Phoenix: International Multidisciplinary Research Journal
Vol 1, No. 2, April-June, 2023
ISSN:2583-6897

INTRUSION DETECTION OF IMBALANCED NETWORK TRAFFIC BASED ON MACHINE

[1] **Ankita,** Master of Computer Application BKIT-Bhalki
[2]**Prof. Poojarani,** Master of Computer Application BKIT-Bhalki

Abstract -The security of today's wireless networks is under constant danger from a growing number of faults, vulnerabilities, and assaults due to the fast increase of these types of problems. Network security is becoming an increasingly essential topic as a direct result of the fact that computer networks and applications are constructed on different platforms. Operating systems that are both complicated and costly are more likely to have flaws in their security. efforts to breach security, completeness, or availability are referred to as "intrusions," and the word "intrusion" describes such efforts. An intrusion detection system (IDS) is useful for detecting flaws in network security as well as unusual activity. Despite the fact that it is sometimes seen as premature and not as an ultimately complete technique of combating intrusions, the development of technology for detecting intrusions has been a sector that has had significant growth in recent years. The completion of this assignment has been elevated to the status of a top priority by both security professionals and network administrators.

This indicates that even more secure systems are unable to totally take its place. IDS is able to anticipate future intrusions based on intrusions that have already been identified since it uses data mining to identify intrusions. In this study, an exhaustive literature evaluation on the use of data mining techniques for IDS is offered. In the first step of this process, we will examine data mining techniques for the purpose of identifying intrusions using real-time and benchmark information. This article provides a comparison of several approaches of detecting intrusions in a network, along with an analysis of the benefits and drawbacks of each strategy. Within the scope of this research, we suggest many methods that might enhance network intrusion detection.

***Key Words*:** Block chain, Security, block size, hash code

1.INTRODUCTION

The significance of network security is rapidly growing as a direct result of the exponential rise in the number of applications and businesses that make use of computer networks. The majority of businesses defend themselves against threats to their networks by using network security technologies such as anti-virus and anti-spam software. These methods are unable to identify sophisticated or newly developed assaults; nevertheless, an intrusion detection system (IDS) [1] allows computer networks and computers to

for the purpose of identifying and removing unwanted intrusions. Identifier Information may be gathered and processed from a variety of sources by systems. sources inside a network or computer, recognizing hazards that might render individuals susceptible, such as misuse and infiltration. sources within a network or computer. IDSs, also known as Intrusion Detection Systems [2,] are computer programs that keep a close eye on all of the activity that takes place

on a network and analyze it in order to identify any potentially harmful behavior. Intruder detection systems (IDS) are now generally recognized as an essential component of the security architecture in most businesses. Companies are able to prevent assaults on their networks if they notice and investigate network breaches. This strategy might be used by security experts to cut down on the complexity of existing threats and lessen the risks currently associated with network security.

2. Literature survey:

1.A deep auto-encoder based approach for intrusion detection system

AUTHORS: F. Farahnakian and J. Heikkonen

Identification of network assaults is one of the most difficult challenges that network operators face in the modern day. This is due to the large number of vulnerabilities that exist in computer systems as well as the inventiveness of those who launch attacks. We describe a deep learning strategy for intrusion detection systems as a solution to this challenge. The Deep Auto-Encoder (DAE), which is one of the most well-known models for deep learning, is used by our method. In order to prevent overfitting and reach a global optimal solution, the proposed DAE model is trained using a greedy layer-wise approach. The results of our experiments on the KDD-CUP'99 dataset indicate that our method offers a significant improvement over previous deep learning-based methods in terms of accuracy, detection rate, and false alarm rate. These findings were obtained through our experiments.

2) "An intrusion detection model based on deep belief networks

AUTHORS: N. Gao, L. Gao, Q. Gao, and H. Wang, In this study, we concentrate on a significant research challenge involving the categorization of big data in intrusion detection systems. The concept of Deep Belief Networks is presented to those working in the area of intrusion detection, and a model of intrusion detection that is based on Deep Belief Networks is offered for use in the intrusion recognition domain. The deep hierarchical model is a deep neural network classifier that is comprised of a mixture of multilayer unsupervised learning networks, one of which is known as a Restricted Boltzmann Machine, and a supervised learning network, one of which is known as a Back-propagation network. This network is used to train in an unsupervised manner. The experimental findings on the KDD CUP 1999 dataset show that the performance of the Deep Belief Networks model is superior to that of SVM and ANN. This conclusion was reached as a consequence of comparing these three models.

3)Improvement of Network Intrusion Detection Accuracy by Using Restricted Boltzmann Machine

AUTHORS: KanS. Seo, S. Park, and J. Kim

A significant amount of noise and data that are considered to be outliers are mixed in with the different types of network intrusion detection data that are utilized for the learning of classification algorithms. Any high-performance network intrusion detection model becomes difficult to foresee in the event that learning is carried out using data that has a high level of impurities. This is true even if the performance of the classification method is exceptional. Not only should the performance of the classification algorithm be improved, but also the management of noises and outliers in the data that is used for the learning of the classification algorithm. This will help to improve the accuracy of the network intrusion detection process. RBM stands for restricted Boltzmann machine and is a kind of unsupervised learning that does not use the usage of class labels. RBM is a probabilistic generative model that creates new data based on the training probability by composing the new data based on the original input data. The new data that were generated using RBM demonstrate that the noises and outliers that were present in the input data have been eliminated. The negative impacts that the noise and outlier data had on the learning process are removed when the freshly constructed data are applied to the network intrusion detection model. In this investigation, sounds and outliers in the KDD Cup 1999 data were examined. The application of RBM on the data and the creation of new data are both required to delete the data. After that, compare the findings of the current data with the data after noises and outliers have been eliminated from it. In conclusion, this research shows that using RBM to remove noises and outliers from the data leads to an increase in the effectiveness of network intrusion detection.

4)Toward an online anomaly intrusion detection system based on deep learning

AUTHORS: K. Alrawashdeh and C. Purdy

Over the course of the last two decades, there has been consistent but incremental development in the area of intrusion detection. The most difficult task is to identify new threats as they occur in real time. via a Restricted Boltzmann Machine (RBM) and a deep belief network, a method to deep learning is constructed in this study for the purpose of anomaly detection via deep learning. In order to accomplish unsupervised feature reduction, our approach utilizes an RBM with one hidden layer. The resulting weights from this RBM are then input into another RBM, which ultimately results in a complex belief network. The weights that have been pre-trained are then sent into a fine tuning layer that is comprised of a Logistic Regression (LR) classifier with multi-class soft-max. We have used the DARPA KDDCUP'99 dataset in order to test the effectiveness of the deep learning architecture that we have developed and implemented in C++ inside Microsoft Visual Studio 2013. Previous deep learning algorithms developed by Li and Salama are outperformed by our architecture, which excels in both the speed and accuracy of detection. On the whole 10% KDDCUP'99 test dataset, we are able to obtain a detection rate of 97.9%. We have been able to reduce the number of false negatives to 2.47% because to the enhancements we have made to the training phase of the simulation. Despite the fact that the flaws in the KDDCUP'99 dataset are widely known and documented, it nonetheless provides methods of machine learning with a challenging obstacle to overcome when attempting to forecast assaults. In the work that we have planned for the future, we will be applying our

machine learning method to datasets that are bigger, more difficult, and include a wider variety of types of assaults.

3. OBJECTIVE:

An analysis and comparison of the benefits and drawbacks of the various methods for detecting intrusions that were covered in the part before this one. As seen in Table 1, each strategy offers perks and drawbacks, depending on how you look at it. The tabular data makes it easy to decide which technique is the most effective and which offers the greatest number of benefits. We can identify how the approaches that are currently being used are defective using this observation table, and we can come up with fresh solutions to fix these problems.
This observation table offers some insight into the investigation of IDS by using several machine learning approaches. Based on fuzzy entropy, which provides an accuracy of 99.5% after a period of time allowing ACO to converge. However, the identification of functions in SVM may be challenging. Chi-square feature selection approaches have an accuracy of 95.8%. The genetic algorithm was unsuccessful in its attempt to find the ideal subset 99.9% of the time.

4. SYSTEM ANALYSIS:

Existing System:

a method for the detection of intrusion that takes into account a number of challenges, including the vastness of the network traffic dataset, feature selection, a low rate of accuracy, and a large rate of false alarms [2].

[5] In order to analyze the network traffic information and identify intrusions, an Online Sequential Extreme Learning Machine (OS-ELM) is used. It is a single hidden layer feed forward neural network (SHLFN) that is both quick and accurate, and it can handle network instances either one at a time or in chunks. Through its performance in a single iteration, it has shown that it is applicable in categorization.

Disadvantages in Existing System:

1. The feature selection method is not good, irrelevant and redundant features are present.

2. The classifier does not work well for limited training data set

Proposed System

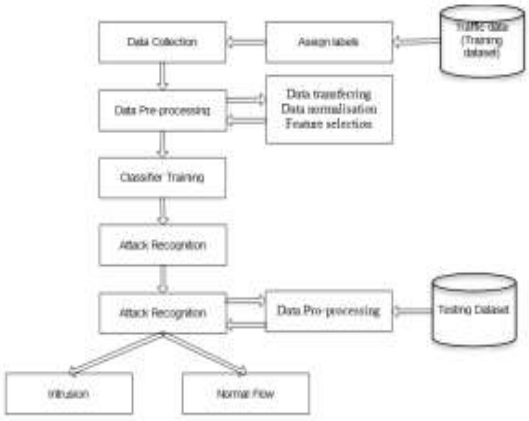
We provide a solution that addresses both of the issues.
1. Feature selection using symmetric uncertainty as the criterion.
Instead of using the CFSmethod that was discussed in the base work, we suggest using a novel formula for feature selection that is based on symmetric uncertainty. The symmetric uncertainty is a better indication for the link between the variables that were used to create the model and

the classification result that was produced (intrusion or no intrusion). Consequently, by employing this strategy, we are able to get the greatest characteristics, all of which are relevant and do not include repetitive information.

2. a grouping device.

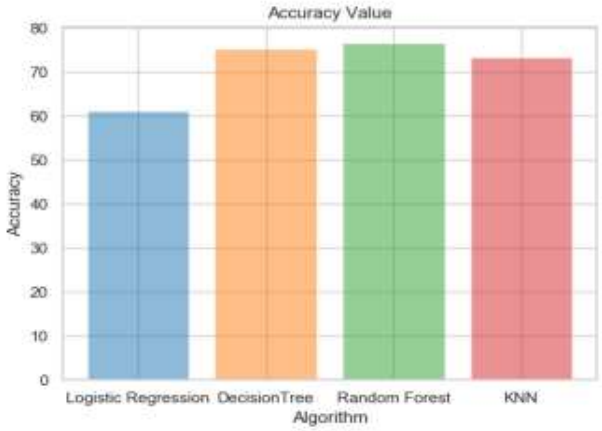
We employ a technique called data expansion so that we can improve the accuracy of the classifier even if we only have a restricted sample set. We first learn MLR, which stands for multivariate linear regression, based on the input dataset, and then we utilize the MLR model that we've learned to create further datasets and train the classifier. Classifiers may be taught effectively using this method, which also helps enhance accuracy**.**

4. ARCHITECTURE



6. Results and Analysis:

The greater incidence of false alarms that is present in contemporary IDSs is the primary drawback of these systems. We need to reduce the number of false alerts and false positives that are generated by a security system in order to improve its overall dependability. Because of the four different machine learning algorithms that have been included into the system, the one that has been suggested may assist in the detection of the genuine assaults. The detection rate of the incursion is high because even if a specific attack is successful in avoiding detection by one algorithm, it will still be discovered by one of the other algorithms; hence, the detection rate of the assault is high. The Vulnerability Assessment is carried out with the assistance of Snort by using predefined rule-sets, in accordance with which an action is carried out for an intrusion that has been recognized. Through conducting this comparative research, we are able to have a better understanding of which method is more effectively detecting assaults. The use of machine learning algorithms enables us to differentiate between genuine threats and erroneous alerts, therefore assisting us in the development of effective intrusion detection systems that are able to reduce the number of false positives, as well as to guarantee both dependability and safety.



Conclusion:

In this paper, we present a comprehensive review of data mining methodologies that are mostly network-based and are primarily focused on IDS. In order to provide future choices for increasing intrusion detection performance and, by extension, IDS, both the benefits and drawbacks of these tactics are investigated as well. The outcomes of the comparative analysis showed that the identification of insider threats using deep hierarchical networks had a higher level of accuracy, clarity, and recall. However, the amount of time needed for training the algorithm of the intrusion detection system is considerable. A network detection model that is based on machine learning is offered here since the effectiveness of wireless intrusion detection systems was not well assessed in the earlier comparisons. You are able to pass over the step of conventional attribute selection thanks to the capabilities of machine learning that automatically extract and choose features. This makes it easier to calculate domain-specific, manually created features and reduces the complexity of doing so. Deep learning (DL), which is utilized often in a diverse range of industries and has been shown to be beneficial, is another example. As a result, throughout the course of the next several years, we will use machines and deep learning techniques in order to avoid DNNs, prevent overfitting with zero elements, and handle model training concerns with a restricted number of attack classifications. Enhances the efficiency with which intrusion detection and prevention are carried out.

Misunderstandings caused by the construction of contentious input, and eventually the solution to the issue of instability in cyber assaults.

ACKNOWLEDGEMENT

The heading should be treated as a 3rd level heading and should not be assigned a number.

REFERENCES

[1] Mohit, S. D., Gayatri, B. K., Vrushali, G. M., Archana, L. G., and Namrata, R. B. (2015). Utilizing Artificial Neural Network Classification and the Invention of an Intrusion in a Network Intrusion Detection System [Using Artificial Neural Network Classification and

Inventing an Intrusion in a Network]. International Journal of Innovative Research in Computer and Communication Engineering, Volume 3(2). International Journal of Innovative Research in Computer and Communication Engineering.
[2] Zaman S., El-Abed M., and Karray F. (2013, January)., Karray F. Features for selecting methods for intrusion detection systems that are based on evolving algorithms are called selection approaches.
[3] Nazir A. I. (2013). A study that looks at the similarities and differences between several artificial neural network–based intrusion detection systems. iInternationaliJournaliofiScientificiandiResearchiPublicationsi.
[4] Varma et al. P. R. K., Kumari et al. V., and Kumar et al. S. I. (2016).
iFeatureiselectioniusingirelativeifuzzyientropyiandianticolonyioptimizationiappliedittoireal-time intrusion detection system.
Featurei selection using relative fuzzy entropy andianticolonyioptimization. i85, i503-510 of the iProcediaiComputeriScience journal. i [5] Thaseen et al. (2017) and Kumar et al. (2017). Intrusion detection model created utilizing a combination of chi-square feature selection and multi-class support vector machines (SVM). iJournal of King Saud University-Computer and Information Sciences, 29(4), 462-472. iJournal of King Saud University-Computer and Information Sciences. i [6] Khammassi and Krichen (2017). iAiGA-LR iWrapperiApproach for iFeatureiSelection in iNetworkiIntrusion Detection. iAiGA-LR iWrapperiApproach for iFeatureiSelection in iNetworkiIntrusion Detection. iComputers and iSecurity, the i70 and the i255-277.
i [7] iAljawarnehiS, iAldwairiiM, and iYasseiniMiB (2018).
iAnomaly-based
iintrusionidetectionisystemithroughifeatureiselectionianalysisiandibuildingihybridiefficientimodel. iAnomaly-based intrusion detection system through feature selection analysis and model construction. iJournal of Computational Science, Volume 25, Issues 152-1613] [i25, i152-1613]
iKabir et al., iHu et al., iWang et al., &iZhuo et al. (2018). a brand-new statistical approach that may be used to intrusion detection systems.
i79, i303-318 iFutureiGenerationiComputeriSystems. i