# Modified Naive Bayes Intrusion Detection System (MNBIDS)

Karuna S. Bhosale– Research Student
Faculty of Telecommunication
Technical University of Sofia
Sofia, Bulgaria
E-mail: bhosale.karuna@gmail.com

Assoc. Prof. Maria Nenova, PhD – Research Guide
Prof. Georgi Iliev, PhD—Research Guide
Faculty of Telecommunication
Technical University of Sofia
Sofia, Bulgaria,
E-mail: mvn@tu-sofia.bg, gli@tu-sofia.bg

***Abstract*—in this result paper we presenting Modified Naïve Bayes Intrusion Detection System (MNBIDS), it is based on the existing Naïve Bayes system. In that we perform the data pre-processing, data normalization and feature extraction. Now a day's Network security is turning into an expanding vital issue, since the fast advancement of the Web. Information mining and machine learning innovation have been broadly connected to network interruption recognition and anticipation frameworks by finding client behavior standards from the network traffic information. In this system, we use real time packet, which is used to real time analysis and also the KDD Cup 99 dataset for the execution. In this system we use the different classifiers on this real time packets and KDD dataset for the comparison of obtained results. In this system we use the Data Pre-processing algorithm, Hybrid Feature Selection Algorithm and Modified Naïve Bayes Algorithm. Using these algorithms we improve the system accuracy and execution performance. After that we compare the SVM, CNN, KNN, ANN, and Proposed Naïve Bayes Classifiers to obtain the better results. We use LOIC DDoS assault generator instrument to make the assault at bundle getting time what's more use KDD compartment dataset for various assault pursue, for instance, DoS, R2L, L2R, Test and whatnot of correspondence orchestrate. The test results that proposed MNBIDS Strategy is progressively precise and valuable.**

**Keywords- Modified Naïve Bayes, DDoS, KDD Cup 99, SVM, CNN, KNN, ANN.**

## I. INTRODUCTION

Nowadays, the intrusion happens in a small number of seconds. Gate-crashes keenly utilize the updated version of command and in this way eradicating their impressions in review and log documents. Effective IDS intellectually separate both intrusive and nonintrusive records [1]. The vast majority of the current frameworks have security breaks that make them effortlessly powerless and couldn't be comprehended. In addition, generous research has been going on intrusion detection method which is as yet assumed as youthful and not an accurate instrument against intrusion. It has additionally turned into a most need and challenging assignments for network overseers and security specialists. So it can't be supplanted by more secure frameworks.

Currently the network security is important point to research because all peoples are using or connected to each other using internet and also they share their data on it. The some attacks or disturbance happen in the network connection and user data has been the broken and miss used it [2]. To avoid the DDoS attacks many researchers are developed the different methods but anyone not avoid this attack accurately. The Naïve Bayes method is better to avoid but not accurate that why we need to modify the Existing Naïve Bayes system to make it more accurate and solve the real time DDoS attacks on the packets captured or KDD Cup 99 dataset [3].

In this system we capture the real time network packets to get the input of our modified system. The JPCAP library is used to capture and store the real time network packets. To remove the Noise of this packets and KDD Cup 99 we use the Noise Removal Algorithm [4].

The need to address this kind of clamour is clear as it is impeding to any kind of information examination. Regardless, basic information questions that are not fitting or only material to a specific information investigation yet not indispensable can in like [5] way significantly pulverize the information examination, and as such, these articles should likewise be considered as clamour, in any occasion concerning a specific examination [6]. For example, in report educational lists that contain news stories, there are particular records that are just pitifully identified with trade news stories. On the off chance that the centres is to utilize gathering to find the solid subjects in a great deal of chronicles, by then the examination will hold on beside if unessential can be disposed of. Thusly, it is central for data cleaning systems that vacant all sort of commotion [7].

These models consolidate the web information starting late made reference to record informational indexes [8]. In hence,

978-1-5386-7709-4/18/$31.00 ©2018 IEEE

information cleaning techniques for the update of information investigation similarly ought to have the ability to lessen the time range likewise improve the exactness [9]. Nevertheless the continuous information parcel getting is performed with various immaterial messages in each stream. The proximity of such insignificant information in got information may prompts mixed up area assaults [10]. Thusly, beforehand planning to feature extraction and assaults disclosure utilizing information mining calculations, we have to play out the information pre-dealing with to empty such information content information content from catch information with the true objective to upgrade the adequacy and precision of security investigation. In this paper, we showed the calculation to get the dependable framework bundles which contain with assault streams. We used the assault generator contraption to unequivocally make the unquestionable sorts of assaults, for instance, R2L, U2R, DoS and Test [11].

The MNBIDS technique is based on the existing Naïve Bayes method. In this system we use a Noise Removal method, Feature Extraction Method and proposed or modified algorithm to obtain the more accuracy and better performance in the security on Network from DDoS attacks.

In this paper we define the literature survey in section II. In section III we introduced the Problem Definition. In section VI it presents the proposed algorithms and section V describes the results. The last section VI it presents the conclusion.

## II. Related Works

In this section we presenting the all recent techniques and presents its features, works advantages, disadvantage.

In [1] the maker is arranged Classification is a commendable data mining technique dependent on machine learning. Portrayal is utilized to orchestrate everything in a great deal of data into one of predefined set of classes or gatherings. Naïve Bayes is an every now and again utilized gathering directed learning methodology to foresee the class likelihood of having a place. This paper proposes another methodology for Naïve Bayes Algorithm in which we attempted to discover impacting district rate and false positive rate of given data. We endeavored the execution of our proposed algorithm by utilizing KDD99 benchmark orchestrate intrusion seeing confirmation dataset, and the exploratory outcomes demonstrated that it updates an area rates and besides decreases false positives for different sorts of framework interferences.

In [2] this paper, maker displays another a learning algorithm for adaptable intrusion acknowledgment utilizing boosting and naïve Bayesian classifier, which considers the headway of classifiers and joins the votes of every individual classifier for arranging a dull or known model. The proposed algorithm makes the likelihood set for each round utilizing genuine Bayesian classifier and updates the heaps of planning points of view subject to the misclassification mess up rate that made by the organizing models in each round. This algorithm looks out for the issue of organizing the expansive intrusion ID dataset, which enhances the acknowledgment rates (DR) and decreases the false positives (FP) at an honorable measurement in interference revelation. We attempted the execution of the proposed algorithm with existing data mining algorithms by utilizing on the KDD99 benchmark intrusion distinguishing proof dataset, and the foundation results demonstrated that the proposed algorithm accomplished high area rates and supreme decreased the proportion of false positives for different sorts of system interferences.

In [3] the author designed network security has turned into a key issue in information technology as there is expanding in security threats. A network interruption identification system screens movement on a network searching for suspicious action, which could be an assault or unapproved action. The majority of the scientists were utilized KDD99 dataset yet in this paper we are making our own dataset by constant packet catching. We utilize Naïve Bayes classifier for interruption discovery which characterizes whether the attack is available or not.

In [4] this paper, creator present and dissect a way to deal with interruption recognition utilizing naïve Bayes network. In addition, we think about various routes for dealing with persistent variables. The distinctive trial results demonstrate that guileless Bayesian networks, Bayesian networks, notwithstanding having a direct structure, can give compelling and correct classifiers to effective and exact classifiers to identify interruptions.

In [5] the creator applies discriminative multinomial Naïve Bayes with various filtering examination with the ultimate objective to build a system intrusion revelation framework. For our test investigation, we used the new NSL-KDD dataset, or, at the end of the day a changed dataset for KDD Cup 1999 interference ID benchmark dataset. We perform 2 class orders with 10-wrinkle cross approval for building our proposed model.

In [6] the author focuses on the development of internet condition has in like manner accomplished to increment in end-client suspicious exercises. Each client gets related to the network condition which progressions unapproved practices in the system.

In [7] the creator exhibits the tremendous computational regard has reliably been an imprisonment in taking care of giant system intrusion data. This issue can be lightened through component choice to condense the degree of the system data included. In this paper they first arrangement existing component affirmation strategies that are computationally executable for dealing with epic system impedance datasets. In this paper, they study and examination of four machine learning algorithms (J48, BayesNet, OneR, NB) of data digging for the errand of recognizing obstructions and look at their relative execution. In light of this examination, it might be shut the J48 choice tree is the most sensible related figuring then the other three algorithms with high true positive rate (TPR) and low false positive rate (FTR) and low estimation time with high accuracy.

In [8] the creator exhibits the vast of Internet benefits wherever all through the world, various sorts and broad various security dangers are developing. Since it isn't very to make a system without any vulnerabilities, Intrusion Detection System (IDS), which can attainably recognize intrusion finds the opportunity to have pulled in thought. Intrusion detection can be described as the way toward seeing odd lead those objectives a system and its advantages. An ID looks data highlights to perceive unapproved or unapproved advancement identified with Network Security. It acknowledges a key occupation for isolating the system action log and each log is portrayed by a gigantic arrangement of highlights and it requires huge computational preparing force and time for the social affair process. This paper envisions to perceive basic reduced features with select by highlight Quintile channel and Chi-Squared. The dimension of this work is to see grouped classes of assaults using Naïve Bayes, Radial Basis and J-48 classifiers are sorted out and attempted uninhibitedly and the get-together rates for different classes are seen. The Naïve Bayes classifier has defeated well concerning Exactness and Arrangement mistake rate differentiated what's more, J-48 and RBF classifier. Test outcomes demonstrate that picked credits give better execution to course of action inducing IDS.

In [9] this paper gives a record of the observational appraisal of five machine learning algorithm, for instance, J48, BayesNet, OneR, NB, and ZeroR using execution criteria: precision, exactness, survey, F-Measure, inaccurately described events, kappa estimation, mean aggregate blunder, root mean squared mistake, relative preeminent blunder, root relative squared mistake. The reason for this paper is to discover which classifier is better in its execution for the intrusion detection system. Machine Learning is one of the techniques used in the intrusion detection system. (IDS).Based on this examination, it might be contemplated that a J48 choice tree is the most sensible related algorithm than the other four algorithms. In this paper they looked execution of choice (IDS) Classifiers utilizing seven segments reduce strategies.

In [10] the author presents extreme objective of intrusion detection system (IDS) change is to accomplish the most ideal accuracy for detection attacks. Different hybrid machine learning techniques were created for IDS. The centroid-based classification method is a specific hybrid learning approach that profoundly productive in the preparation and classification stages. This paper considers 60 related papers in the period somewhere in the range of 2010 and 2016 focusing on creating IDS utilizing hybrid classifiers, which 11 papers utilized centroid-based classification. Comparative examinations are looked at by the algorithm utilized in hybrid machine taking in, the dataset utilized, the foundation of the agent highlight, the phases of pre-handling data, and assessment methods considered. The achievements and confinements in creating IDSs utilizing hybrid machine learning and centroid-based classification were exhibited and talked about. A few future research openings were given that may urge intrigued scientists to work around there.

In [11] the author presents intrusion detection system assumes a critical job as it can identify diverse kinds of attacks in the network. The primary thought of the intrusion detection system is to see the pernicious attacks which scare the security from the information system's ordinary exercises. The intrusion detection system can be figured fundamentally as an issue of parallel classification, with the goal that it tends to be comprehended utilizing powerful classification technique. To correct this restriction, an altered form of SVM is presented in this work. In this work, classification is finished utilizing altered SVM and assessment of the proposed technique is finished utilizing KDD dataset by leading investigations. The exploratory outcome demonstrated that the broad time is diminished utilizing changed SVM by performing legitimate dataset.

## III. PROPOSED APPROACH FRAMEWORK AND DESIGN

### *A. Problem Definition*

The research aims at providing security for network so that it can be prevented from many attacks, particularly DoS (Denial-of-Service) or DDoS attacks which could make the server become non-functional at one stage. These attacks can be of three forms.

- Attacks exploiting some vulnerability in the software usage of an administration to cut the server down.
- Attacks that utilization up all the accessible resources on the objective machine
- Attacks that devour all the transfer speed accessible to the unfortunate casualty machine.

They generally consist of the efforts of concerned person or persons to prevent internet sites or services for their efficient functioning temporarily or indefinitely.

These attacks are a serious threat to the internet. When the interruptions go beyond a critical stage, the internet routing mechanism begins to slow down and at one stage it stops altogether responding. This is primarily due to the server's inability to trace the required source. The main objective of this examination is to find out a suitable methodology to prevent DDoS attacks and enhance the performance of the server.

### *B. Proposed Algorithm*

**Algorithm 1 Data Preprocessing**
Info: Catch Packets Data file D

Result: Data file without noise P

Step 1: Acknowledge each line of file D as T[i]

Step 2: Check if (T[i] == Delimiters) at that point Goto step 3 else go to step4 Step 3: Expel Delimiters

Step 4: Check if (T[i] == Stop Words) at that point Goto step 5 else Goto steps 6

Step 5: Expel all stop words

Step 6: if (T[i] ==URL) at that point Goto step 7 else got step 8

Step 7: Expel URL from T[i]

Step 8: if (T[i] == HTTP message) at that point Goto step 9 ...else Goto step 10
Step 9: Evacuate HTTP message
Step 10: If EOF D then Goto step 11 else Goto step 2
Step 11: Stop

**Algorithm 2: Hybrid Feature Selection Algorithm**

Input: Pre-processed data set P= {f_i,i=1,……,n}
Output:F- the Optimized features subset
Initialization: F=∅
Calculate corr(C; f_i ) for each feature, i=1,……,n
n_f=n; Select the feature f_i such that:
argmax(corr(C,f_i )),i=1,……,n_f,
Then, set P P\{f_i };C←CU{f_i }; n_f=n_f-1
while ≠∅ do
Calculate G_corr in to find f_i where i ϵ {1,2,……,n_f };
n_f= n_f-1;
P←P\ {f_i};
If (G_corr>0) then
| F←FU {f_i}.
End
End
Sort F according to the value of G_corrof each selected feature.
Return F

**Algorithm 3: Modified Naïve Bayes Algorithm**

Input:
Naïve Bayes Classifier NB and F,
Training data D (Including all normal and four attacks)
Testing data X (observed data item x)
Output:
Lx - the classification label of x
Start:
Lx = Nbclassify (NB, D, X); // Classify the test data X
If Lx == “Normal”
{
Return “No Attack is detected”;
}
Else if Lx == “dos”
{
Return “DOS Attack is detected”;
}
Else if Lx == “probe”
{
Return “Probe Attack is detected”;
}
Else if Lx == “r2l”
{
Return “R2L Attack is detected”;
}
Else if Lx == “u2r”
{
Return “U2R Attack is detected”;
}
Else if Lx == “bf”
{
Return “Brute Force Attack is detected”;
}
End
End
End
End
End
End
Stop
.

## III. RESULTS AND DISCUSSION

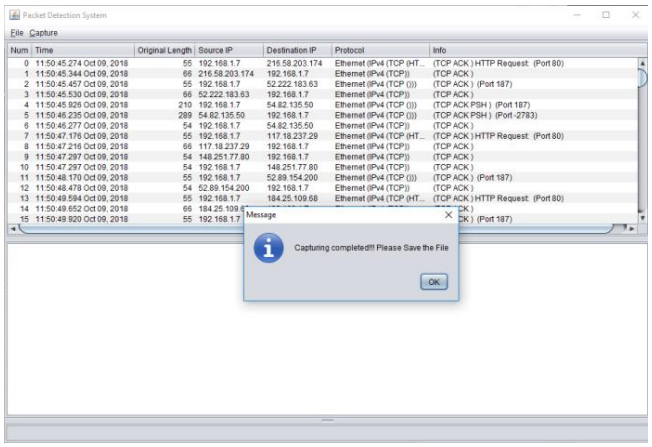


Figure 1 Packet Captured Window

In our system we capture the constant network packets utilizing the JPCAP library. The network packets are listed in the GUI and also stored in the file to execution of our system.

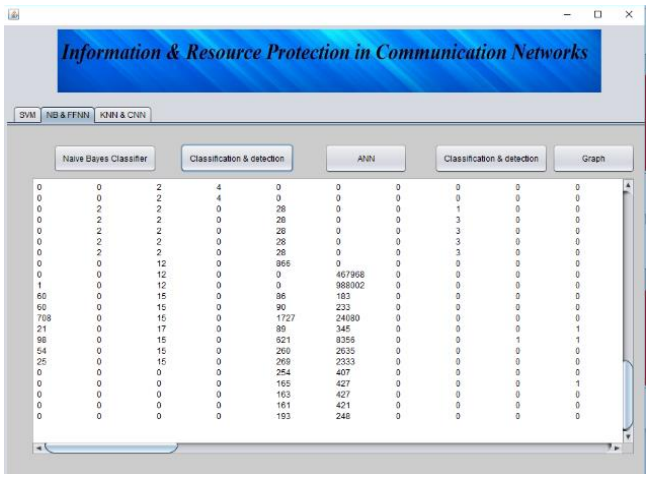


Figure 2 Naïve Bayes Classification and Detection

In this window, it shoes the Naïve Bayes techniques classification and detection results on the system GUI. In this execution we use the KDD Cup 99 dataset. We apply Noise removal and feature selection techniques on it to increase the accuracy and performance.

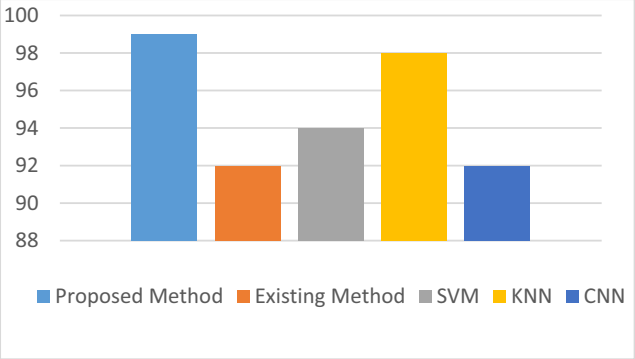


Figure 3 Accuracy Graph

In this figure we compare the some existing techniques with the proposed method. In that the proposed method is gain the better results as compare others.

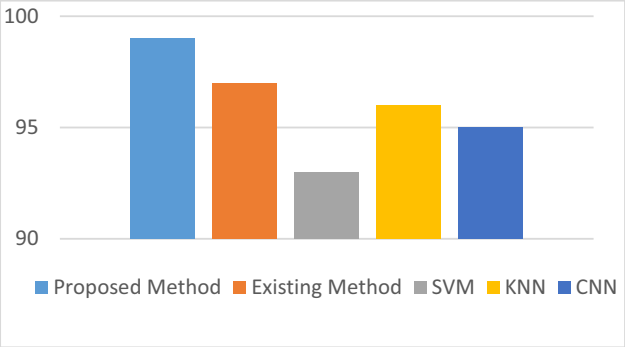


Figure 4 Precision graph

In this precision graph it show the comparison of other techniques with proposed technique.

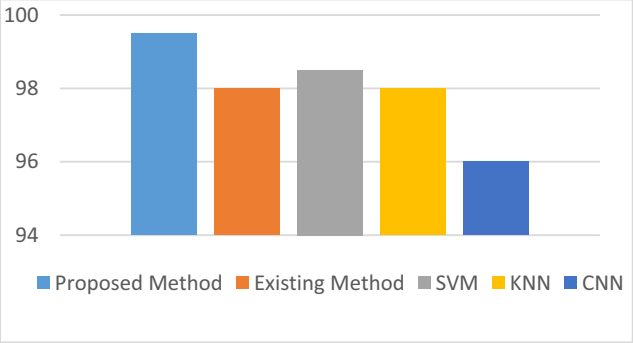


Figure 5 Recall Graph

In this Recall graph it show the comparison of other techniques with proposed technique.

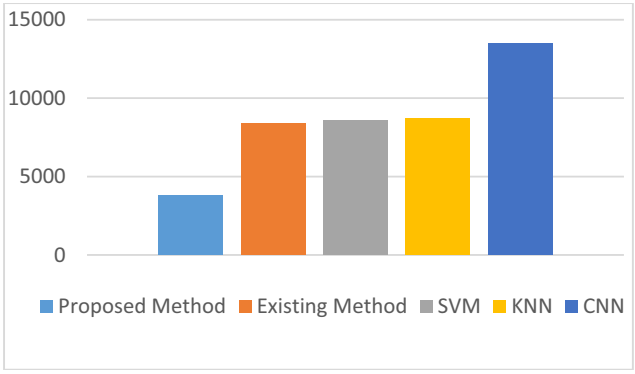


Figure 6 Time Graph

In this graph, we show the required time of execution for every technique and compare with the proposed method.

Table 1: Existing Techniques Comparison

| Method Name | Accuracy (%) |
|---|---|
| CDMT-ML | 93% |
| AIDNBC | 95% |
| NBCID | 88% |
| IDNBN | 90% |
| DMNB | 94% |
| IDS | 82% |
| MNBIDS | 97% |

In this table we compare the related existing methods and our proposed method. In that our method is generating the better results as compare to others.

## IV. CONCLUSION

In this result paper we demonstrated the Modified Naïve Bayes Intrusion Detection System (MNBIDS) to enhance the detection of DDoS attack. In that we use the Real Time Captured Network Packets and KDD Cup 99 dataset for execution. The KDD 99 dataset generated in the laboratory representing HTTP flood attacks have been used for experimentation. In this system we compare the some existing techniques with our proposed system. In our proposed system we use the Noise Removal and Feature selection methods to improve performance and then apply our proposed algorithm MNBIDS on the selected Dataset. The proposed MNBIDS system is more accurate as compare to the existing system. We compare the proposed system and existing system using accuracy, precision, recall and base on execution time of systems on same platform. The result is our proposed MNBIDS system is better as compare to others.

## REFERENCES

[1] ShyaraTaruna R, Mrs. SarojHiranwal, "Enhanced Naïve Bayes Algorithm for Intrusion Detection in Data Mining",ShyaraTaruna R et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (6) , 2013, 960-962.

[2] Dewan Md. Farid, Mohammad Zahidur Rahman, and Chowdhury Mofizur Rahman, "Adaptive Intrusion Detection based on Boosting and Naïve Bayesian Classifier", International Journal of Computer Applications (0975 – 8887) Volume 24– No.3, June 2011.

[3] Shubhangi S. Gujar and B. M. Patil, "INTRUSION DETECTION USING NAÏVE BAYES FOR REAL TIME DATA",International Journal of Advances in Engineering & Technology, May, 2014.

[4] Nahla Ben Amor, Salem Benferhat and ZiedElouedi, "Naive Bayesian Networks in Intrusion Detection Systems", InstitutSup´erieur de Gestion Tunis, 41 Av. de la libert´e, 2000 Le Bardo, Tunisie nahla.benamor@gmx.fr, zied.elouedi@gmx.fr.

[5] Mrutyunjaya Panda, Ajith Abraham and ManasRanjan Patra, "Discriminative Multinomial Naïve Bayes for Network Intrusion Detection", Department of AE&IE Gandhi Institute of Engg. and Tech. Gunupur-765022, India e-mail: mrutyunjaya74@gmail.com.

[6] Manish Kumar Nagle, and Dr.Setu Kumar Chaturved, "Feature Extraction Based Classification Technique for Intrusion Detection System",International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 8, Issue 2 (August 2013), PP. 23-38.

[7] Yogendra Kumar Jain and Upendra, "An Efficient Intrusion Detection Based on Decision Tree Classifier Using Feature Reduction", International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 8, Issue 2 (August 2013), PP. 23-38.

[8] S. Saravanan, Dr. R M. Chandrasekaran, "Intrusion Detection System using Bayesian Approach", International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 7, January 2015.

[9] Upendra and Yogendra Kumar Jain, "An Empirical Comparison and Feature ReductionPerformance Analysis of Intrusion Detection",International Journal of Control Theory and Computer Modelling (IJCTCM) Vol.2, No.1, January 2012.

[10] BambangSetiawan*, SupenoDjanali, Tohari Ahmad, "A Study on Intrusion Detection Using Centroid-Based Classification", 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia.

[11] P. Lakshmi, D. Geetha, "Intrusion Detection System Using Modified Support Vector Machine", Vol 7, No 10 (2015).