodo, E.; Bellekens, X.; Hamilton, A.; Tachtatzis, C.; Atkinson, R. "Shallow and deep networks intrusiondetection system: A taxonomy and survey". arXiv **2017**, arXiv:1701.02145.

[6] Axelsson, S. "Intrusion detection systems: A survey and taxonomy; technical report". In Chalmers University: Goteborg, Sweden, Volume 99, 2000.

[7] Kim, J.; Kim, H. "An e_ective intrusion detection classifier using long short-term memory with gradient descent optimization". In *Proceedings of the IIEEE international Conference on Platform Technology andService (PlatCon)*, Busan, Korea, 13–15 February, 2017.

[8] Hinton, G.E.; Osindero, S.; Teh, Y.W. "A fast learning algorithm for deep belief nets". In *Neural Compute.* vol. 18, pp. 1527–1554, 2006 [CrossRef]

[9] Wu, Z.; Wang, X.; Jiang, Y.G.; Ye, H.; Xue, X. "Modeling spatial-temporal clues in a hybrid deep learning framework for video classification". In *Proceedings of the 23rd ACM International Conference on Multimedia*, Brisbane, Australia, 26–30 October 2015; pp. 461–470.

[10] Tang, D.; Qin, B.; Liu, T. "Document modeling with gated recurrent neural network for sentiment classification". In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Lisbon, Portugal, 17–21 September 2015; pp. 1422–1432.

[11] George, G, Haas, M.R., Pentland, A. (2014). "Big Data and Management", In *Academy of Management Journal*, 57(2): pp. 321–326.

[12] Zhang, D. (2013). "Granularities and Inconsistencies in Big Data Analysis", In *International Journal of Software Engineering and Knowledge Engineering*, 23(6): pp. 887–893.

[13] Manandhar, P. (2014). "A Practical Approach to Anomalybased Intrusion Detection System by Outlier Mining in Network Traffic" (Doctoral dissertation), In *Masdar Institute of Science and Technology)*.

[14] Guillen, E., Sánchez, J., & Paez, R. (2015). "Inefficiency of ids static anomaly detectors in real-world networks". In *Future Internet*, 7(2), pp. 94-109.

[15] Yin, C.; Zhu, Y.; Fei, J.; He, X. "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks". In *IEEE Access 2017*, 5, 21954–21961. [CrossRef]

[16] Kuang, F.; Xu,W.; Zhang, S. "A novel hybrid KPCA and SVM with GA model for intrusion detection". In *Appl. Soft Comput.* 2014, 18, pp. 178–184. [CrossRef]

[17] Reddy, R.R.; Ramadevi, Y.; Sunitha, K.V.N. "E_ective discriminant function for intrusion detection using SVM". In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Jaipur, India, 21–24 Septembert 2016; pp. 1148–1153.

[18] Li, W.; Yi, P.; Wu, Y.; Pan, L.; Li, J. "A new intrusion detection system based on KNN classification algorithm in wireless sensor network". In *J. Electr. Comput. Eng.* 2014, 2014, 240217. [CrossRef]

[19] Farnaaz, N.; Jabbar, M.A. "Random forest modeling for network intrusion detection system". In *Procedia Comput. Sci. 2016*, 89, pp. 213–217. [CrossRef]

[20] Zhang, J.; Zulkernine, M.; Haque, A. "Random-forests-based network intrusion detection systems". In *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 2008*, 38, pp. 649–659. [CrossRef]

[21] Kim, G.; Lee, S.; Kim, S. "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection". In *expert syst. appl. 2014*, 41, pp. 1690–1700. [CrossRef]

[22] Panda, M.; Patra, M.R. "Network intrusion detection using naive bays". In *int. J. comput. sci. netw. secur.* 2007, 7, pp. 258–263.

[23] M Hassan Zaib. "NSL-KDD dataset". [online]. Available: https://www.kaggle.com/hassan06/nslkdd. [accessed: 15 March 2020]

[24] Baranidharan, T. and D.K. Ghosh, "edical image classification using genetic optimized elman network". In *Am. J. applied sci.*, 9: pp. 123-126. 2012.

[25] Guo, Y., B. Wang, X. Zhao, X. Xie and L. Lin et al, "Feature selection based on Rough set and modified genetic algorithm for intrusion detection". In *Proceedings of the IEEE 5th International Conference on Computer Science and Education*, pp: 1441- 1446,2010

[26] Cortes C, "Vapnik VN Support vector networks". In *Machine Learning 2*, pp.:273–297, (1995)

[27] T. H. Hadi and M. R. Joshi, "Handling ambiguous packets in intrusion detection". In *3rd International Conference on Signal Processing, Communication and Networking (ICSCN)*, March 2015, pp. 1–7

[28] Aldhyani H. T and M. R. Joshi, "Analysis of dimensionality reduction in intrusion detection". In *International Journal of Computational Intelligence and Informatics*, Vol. 4: No. 3, 2014, pp. 199-206

[29] Doshi, R.; Apthorpe, N.; Feamster, N. "Machine learning ddos detection for consumer internet of things devices". In *Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 24 May 2018; pp. 29–35.

[30] Meidan, Y.; Bohadana, M.; Mathov, Y.; Mirsky, Y.; Shabtai, A.; Breitenbacher, D.; Elovici, Y. "N-BaIoT— Network-based detection of IoT botnet attacks using deep autoencoders". In *IEEE Pervasive Comput.* 2018, 17, 12–22. [CrossRef].

.[31] Diro, A.; Chilamkurti, N. "Leveraging LSTM networks for attack detection in fog-to-things communications". In *IEEE Commun. Mag.* 2018, 56, 124–130. [CrossRef].

[32] Potluri, S.; Ahmed, S.; Diedrich, C. "Convolutional neural networks for multi-class intrusion detection system". In *Mining Intelligence and Knowledge Exploration*; Springer: Cham, Switzerland, 2018; pp. 225–238.

[33] Zhang, B.; Yu, Y.; Li, J. "Network intrusion detection based on stacked sparse autoencoder and binary tree ensemble method". In *Proceedings of the 2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.

[34] Zhang, H.; Yu, X.; Ren, P.; Luo, C.; Min, G. "Deep Adversarial Learning in Intrusion Detection: A Data Augmentation Enhanced Framework". In arXiv 2019, arXiv:1901.07949.

[35] Teng, S.; Wu, N.; Zhu, H.; Teng, L.; Zhang, W. "SVM-DT-based adaptive and collaborative intrusion detection". In *IEEE/CAA J. Autom. Sin.* 2017, 5, 108–118. [CrossRef].