A Review and Meta-Analysis for Efficient Intrusion Detection on KDD Dataset

Nidhi Shrivastava
M.Tech Student, CSE
LNCTE, Bhopal

Shrish Dixit
Assistant Professor, CSE
LNCTE, Bhopal

Shiv Kumar Sahu
HOD, CSE
LNCTE, Bhopal

ABSTRACT

In any network based system and organization identifying the possible attacks is very crucial and important to perceive the data integrity and security. Researchers are working in this field and several works is in progress. Due to the immense use, frequently updating in the data structure and large number of intrusions nature variability there are lot of scope in this area in terms of intrusion detection and classification. The main aim of this paper is to explore the gaps in the previous techniques and find out the methodologies by which any kind of hybridization is possible which can be capable in improving the classification accuracy.

Keywords

Intrusion detection techniques, KDD, DOS, U2R, R2L and probe

1. INTRODUCTION

The application which is capable of monitoring malicious behavior is called intrusion detection technique [1]. The main types of attacks considered for this study are denial of service (DoS), user to root (U2R), remote to login (R2L) and probe.

The procedure of pernicious session detection, identification when we are currently correspondence or removing information in the network environment [2] [3]. The key component in adhering and recognizing the components and malicious behavior is higher concern now days [3] [4]. Different methodologies and new algorithms have been proposed in this direction. Different authors presented their own views with some advantages and suggestions. So there are several gaps can be recognized in improving the classification accuracy.

Intrusion detection structure oversees managing the communication events happening in PC or framework circumstances and breaking down them for signs of possible events, which are infringement or unavoidable perils to PC security, or standard security practices Intrusion detection system (IDS) have ascended to distinguish exercises which risk the uprightness, protection or openness of are sourced as a push to give a response for existing security issues [5].

Considering the above facts we have reviewed and analyses few angles in the resulting segments. We likewise examine about information mining and advancement methods, in light of the fact that it can be utilized as a part of shaping the structure which delivers a better recognition framework.

As we analyze this study toward an unrivaled framework with the mix of data mining and streamlining. These strategies are profitable and has been used as a detection approaches like [6][7][8][9][10][11]. So the use of these counts can enhance an impact. The explores have broadened their perspectives in this bearing by a few exploration papers as in [12][13][14][15].There are several reports are reported in the direction of security and intrusion detection [7-20]. We have also discussed the security aspects as it can be easy for the understanding of security threats and their nature.

Sections for the forwarding of this paper are as follows. Section 2 describes the related works which describes the related research works in intrusion detection area. Section 3 describes the gap identification which shows the possible gaps in the traditional technique. Section 4 describes the discussion and analysis for the suggested solution based on the gaps identified. Section 5 shows the conclusions and future works based on the study.

2. RELATED WORKS

In 2011, LI [20] concentrated on intrusion detection and identification taking into account grouping the data. The point is to enhance the recognition rate and reduction the false alert rate. A changed element K-implies calculation called MDKM to recognize irregularity exercises is proposed and relating recreation investigations are introduced. Firstly, the MDKM calculation channels the clamor and secluded focuses on the information set. Furthermore by computing the separations between all example information focuses, they acquire the high-thickness parameters and bunch segment parameters, utilizing dynamic iterative procedure we get the k grouping focus precisely, then an inconsistency location model is introduced. They utilized KDD CUP 1999 information set to test the execution of the model. Their outcomes demonstrate the framework has a higher location rate and a lower false alert rate, it accomplishes hopeful point.

In 2011, Muda et al. [21] talk about the issue of current irregularity location that it not able to recognize a wide range of assaults effectively. To defeat this issue, they propose a crossover learning approach through mix of K-Means grouping and Naïve Bayes arrangement. The proposed methodology will bunch all information into the relating bunch before applying a classifier for order reason. A test is completed to assess the execution of the proposed approach utilizing KDD Cup '99 dataset. Result demonstrates that the proposed approach performed better in term of precision, identification rate with sensible false alert rate.

In 2012, LI [22] presented an improved FP-growth algorithm. They have suggested data preprocessing of which is capable in increasing efficiency in searching the prefix node so that time complexity is reduced. This technique has been applied for intrusion detection and the achieved results are effective and feasible.

In 2012, P. Prasenna et al. [23] recommended that in customary system security just depends on scientific calculations and low counter measures to taken to avoid interruption discovery framework, albeit the greater part of this methodologies as far as hypothetically tested to execute. Creators propose that as opposed to producing vast number of principles the advancement streamlining strategies like genetic network programming (GNP) can be utilized .The

GNP depends on coordinated diagram. They concentrate on the security issues identified with convey an information mining-based IDS in a constant situation. They sum up the issue of GNP with affiliation guideline mining and propose a fluffy weighted affiliation principle mining with GNP system appropriate for both ceaseless and discrete characteristics.

In 2014, Deshmukh et al. [24] presents a data mining technique in which different preprocessing strategies are included, for example, Normalization, Discretization and Feature determination. With the assistance of these strategies the information is preprocessed and required elements are chosen. They utilized NaIve Bayes technique as a part of an administered learning strategy which groups different system occasions for the KDD cup'99 Dataset.

In 2014, Benaicha et al. [25] present a Genetic Algorithm (GA) approach with an enhanced starting populace and choice administrator, to productively distinguish different sorts of system interruptions. They utilized GA to upgrade the pursuit of assault situations in review records, because of its great parity investigation/abuse; as indicated by the creators it gives the subset of potential assaults which are available in the review document in a sensible preparing time. The testing period of the Network Security Laboratory Knowledge Discovery and Data Mining (NSL-KDD99) benchmark dataset has been utilized to distinguish the abuse exercises. Their methodology of IDS with Genetic calculation expands the execution of the location rate of the Network Intrusion Detection Model and decreases the false positive rate.

In 2014 Kiss et al. [26] propose that Modern Networked Critical Infrastructures (NCI), including digital and physical frameworks, is presented to wise digital assaults focusing on the steady operation of these frameworks. To guarantee inconsistency mindfulness, their watched information can be utilized as a part of understanding with information mining strategies to create Intrusion Detection Systems (IDS) or Anomaly Detection Systems (ADS). They proposed a bunching based methodology for distinguishing digital assaults that bring about oddities in NCI. Different bunching methods are investigated to pick the most appropriate for grouping the time-arrangement information highlights, subsequently ordering the states and potential digital assaults to the physical framework. The Hadoop usage of MapReduce worldview is utilized to give an appropriate handling environment to vast datasets.

In 2014, Thaseen et al. [27] proposed a novel strategy for incorporating essential segment investigation (PCA) and bolster vector machine (SVM) by improving the bit parameters utilizing programmed parameter determination method. Their methodology decreases the preparation and testing time to distinguish interruptions in this manner enhancing the precision. Their proposed strategy was tried on KDD information set. The datasets were deliberately isolated into preparing and testing considering the minority assaults, for example, U2R and R2L to be available in the testing set to recognize the event of obscure assault. Their outcomes demonstrate that the proposed strategy is fruitful in distinguishing interruptions. Their test results demonstrate that the grouping exactness of the proposed strategy beats other order methods utilizing SVM as the classifier and other dimensionality lessening or highlight determination strategies.

In 2014, Wagh et al. [28] recommended Network security is a vital part of web empowered frameworks in the present world situation. As indicated by the creators because of mind boggling chain of PCs the open doors for interruptions and assaults have expanded. Accordingly it is need of great importance to locate the most ideal courses conceivable to secure our frameworks. So the creators propose interruption location frameworks are assuming indispensable part for PC security. The best strategy used to take care of issue of IDS is machine learning. Thy watched that the rising field of semi administered learning offers a guaranteed route for correlative exploration. So they proposed a semi-directed technique to lessen false alert rate and to enhance location rate for IDS.

In 2014, Masarat et al. [29] presented a novel multistep structure in view of machine learning systems to make an effective classifier. In first step, the component choice strategy will actualize taking into account pick up proportion of elements by the creators. Their technique can enhance the execution of classifiers which are made taking into account these components. In classifiers blend step, we will introduce a novel fluffy gathering strategy. Thus, classifiers with more execution and lower cost have more impact to make the last classifier.

In 2015, Bahl et al. [30] suggested U2R attack class is an open research problem. Their purpose of this study is to identify the important features to improve the detection rate and reduce the false detection rate. The researched highlight subset choice methods enhance the general exactness, discovery rate of U2R attack class furthermore decrease the computational expense. The experimental results have demonstrated a discernible change in location rate of U2R attack class with highlight subset determination systems.

In 2015, Yan et al. [31] proposed an intelligent intrusion detection model. Taking into account the attributes of worldwide predominance of hereditary calculation and area of nerve, the model upgrades the weights of the neural system utilizing genetic algorithm. Test results demonstrate that the insightful way can enhance the proficiency of the interruption identification.

In 2015, Haidar et al. [32] emphasizes the significance of abnormality based interruption recognition procedures, the vital results of these frameworks, most recent created techniques and what is normal from the future trials in this field. In addition, the method of learning client profiles impacts in recognizing interruptions can be explored. Finally, the lights will be shed on an offline approach using Multi-Layer Perceptron (MLP) and Self Organizing Maps (SOM) which is a distinguished method in intrusion detection.

3. GAPS IDENTIFICATION

Based on the study and analysis presented in this paper we came with the following identification:

1) There is a need of improving in classification accuracy.
2) Classification based on attacks is also needed as it can validate the attack properly.
3) The classification can be based on optimization so that optimal solution can be achieved.
4) Machine learning approaches can be helpful in detection.
5) There are variant accuracies in different attacks.
6) Maintain long log files for detection.

7) The detection rates are low in case of U2R and R2L, so there is the need of an improvement in the concern attacks.

4. DISCUSSION AND ANALYSIS

There is several research works are analyzed and discussed here. Based on the analysis there is need of improvement in all attacks like DoS, U2R, R2L and Probe. But special concentration is needed is needed in case of U2R and R2L. Attack subset properties can be extracted by the subset superset approach [33]. This extraction will help in validation the attacks classification. Below table shows the detail analysis of the results obtained by different methodology.

Table 1: Analysis

S.no	Approach	Accuracy (DOS %)	Accuracy (U2R %)	Accuracy (R2L %)	Accuracy (Probe %)
1	Neural Network using Genetic Algorithm [31]	97.1	83.9	80.5	82
2	Multi-Layer Perceptron (MLP) and Self Organizing Maps (SOM) [32]	Only discuss intrusion attempts	Only discuss intrusion attempts	Only discuss intrusion attempts	Only discuss intrusion attempts
3	Feature Subset Selection [30]	80.80	80.80	86.02	86.02
4	Genetic Algorithm and Fuzzy Logic [34]	Efficient rule set generations	Efficient rule set generations	Efficient rule set generations	Efficient rule set generations
5	K-Means clustering and Naïve Bayes classification [21]	97.14	96.11	94.10	NA
6	K-Means clustering and Naïve Bayes classification [21]	89.90	84.32	83.23	NA
7	K-Means clustering and Naïve Bayes classification [21]	95.23	92.14	91.21	NA
8	K-Means clustering and Naïve Bayes classification [21]	91.34	86.14	85.11	NA
9	K-Means clustering and Naïve Bayes classification [21]	95.12	93.21	91.13	NA
10	Fuzzy ensemble of classifiers [36]	93.00	NA	NA	NA
11	Data mining approach [37]	97.5	48.0	95.0	92.7

5. CONCLUSIONS AND FUTURE WORKS

This paper provides a systemic review and analysis in the direction of creating an efficient intrusion detection framework. The analysis of the paper provides an intrinsic and wide idea behind several intrusion detection techniques and the comparative study which is capable of understanding the gaps and identification. Based on the analysis and understanding the following future implications can be suggested.

- Separate classification is needed to done in each intrusion case.
- Combination of evolutionary technique can be applied simultaneously.
- A rule can be classified dynamically with positive and negative association.
- Random classification can be done on several iteration to check the improvement.
- There is a need of improvement in case of U2R and R2L.

6. REFERENCES

[1] Farhaoui Y. How to secure web servers by the intrusion prevention system (IPS)? International Journal of Advanced Computer Research. 2016 Mar 1; 6(23):65.

[2] Jianliang M, Haikun S, Ling B. The application on intrusion detection based on k-means cluster algorithm. InInformation Technology and Applications, 2009. IFITA'09. International Forum on 2009 May 15 (Vol. 1, pp. 150-152). IEEE.

[3] Kabiri P, Ghorbani AA. Research on Intrusion Detection and Response: A Survey. IJ Network Security. 2005 Sep; 1(2):84-102.

[4] Park HA. Secure chip based encrypted search protocol in mobile office environments. International Journal of Advanced Computer Research. 2016; 6(24):72-80.

[5] Tiwari R, Sinhal A. Block based text data partition with RC4 encryption for text data security. International Journal of Advanced Computer Research. 2016; 6(24):107-13.

[6] Tian L, Jianwen W. Research on network intrusion detection system based on improved k-means clustering algorithm. In Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on 2009 Dec 25 (Vol. 1, pp. 76-79). IEEE.

[7] Devaraju S, Ramakrishnan S. Analysis of Intrusion Detection System Using Various Neural Network classifiers. IEEE 2011. 2011:1033-8.

[8] Conteh NY, Schmick PJ. Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks. International Journal of Advanced Computer Research. 2016 Mar 1; 6(23):31.

[9] Lee HY, Wang NJ. The implementation and investigation of securing web applications upon multi-platform for a single sign-on functionality. International Journal of Advanced Computer Research. 2016 Mar 1; 6(23):39.

[10] Ishida M, Takakura H, Okabe Y. High-performance intrusion detection using optigrid clustering and grid-based labelling. InApplications and the Internet (SAINT), 2011 IEEE/IPSJ 11th International Symposium on 2011 Jul 18 (pp. 11-19). IEEE.

[11] Brugger ST. Data mining methods for network intrusion detection. University of California at Davis. 2004 Jun 9.

[12] Lee W, Stolfo SJ. Data mining approaches for intrusion detection. In Usenix security 1998 Jan 26.

[13] Nalavade K, Meshram BB. Mining Association Rules to Evade Network Intrusion in Network Audit Data. International Journal of Advanced Computer Research. 2014 Jun 1;4(2):560.

[14] Naoum R, Aziz S, Alabsi F. An Enhancement of the Replacement Steady State Genetic Algorithm for Intrusion Detection. International Journal of Advanced Computer Research. 2014 Jun 1; 4(2):487.

[15] Lee W, Stolfo SJ, Mok KW. A data mining framework for building intrusion detection models. InSecurity and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on 1999 (pp. 120-132). IEEE.

[16] Kumari S, Shrivastava M. A Study Paper on IDS Attack Classification Using Various Data Mining Techniques. International Journal of Advanced Computer Research. 2012; 2(3).

[17] Venkatesan R, Ganesan R, Selvakumar AA. A Comprehensive Study in Data Mining Frameworks for Intrusion Detection. International Journal of Advanced Computer Research (IJACR). 2012; 2: 29-34.

[18] Patel R, Bakhshi D, Arjariya T. Random Particle Swarm Optimization (RPSO) based Intrusion Detection System. International Journal of Advanced Technology and Engineering Exploration. 2015; 2(5): 60-66.

[19] Sperotto A, Schaffrath G, Sadre R, Morariu C, Pras A, Stiller B. An overview of IP flow-based intrusion detection. Communications Surveys & Tutorials, IEEE. 2010 Jul 1; 12(3):343-56.

[20] Han LI. Using a dynamic K-means algorithm to detect anomaly activities. In Computational Intelligence and Security (CIS), 2011 Seventh International Conference on 2011 Dec 3 (pp. 1049-1052). IEEE.

[21] Muda Z, Yassin W, Sulaiman MN, Udzir NI. Intrusion detection based on K-Means clustering and Naïve Bayes classification. In Information Technology in Asia (CITA 11), 2011 7th International Conference on 2011 Jul 12 (pp. 1-6). IEEE.

[22] Yin-huan LI. Design of intrusion detection model based on data mining technology. In2012 International Conference on Industrial Control and Electronics Engineering 2012 Aug 23.

[23] Prasenna P, RaghavRamana AV, Krishnakumar R, Devanbu A. Network programming and mining classifier for intrusion detection using probability classification. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on 2012 Mar 21 (pp. 204-209). IEEE.

[24] Deshmukh DH, Ghorpade T, Padiya P. Intrusion detection system by improved preprocessing methods and Naïve Bayes classifier using NSL-KDD 99 Dataset. In Electronics and Communication Systems (ICECS), 2014 International Conference on 2014 Feb 13 (pp. 1-7). IEEE.

[25] Benaicha SE, Saoudi L, Guermeche B, Eddine S, Lounis O. Intrusion detection system using genetic algorithm. InScience and Information Conference (SAI), 2014 2014 Aug 27 (pp. 564-568). IEEE.

[26] Kiss I, Genge B, Haller P, Sebestyen G. Data clustering-based anomaly detection in industrial control systems. In Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on 2014 Sep 4 (pp. 275-281). IEEE.

[27] Thaseen IS, Kumar CA. Intrusion detection model using fusion of PCA and optimized SVM. In Contemporary Computing and Informatics (IC3I), 2014 International Conference on 2014 Nov 27 (pp. 879-884). IEEE.

[28] Wagh SK, Kolhe SR. Effective intrusion detection system using semi-supervised learning. In Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on 2014 Sep 5 (pp. 1-5). IEEE.

[29] Masarat S, Taheri H, Sharifian S. A novel framework based on fuzzy ensemble of classifiers for intrusion

detection systems. In Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on 2014 Oct 29 (pp. 165-170). IEEE.

[30] Bahl S, Sharma SK. Improving Classification Accuracy of Intrusion Detection System Using Feature Subset Selection. In Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on 2015 Feb 21 (pp. 431-436). IEEE.

[31] Yan C. Intelligent Intrusion Detection Based on Soft Computing. In Measuring Technology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on 2015 Jun 13 (pp. 577-580). IEEE.

[32] Haidar GA, Boustany C. High Perception Intrusion Detection Systems Using Neural Networks. Ninth International Conference on Complex, Intelligent, and Software Intensive Systems 2015 (pp. 497-501). IEEE.

[33] Dubey AK, Dubey AK, Agarwal V, Khandagre Y. Knowledge discovery with a subset-superset approach for Mining Heterogeneous Data with dynamic support. In Software Engineering (CONSEG), 2012 CSI Sixth International Conference on 2012 Sep 5 (pp. 1-6). IEEE.

[34] Hassan MM. Current studies on intrusion detection system, genetic algorithm and fuzzy logic. arXiv preprint arXiv:1304.3535. 2013 Apr 12.

[35] Kumar A, Maurya HC, Misra R. A Research Paper on Hybrid Intrusion Detection System. International Journal of Engineering and Advanced Technology (IJEAT). 2013;2: 2249-895.

[36] Masarat S, Taheri H, Sharifian S. A novel framework, based on fuzzy ensemble of classifiers for intrusion detection systems. InComputer and Knowledge Engineering (ICCKE), 2014 4th International eConference on 2014 Oct 29 (pp. 165-170). IEEE.

[37] Mukkamala S, Abraham AS. Designing Intrusion Detection Systems: Architectures, Challenges and Perspectives. Department of Computer Science, Oklahoma State University, USA. 2003.