International Journal of Engineering & Technology, 7 (2.33) (2018) 536-540

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Comparison of intrusion detection system based on feature extraction

Pradeep Laxkar [1] *, Prasun Chakrabarti [2]

[1] PhD Scholar, Department of CSE, SPSU Udaipur SPSU, Udaipur Udaipur, India
[2] Professor, Department of CSE, SPSU Udaipur SPSU, Udaipur Udaipur, India
**Corresponding author E-mail: Pradeep.laxkar@gmail.com*

Abstract

In network traffic classification redundant feature and irrelevant features in data create problems. All such types of features time-consuming make slow the process of classification and also affect a classifier to calculate accurate decisions such type of problem caused especially when we deal with big data. In this paper, we compare our proposed algorithm with the other IDS algorithm.

Keywords: *IDS; Big Data; Feature Selection; Spark*

1. Introduction

An intrusion detection system (IDS) is a device with a firmware or software combination of both that monitors a computer network or computer systems for malevolent activity or policy violations. From the observation of IDS such detected activity or violation is typically reported either to a system administrator for further action. An [1] IDS performs various functions like monitoring of users and system activity, recognizing anomalous action through numerical study, identifying known attack patterns in any node activity , by changing system arrangement errors in a right way , configuring and operating traps to save information about intruders , checking node configuration for vulnerabilities and misconfigurations , IDS do management of audit trails and stress user breach of rule or normal activity also it eases the integrity of significant node and important data files.

IDS are divided into two [2] fold, one is based on the method of detection which includes signature-based and anomaly-based IDS. Other IDS ARE host-based (HIDS) versus network-based IDS (NIDS).

i) Signature-based IDS: The IDS identifies known intrusive behavior. Other behavior is by default not reported, that is, these systems provide a default outcome of the permit (or legal). Signature based IDS work on predefine libraries of signature which are made in real time and static network attacks. The signature of the attack is matched with network data and IDS take a decision. These types of IDS constrain the range of attacks that can possibly be detected in return for an acceptable error rate in detection. It covers the entire attack space, at the cost of increased error rates. The latter is due to the fundamental problem that an anomaly is not necessarily an attack, something alluded to earlier. It is indeed often not an attack, and this leads to the major failing of many such systems, that is, the problem of a high false positive or false alert rate [2].

ii) Anomaly-based systems: this type of IDS identifies deviations from normal behavior. They use a model of normal behavior and report any activity which does not conform to the normal behavior, thus providing a default outcome of deny (or illegal).

Feature Selection: Feature extraction algorithm [3] can be divided into two steps:

Step 1: Feature construction:

Feature selection is a method for removing inappropriate and redundant features. It selects a finest subset of features. Subset of feature is given as input to produce a better characterization of patterns, which belongs to different classes.

Step 2: Feature selection:

Feature construction is the most important step in the data demonstration method for lots of tasks. We can use feature construction to perform many operations like classification and regression, for the most part the achievement of any consequent value or modeling of an input raw data. Such development refers to shaping representative features of the original data. We can develop automatic feature construction by using several methods like as n-grams, association rule learning and frequency episode methods.

2. Literature review

1) In [4] this paper authors modifies FVBRM techniques for feature selection and compared with original FVBRM and other feature selection methods. In proposing methods they have done feature selection based on vitality.
2) In [5] this paper authors used various filtering examinations with discriminative multinomial Naïve Bayes to construct a network intrusion detection system. To perform an experimental analysis, they used the new NSL-KDD dataset. They have performed [2] class classifications with 10-fold cross validation for building their proposed model.
3) In [6] this Paper author used Deep belief neural (DBN) networks to build IDS.DBN network are most dominant deep neural nets it is also important neural networks that stack controlled Boltzmann Machines. In this paper author used all the capabilities of DBN's to develop an effective intrusion detection system. They have performed a series of experiments after training it with NSL-KDD dataset.

4) In [7] this paper authors proposed a Bi-Layer behavioral-based feature selection approach. They divided their approach in two layers
   i) In the first layer authors used information gain to identify the rank of features and then the selected a new set of features which are depending on a global maxima classification precision.
   ii) In second layer authors selected a new set of features from data which were redacted in the first layer. In redacting data they searched a group of local maximum classification accuracy. Which leads to increase the quantity of compact features.
5) In this paper [8] authors proposed and validate an IDS method for selecting an optimal feature subset. The proposed method is based on the study of the Pearson correlation coefficients. They have used the correlation analysis as a base and identified analysis between two variables. The correlation analysis is used to decide the quality of future goods.

3. Proposed method

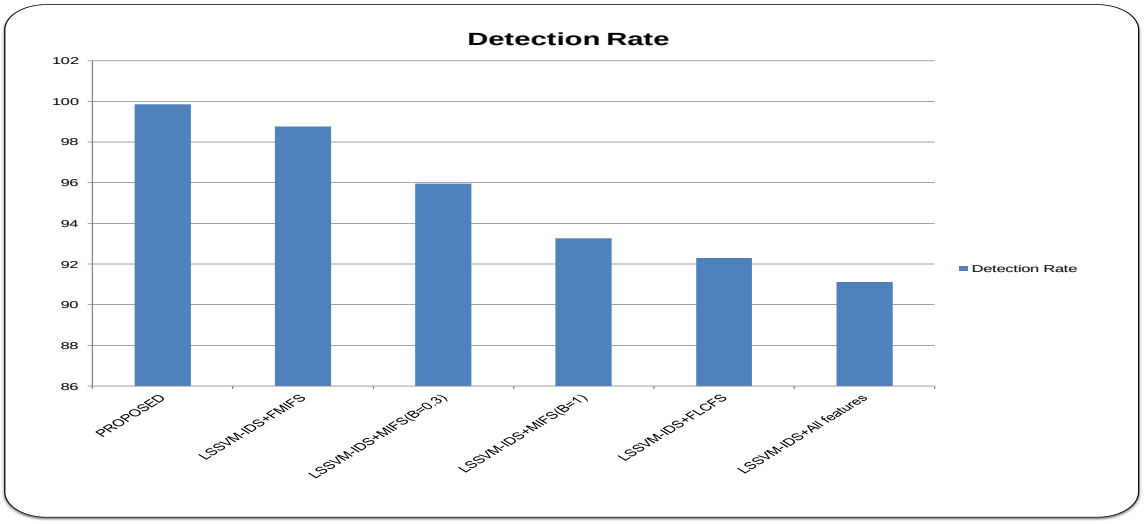
In our [9] method we have used a packet collector (Wireshark) capture incoming packet. We have converted a packet to NSL-KDD format. The data are pre-processed and loaded into apache spark RDD. We have applied feature extraction and selection method of data using sparks machine learning library. Performance is calculated using Spark Mllib classifier mode.

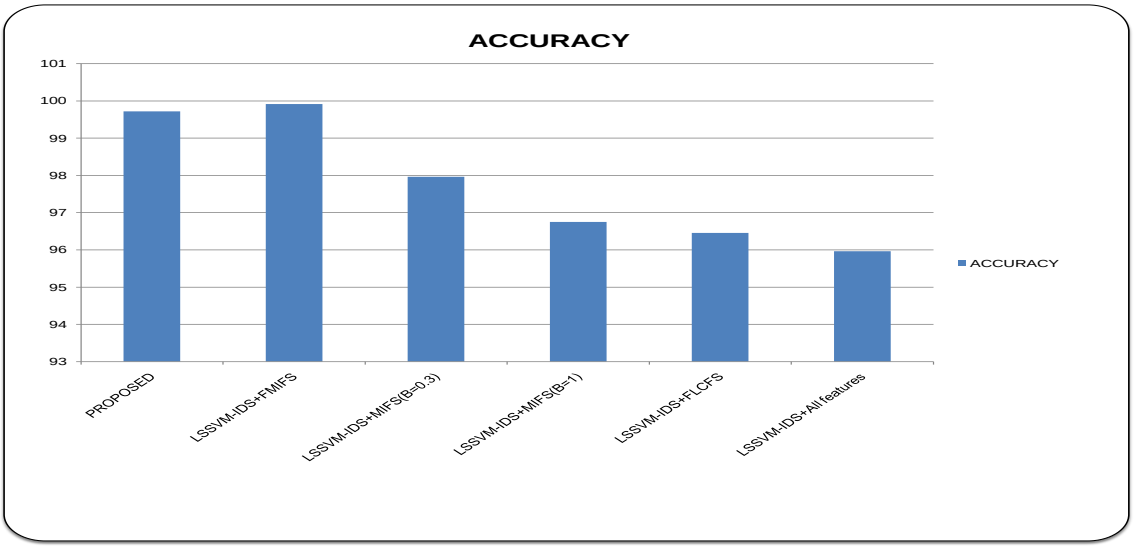
4. Comparison

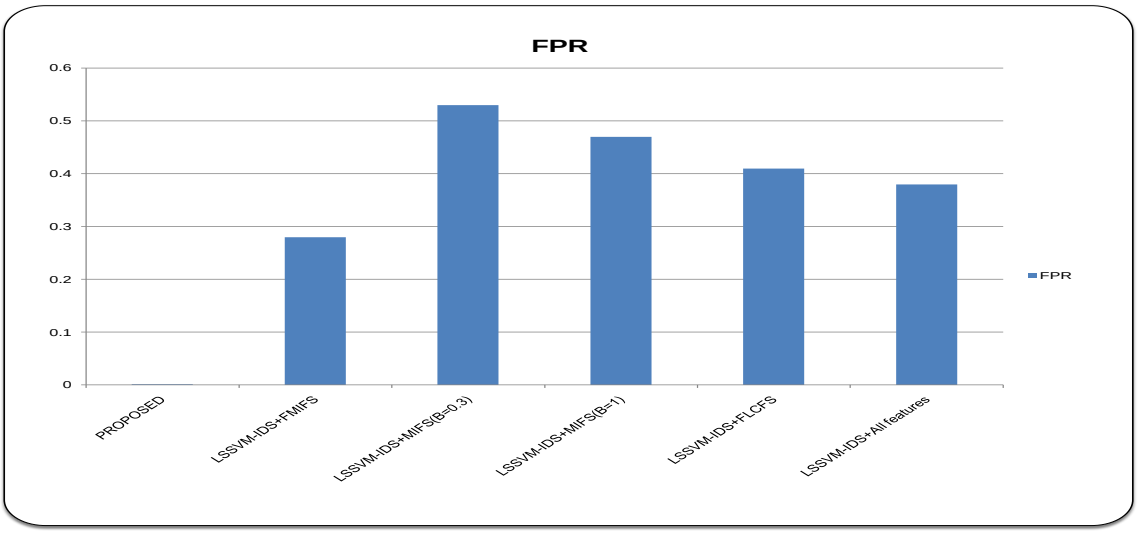
Comparison is done with following algorithms:

i) LSSVM-IDS+FMIFS [3].
ii) LSSVM-IDS+MIFS (B=0.3)
iii) LSSVM-IDS+MIFS (B=1) [3].
iv) LSSVM-IDS+FLCFS.
v) LSSVM-IDS+All features.
vi) DMNB.
vii) Multinomial Naïve Bayes + N2B.
viii) Multinomial Naïve Bayes updateable + N2B [5].
ix) Discriminative Multinomial Naïve Bayes +PCA [5].
x) Discriminative Multinomial Naïve Bayes +RP [5].
xi) Discriminative Multinomial Naïve Bayes +N2B [5].
xii) SVM.
xiii) DBN.
xiv) DBN-SVM.
xv) Bi-layer behavioral based.
xvi) CFS+Best First.
xvii) GR+Ranker.
xviii) IG+ Ranker.
xix) FVBRM C4.5.
xx) Hybrid C4.5 with linear correlation-based.
xxi) Bi-layer behavioral based.

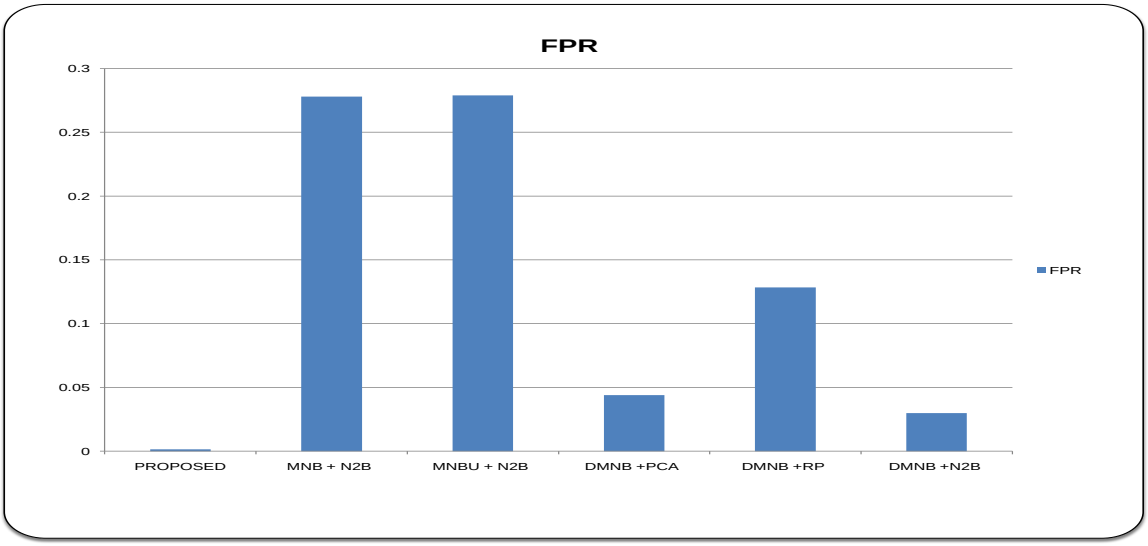
a) Comparison with Least Square Support Vector Machine Based IDS (LSSVM-IDS).



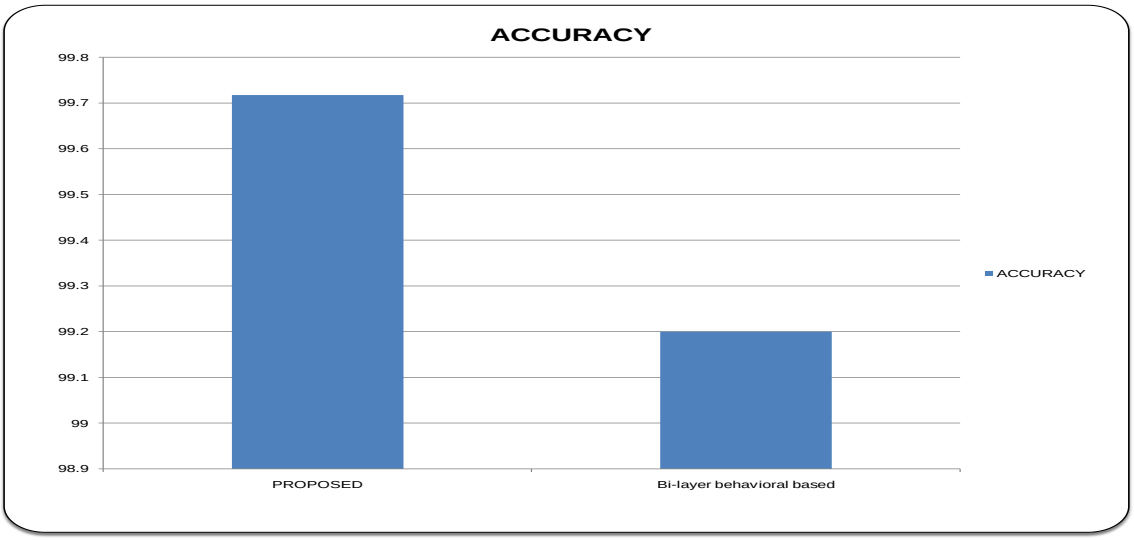




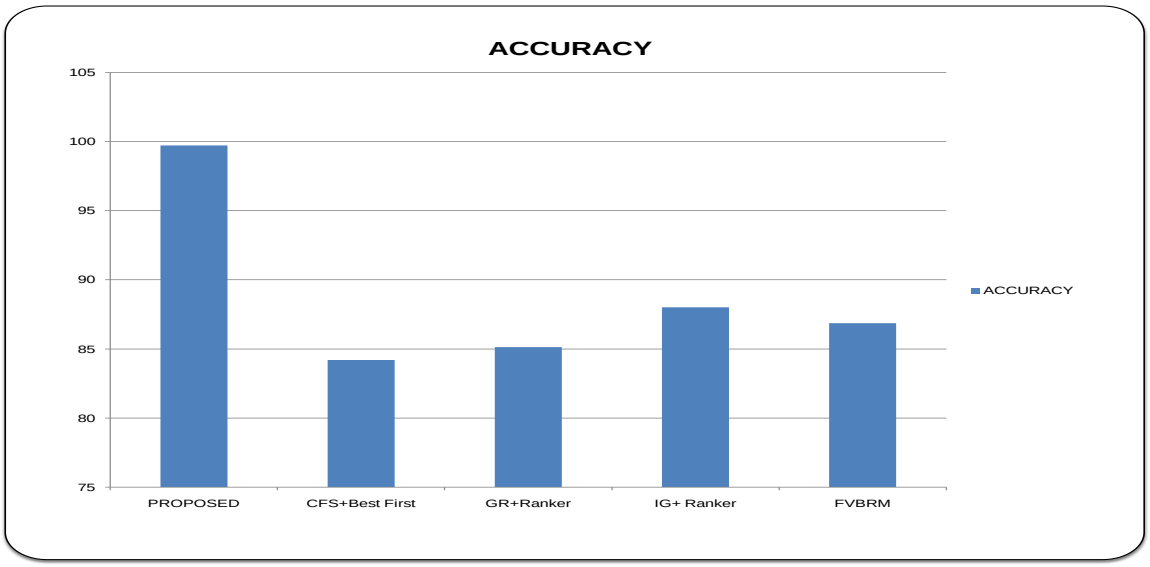
b) Comparison with Discriminative Multinomial Naïve Bayes (DMNB)

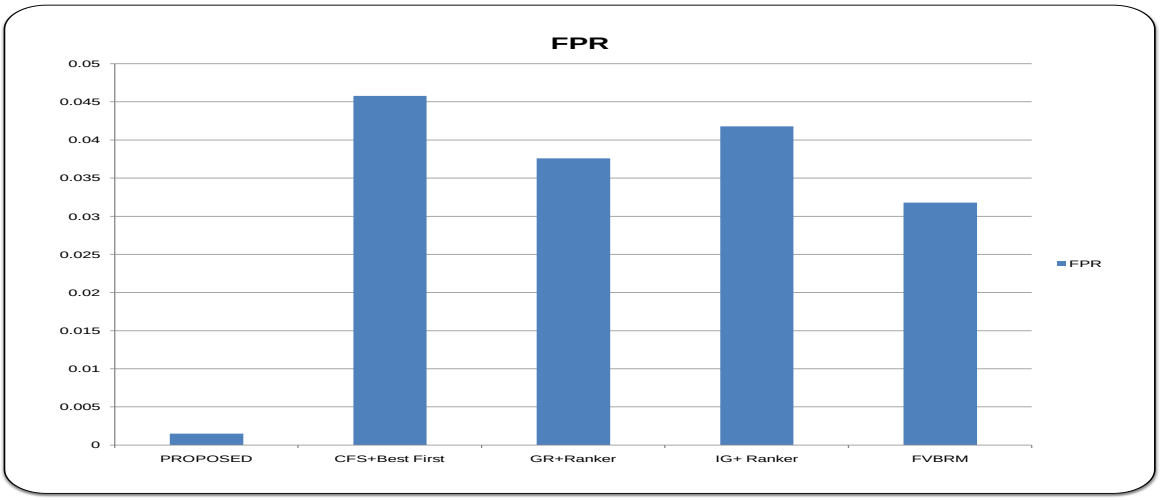


c) Comparison with Bi-layer behavioral based IDS.

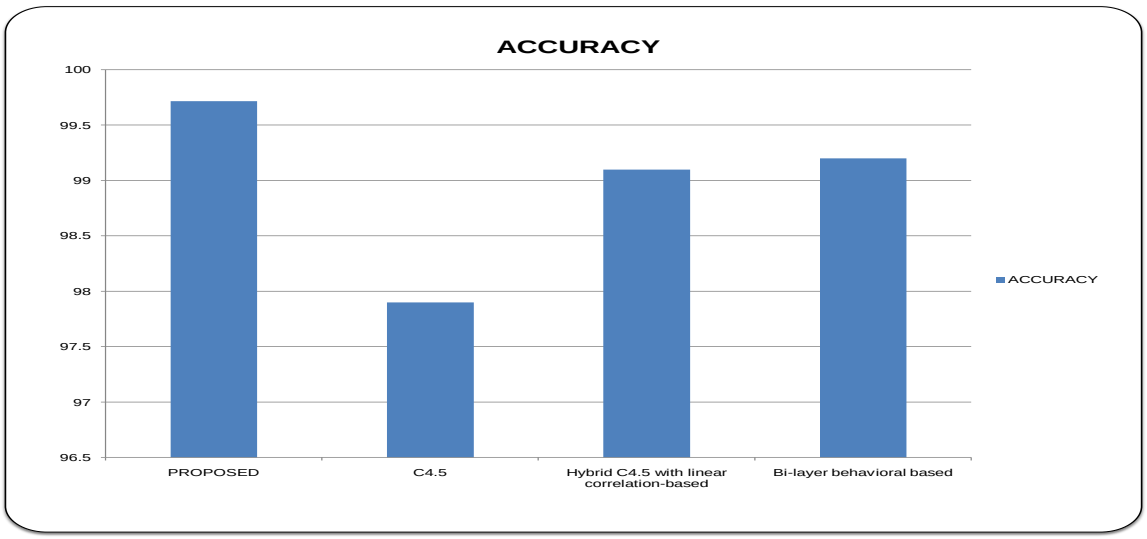


d) Comparison with Other Feature selection Methods.





e) Comparison with Correlation-Based Feature Selection IDS.



5. Conclusion

In this paper we have compare our proposed algorithm with various IDS algorithms like LSSVM-IDS+FMIFS , LSSVM-IDS+MIFS(B=0.3), LSSVM-IDS+MIFS(B=1) , LSSVM-IDS+FLCFS , LSSVM-IDS+All features ,DMNB, Multinomial Naïve Bayes + N2B ,[5]Multinomial Naïve Bayes updateable + N2B , Discriminative Multinomial Naïve Bayes +PCA , Discriminative Multinomial Naïve Bayes +RP , Discriminative Multinomial Naïve Bayes +N2B ,SVM , DBN , DBN-SVM , Bi-layer behavioral based , CFS+Best First , GR+Ranker , IG+ Ranker , FVBRM C4.5,Hybrid C4.5 with linear correlation-based ,Bi-layer behavioral based algorithm. We have used various parameters for comparison like False alarm rate, Accuracy, detection rate and F-1

score. We have observed that the proposed method is more efficient than all other methods.

References

[1] Menu Bijone, A Survey on Secure Network: Intrusion Detection & Prevention Approaches in American Journal of Information Systems Vol. 4, No. 3, 2016, pp 69-88. Do: 10.12691/ages-4-3-2

[2] Mohay, George M. Computer and intrusion forensics. Artech House, 2003.

[3] Nguyen, Hai Thanh, Katrin Franke and Slobodan Petrovic. "Feature Extraction Methods for Intrusion Detection Systems." Threats, Countermeasures, and Advances in Applied Information Security. IGI Global, 2012. 23-52. Web. 13 Feb. 2018. doi:10.4018/978-1-4666-0978-5.ch002

[4] Jupriyadi and A. I. Kistijantoro, "Vitality based feature selection for intrusion detection," 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA), Bandung, 2014, pp. 93-96.

[5] 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, 2010, pp. 5-10.

[6] M. Z. Alom, V. Bontupalli and T. M. Taha, "Intrusion detection using deep belief networks," 2015 National Aerospace and Electronics Conference (NAECON), Dayton, OH, 2015, pp. 339-344.

[7] Heba F. Eid , Mostafa A. Salama , boul Ella Hassanien , Tai-hoon Kim Bi-Layer Behavioral-Based Feature Selection Approach for Network Intrusion Classification" , International Conference on Security Technology SecTech 2011: Security Technology pp 195-203

[8] Eid H.F., Hassanien A.E., Kim T., Banerjee S. (2013) Linear Correlation-Based Feature Selection for Network Intrusion Detection Model. In: Awad A.I., Hassanien A.E., Baba K. (eds) Advances in Security of Information and Communication Networks. Communications in Computer and Information Science, vol 381. Springer, Berlin, Heidelberg.

[9] Laxkar P., Chakrabarti P.,Ghosh A. and Panwar P., "An effective Intrusion Detection System Using Machine Learning Library of Spark", International Journal of Emerging Technology and Advanced Engineering, 8(2),pp.48-52,2018