IJHRMOB

International Journal of HRM and Organizational Behavior

www.ijhrmob.com

editor@ijhrmob.com

International Journal of
HRM and Organizational Behavior
ISSN 2454 - 5015

NETWORK INTRUSION DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUE WITH FEATURE SELECTION

[1]Mr.T.SATHISH,[2]VARUN KUMAR,[3]R BHARGAVAN,[4]SYED JAWED ALI JAFFER,[5]SRIRAM HARSHIT

[1]Assistant Professor,Department Of CSE,Malla Reddy Institute Of Engineering And Technology(autonomous),Dhulapally,Secundrabad, Telangana, India,sathish46nkl@mriet.ac.in

[2,3,4,5]UG Students, Department Of CSE,Malla Reddy Institute Of Engineering And Technology(autonomous),Dhulapally,Secundrabad, Telangana, India.

ABSTRACT

Two machine learning techniques, namely SVM (Assist Vector Machine) and ANN (Artificial Neural Networks), are evaluated in this study for their efficiency. The presence or absence of regular or anomalous signatures in the request data will be determined using machine learning techniques. Internet intrusion detection systems (IDS) monitor request data and check if it contains typical or assault signatures; if it does, demand is reduced. This is necessary because nowadays all services are accessible online and malicious individuals can attack client or web server machines through this net. When new request signatures come, the IDS will be educated with all possible strikes signatures using AI techniques and then used to determine whether the new request comprises regular or assault trademarks. Here, we compare and contrast two AI algorithms, Support Vector Machines (SVM) and Artificial Neural Networks (ANN), and we find, experimentally, that ANN is more accurate than the current state-of-the-art SVM. Examining how well SVM and ANN work is the focus of this article. By utilising Relationship Based and Chi-Square Based function option formulas, the author has reduced the dataset dimension, eliminated irrelevant data, and loaded the model with important attributes. As a result of these features choice formulas, the dataset dimension will decrease and the forecast accuracy will increase.

Keywords: ANN, SVM, IDS, Attacks, UTM, IPS.

I.INTRODUCTION

The prevalence of cybercrime is directly proportional to the exponential growth in internet use and the ease with which people may access various online resources. first two The first line of defence against a safety strike is intrusion detection. Research investigations are therefore focusing heavily on safety and security services such Intrusion Avoidance Systems (IPS), Unified Hazard Modelling (UTM), Firewall, and Invasion Discovery System (IDS). By gathering data and analysing it for potential security breaches, intrusion detection systems (IDS) may identify attacks from a variety of systems and networks [3]. There are two ways that network-based intrusion detection systems (IDS) evaluate data packets as they travel across a network. Anomaly based detection is a substantial field of study because, even now, it lags significantly behind signature based detection [4-5]. Among the difficulties of anomaly-based invasion detection is the fact that it must deal with novel attacks for which there is no precedent in order to identify the abnormality. As a result, the system needs some kind of intelligence to distinguish between benign and malicious communication, and researchers have been uncovering artificial intelligence ways to do just that in recent years [6]. Still, intrusion detection systems aren't a panacea for all problems with security. For instance, in the event that there is a hole in the network's procedures or a poor identity and authentication system, IDS will not be able to patch it.

Invasion discovery research started in 1980, and the first version was published in 1987 [7]. Despite heavy investment from businesses and academics over the last several decades, intrusion detection innovation remains in its infancy and has not yielded satisfactory results [7]. Commercial success and widespread acceptance by innovation-based companies throughout the globe have been seen by signature-based network intrusion detection systems, but anomaly-based systems have not achieved the same level of success. For this reason, anomaly based discovery is now a hotspot for research and development in the area of intrusion detection systems [8]. Important issues still need to be addressed before a widespread deployment of anomaly based intrusion detection systems can be considered [8]. When it comes to

comparing the efficacy of breach detection using supervised equipment learning approaches, however, there is a dearth of current work [9]. An improvement, anomaly-based network intrusion detection systems safeguard specific networks and systems against malicious actions. Anomaly detection capabilities have enabled safety and security tools to emerge, but some important issues have not yet been resolved, despite the selection of anomaly-based network invasion discovery methods described in recent literature [8]. Linear regression, support vector machines (SVMs), genetic algorithms, k-nearest neighbour formulas, ignorant bayes classifiers, decision trees, and gaussian mix models are just a few of the anomaly-based approaches proposed [3,5]. Since it has already shown itself on several kinds of problems, support vector machines (SVMs) are among the most used learning formulas [10]. A major problem with anomaly based discovery is that, although all of the above tactics may find new attacks, they all have a high dud rate. This is due to the fact that it is not easy to derive explanations for reasonable, typical behaviour from training data sets [11]. Back propagation has been known since 1970 as the inverse setting of automated differentiation [12], and it is still often used today to train Artificial Semantic Networks (ANNs). Without a complete collection of network-based data, it is very difficult to evaluate the efficacy of network intrusion detection systems [13]. The bulk of the abnormality-based techniques proposed in the literature were tested using the KDD MUG 99 dataset [14]. This study used two machine learning strategies—support vector machines (SVMs) and artificial neural networks (ANNs)—to the well-known benchmark dataset for network intrusion, NSLKDD [15].

II RESEARCH STUDY

A macro-social exploratory study of the cyber-victimization rate across states [1]. This study looks at the relationship between cyber-theft victimisation and signs of macro-level possibilities. In line with the arguments put forward by criminal chance theory, patterns of net accessibility at the state level are used to quantify direct exposure to run the risk. To find out if cyber-victimization varied between states due to differences in social structure, we looked at a number of other structural characteristics of states. Where people get their internet connection is correlated with structural

factors like unemployment and the percentage of the population that lives outside of major cities, according to the current research. Additionally, this study found a positive correlation between the proportion of people who solely use their home internet connection and cyber-theft victimisation rates at the state level. Theoretical considerations about these results are addressed.

[2] Use of limited recognised information in a step-by-step anomaly-based breach detection system, As the internet continues to grow and more people across the globe have access to online media, the prevalence of cybercrime is also on the rise. Cybercriminals target both people and businesses nowadays. You may protect yourself with a variety of tools, such as firewall programmes and Invasion Discovery Systems (IDS). In order to prevent packages from passing through unchecked, firewall software acts as a checkpoint. It may even partition the whole network's web traffic in the worst-case scenario. In contrast, intrusion detection systems automate network monitoring. Building intrusion detection systems is quite challenging due to the streaming nature of data in computer networks. Online dataset categorization is suggested as a solution to this issue in this research. This is achieved by using a naïve Bayesian classifier that is trained step-by-step. Moreover, active discovery enables issue solving with a small amount of identifiable data points, which are sometimes quite expensive to gather. Two groups, one dealing with offline activities and the other with online ones, make up the suggested approach. The first one describes data preparation while the final one shows the NADAL online method. We evaluate the suggested method against the NSL-KDD typical dataset-using step-by-step naive Bayesian classifier. The suggested method has three benefits over the step-by-step ignorant Bayesian approach:(1) it overcomes the streaming data challenge;(2) it reduces the high expenditure of instance labelling; and(3) it boosts accuracy and Kappa. Therefore, the method works well for intrusion detection systems.

the third A method for assessing the security breach detection system via modelling and application, The goal of intrusion detection systems (IDSs) is to identify strikes either in progress or after they have already occurred. Research aimed to do two things: first, examine the system IDS; and second, reduce the impact of attacks. Undoubtedly,

intrusion detection systems (IDSs) collect data on website traffic from many network or computer system resources and use it to make systems safer. On the other side, evaluating IDS is a vital task. Considering the components of a system is different from studying the system as a whole in terms of performance. We provide an approach to IDS evaluation in this research that relies on element efficiency determination. To begin, we have proposed an embedded systems-based equipment system to ensure the safe implementation of the IDS SNORT components. Next, we ran it through a test that mimics real-world website traffic and attacks using the Metasploit 3 Framework and Linux KALI (Backtrack). The obtained data demonstrates that the characteristics of these components have a significant impact on the IDS efficiency.

III.EXISTING SYSTEM

The lack of a comprehensive network-based data collection is one of the main obstacles to assessing the efficacy of network intrusion detection systems [3]. The majority of the abnormality-based techniques proposed in the literature were tested on the KDD CUP 99 dataset. Here, we use two machine learning techniques—support vector machines (SVMs) and artificial neural networks (ANNs)—to the well-known standard dataset for network invasion, NSLKDD. Interesting are the claims made and the contributions AI has made up to this point. Today, machine learning has presented us with a plethora of reality applications. It would seem that AI will definitely become global ruler in the near future. So, we tested the notion that machine learning techniques may overcome the challenge of discovering new attacks, sometimes known as zero-day attacks, that contemporary technology-enabled enterprises face. Based on the information gathered from the observed traffic, we were able to develop a version of the monitored maker that can detect undiscovered network website traffic. We used the SVM and ANN learning algorithms together to get the best classifier in terms of accuracy and success rate.

IV.PROPOSED SYSTEM

As shown in Figure 1, the suggested system is composed of an algorithm for detecting attributes and an algorithm for selecting them. In order to assign a given set of circumstances to a certain class, attribute selection elements are responsible for extracting the most

relevant functions or attributes. Using the results found in the function selection component, the finding formula component builds the necessary knowledge or expertise. The design learns and becomes qualified with the help of the training dataset. After that, the testing dataset is used to evaluate how well the intelligences were able to classify unknown data. Unattended device learning for anomaly detection in network website traffic Two supervised device-finding techniques, namely Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs), are evaluated in this study. The presence of attack (abnormality) fingerprints in demand data may be detected using artificial intelligence algorithms. IDS (Network Invasion Detection System) is used to prevent malicious users from launching attacks on customer or web server devices. The system monitors request information and checks for attack signatures; if it finds any, the request is rejected. Nowadays, everything can be found on the internet. The IDS will be trained with all potential attack signatures using AI formulae, and then a train version will be created. When new demand trademarks appear, this design will be used to determine whether the new demand has typical or attack signatures. The research compares the effectiveness of support vector machines (SVMs) and artificial neural networks (ANNs), and the results show that the latter is more accurate than the former. In order to prevent all attacks, intrusion detection systems have developed procedures to examine each incoming request for signs of attacks. If the request appears to be from legitimate users, the request will be forwarded to the web server for processing. However, if the request contains attack trademarks, the request will be rejected and the information will be recorded in the dataset for future use in discovery. For intrusion detection systems (IDS) to detect these types of assaults, they must first be trained with all possible attack trademarks originating from malevolent individuals' requests. Only then can they create a training model. In order to determine whether a newly received request is part of the normal class or the strike course, IDS will apply the request to the specific train model. In order to train these designs and make predictions, a plethora of information mining prediction algorithms will be used. Reviewing the effectiveness of SVM and ANN is the focus of this research. To reduce the size of the dataset and improve the accuracy

ISSN 2454 - 5015

of the predictions, the author has used Relationship Based and Chi-Square Based function selection formulas in this algorithm. The feature option algorithms removed unnecessary data from the dataset and replaced it with design with vital functions.

V.WORKING METHODOLOGY

The creator of the habits test has made use of the NSL KDD Dataset; here are some sample documents containing request trademarks from that dataset. You can find the dataset I used, which is located in the 'dataset' folder, at the same location.

Here are some examples of variables found in datasets: size, protocol_type, service, flag, src_bytes, dst_bytes, land, incorrect_fragment, pushing, hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_root, num_file_creations, num_shells, quantity of access files, number of outbound commands, is_host_login, is_guest_login, and matter, servers, error rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate, dst_host_srv_rerror_rate, label.
Names of request trademarks are all capitalised and presented in a bright way.

no, tcp, ftp_data, SF,491, zero,0,0,0, zero,0,0,0, absolutely no, no, absolutely no, no, zero,0, absolutely no, zero, absolutely no,2,2, zero,0, zero,0,1,0,0, a hundred and fifty,25, normal,.17,0.03,0.17,0,0,0.05, normal, no.

0, tcp, private, S0,0,0, no, no, no,0,0, absolutely no, absolutely no, no,0,0, no, absolutely no, zero,0, absolutely no, absolutely no,Anamoly, 166,9,1,1, zero, zero.05, zero.06, no, 255, 9,0.04, zero.05, zero,0,1,1,0,0, zero.

The values of the signatures are the facts mentioned above, and the class tag that makes up the staying charge is either a standard request signature or an assault trademark. 'Neptune' is an attack name in the second file. In the same vein, the collection contains over 30 notable assault names.

ISSN 2454 - 5015

Some variables in the dataset documents are still in string format; for example, tcp and ftp_data. These values aren't important for the prediction and may be removed using the PREPROCESSING Idea. Because our algorithm would fail miserably if fed attack names in text format, we want to instead provide each attack a number value. A new document called "easy.Txt" will be created for the purpose of creating the education version when all of this is carried out in the PREPROCESS activities.

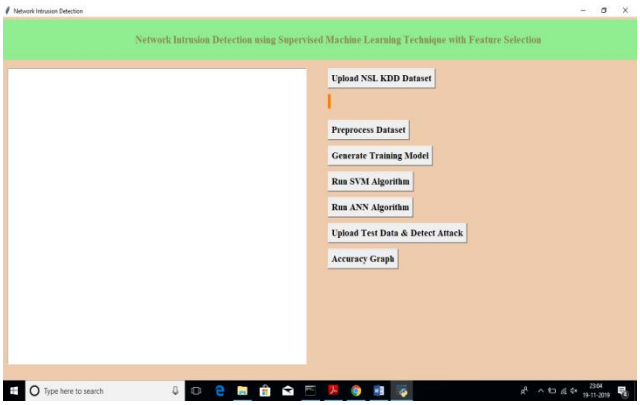
I am assigning number identifiers for each attack in the line below.

In the aforementioned traces, we can see that "day-to-day" has the id "absolutely no" and "anamoly" has the id "1" and exists for all attacks.

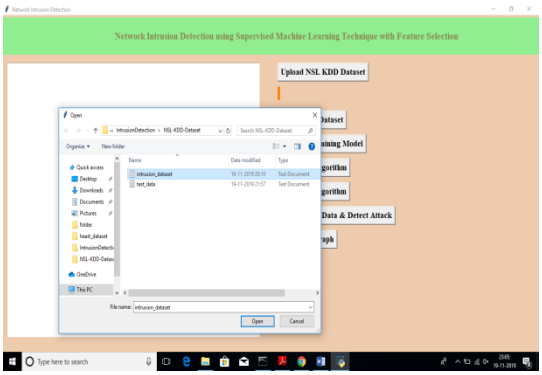
The two instructions below must be executed before any code.

VI.OUTPUT EXPLANATION

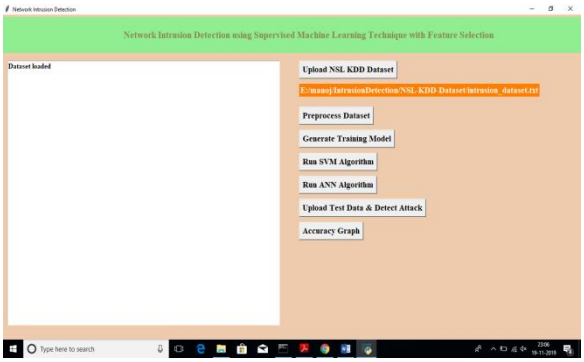
Double click on ‘run.bat’ file to get below screen



Press the "Upload NSL KDD Dataset" button on the previous screen to add the dataset.



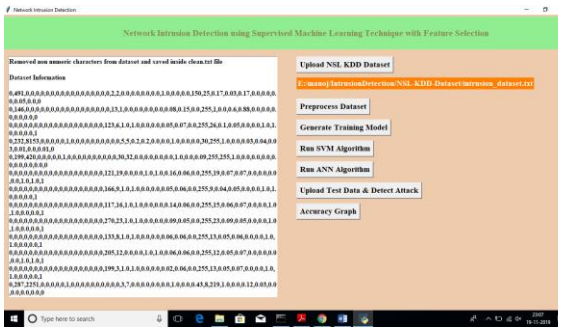
After importing the dataset, the following screen appears underneath the one I was using to submit the "intrusion_dataset.Txt" report:



To clean the dataset by converting attack names to numbers and eliminating string

ISSN 2454 - 5015

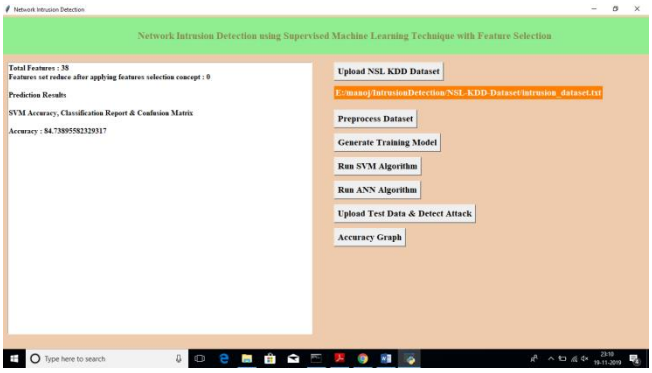
values, choose the "Pre-process Dataset" button.



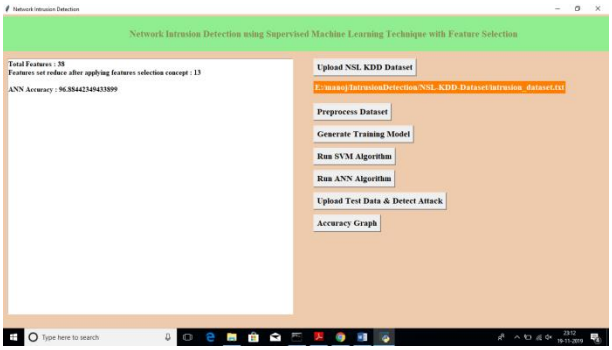
The next step is to pre-process the data by removing any text values and then converting the names of the attacks to numbers. The normal signature has an identifier of zero and the anomalous attack has an identifier of one.

The next step is to use the 'Generate Training Model' button to divide the data into train and examine sets. This will create a version for making predictions using SVM and ANN.

There are a complete of 1244 entries in the dataset, with 995 getting used for schooling and 249 for checking out, as seen in the above display screen. The next step is to construct an SVM model and determine its accuracy through clicking the "Run SVM Algorithm" button.

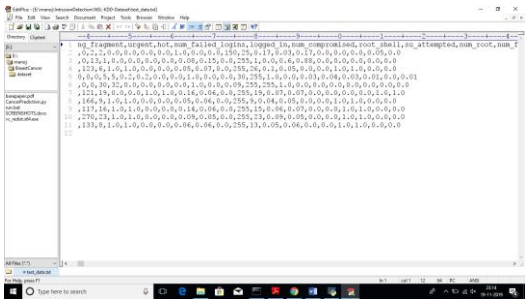


To find the ANN accuracy, click on "Run ANN Algorithm," as shown above; using SVM, we got an accuracy of 84.73%.

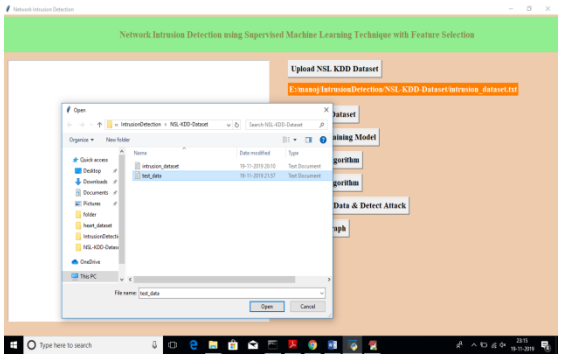


After achieving an accuracy rate of 96.88% in the previous display, we can proceed to submit test data and determine whether it is normal or has a strike by clicking the "Upload Test Data & Detect Strike" button. The programme will undoubtedly make a prediction and provide us with the results, and all of the test data is numeric. See a few documents derived from test data below.

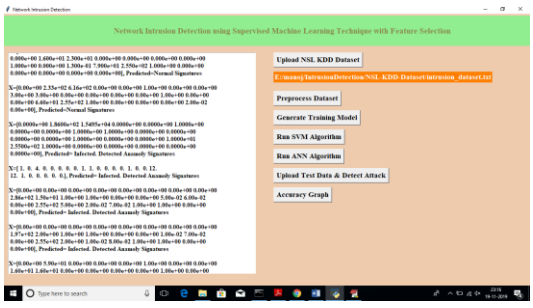
ISSN 2454 - 5015



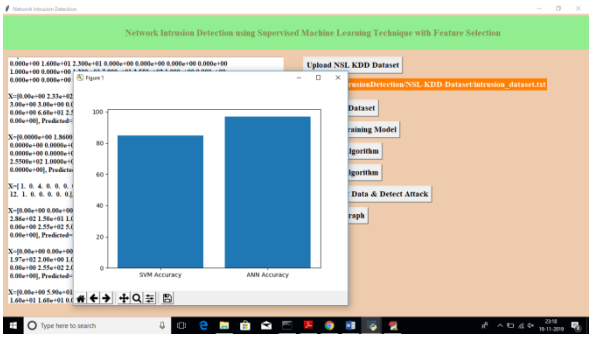
The programme will identify and provide us with results even when the test data above does not include either a 0 or a 1.



I am uploading the 'test_data' file, which includes the test records, to the screen above. After the prediction, I will receive the results below.



We expected each test document to show up in the above display as either "Regular Trademarks" or "contaminated," depending on the information entered. To see a graph comparing the accuracy of SVM and ANN, click the "Precision Chart" button.



The graph above shows that ANN outperformed SVM in terms of accuracy; the x-axis indicates the names of the methods, and the y-axis suggests how properly they achieved.

VII.CONCLUSION

In order to determine the best model, we have provided a number of machine learning options that make use of different maker learning algorithms and function option strategies. Based on the results, the design that used ANN and wrapper feature selection was the most effective in correctly classifying network website traffic, with a detection rate of 94.02%. We anticipate that further research into the topic of building a detection system capable of detecting both known and new attacks will be prompted by these results. At this time, intrusion detection systems

can only identify assaults that have already been discovered. Due to the high false positive rate of current systems, finding new attacks, often known as zero day attacks, is an active area of study.

VIII.REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016

. [10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning

support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.

[11] M. Tavallaee, N. Stakhanova and A. A. Ghorbani, "Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516-524, Sept. 2010.

[12] F. Gharibian and A. A. Ghorbani, "Comparative Study of Supervised Machine Learning Techniques for Intrusion Detection," Fifth Annual Conference on Communication Networks and Services Research (CNSR '07), Fredericton, NB, Canada, 2007, pp. 350-358.

[13] N. Moustafa and J. Slay, "UNSWNB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, 2015, pp. 1-6.

[14] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 2017, pp. 1881-1886.

[15] M. Panda, A. Abraham and M. R. Patra, "Discriminative multinomial Naïve Bayes for network intrusion detection," 2010 Sixth International Conference on Information Assurance and Security, Atlanta, GA, USA, 2010, pp. 5-10.