A Novel Approach for Classification of DDoS Attacks using Naive Bayes

Usham Sanjota Chanu, Khundrakpam Johnson Singh, Yambem Jina Chanu

National Institute of Technology, Manipur, India

Corresponding author: Usham Sanjota Chanu, Email: chanu06atcs012@gmail.com

One of the prevailing Internet attacks that cause havoc in society is the Distributed Denial of Service (DDoS) attack. It is human bad intention that triggers it and is one of the most discussed intrusion detection in the field of Information security and control access. Detection of DDoS attack is quite challenging and required effective classification models. Moreover, before any dataset is fed into the classification algorithm, it requires certain pre-processing to decrease the dimensions of the dataset. The original attack datasets contain features that have no or very less significance in classification. Information gain which is a feature selection algorithm is applied to decrease the dimension of the dataset which in turn helps in selection of important features. The Naïve Bayes classifier which works on the principle of bayes theorem is deployed as a classifier to the selected features to classify the class categories within a short duration with improve performance parameters.

Keywords: Naïve Bayes, Information Gain, DDoS attack, Information Security, Ranking Algorithm.

2023. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 429–436. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-40

1 Introduction

The human intensions are the lethal weapon in the cyber space. The good environment of Internet which gives enormous amount of benefits to human society can be silent battle ground for multiple intrusions. Intrusions are the illegal activities on the digital world which includes accessing other information without authorized permission. The attack dataset such as NSL-KDD which is a more filtered version of KDD dataset contain multiple network attacks and normal request. The overview of the attacks present in the NSL-KDD dataset is illustrated in Fig.1. Even though NSL-KDD may not be the perfect real time traffic, most of the researchers still prefer it for their research work until more reliable dataset come into public domain. Among these attacks the most prominent one is the DDoS attack which is a collective effort of Denial of Service (DoS) attacks.

DoS attack can be defined as the deliberate effort of a malicious client sending enormous amount of request to the target server thereby exhausting the resources to ultimately deny benign client's [1] requests. DoS attacks are executed after installing malicious or automated programs to generate that abundance of illegitimate request. Once the DoS attacks are carried out the target server exhaust its resources such as memory, bandwidth and runtime thereby making the resources unavailable to normal clients.

DDoS attacks [2] have the similar methods with that of DoS attack by deploying multiple malicious clients located in different geographical locations. It is more dangerous than the DoS attack and achieve the goal of resource exhaustion in short duration. The motives of carrying the DDoS attack are due to competitions, enmity, politics, ransom and entertainment [3]. There is a need to detect such types of attacks and prevent the target server from crashing.

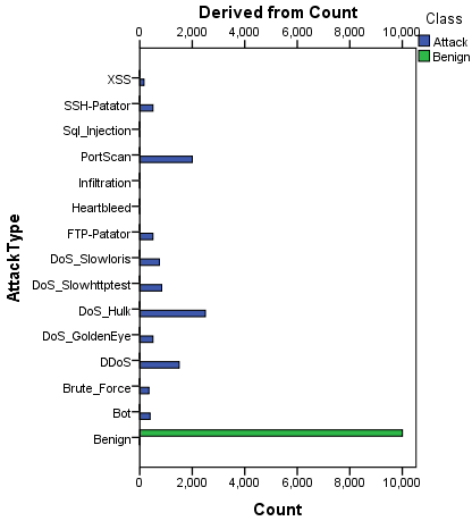

Fig. 1. Attack types in NSL-KDD dataset

In the paper, we use NSL-KDD [4] dataset and pre processed to clean the data, normalize the data and select a subset of features. This step is crucial because the noise present in the data could degrade the

performance. We then deploy Information Gain feature selection algorithm to determine a group of features that helps in depicting class categories effectively. Deploying the classification algorithm to the original dataset will consume more computation time and resources. Using all the features will also generate irrelevant class categories. In the paper, we deploy Naïve Bayes classifiers due to its property of building the faster model and detecting the class categories in short duration of time.

The overview of this paper is as follows: Section 2 contributes some of the related works. The proposed methodology with feature selection and classification model is described in Section 3. Experimental results and discussions are illustrated in Section 4. The concluding remarks and future scope are finally presented in Section 5.

2 Related Works

There are vast researches in the field of DDoS attacks which helps us to collect some of the related works that will help in our works. Some of the existing works are given below:

Saied et al. [5] proposed a supervised type of Artificial Neural Network (ANN) classification model for detection of DDoS attack. The ANN used are Feed-forward, Error Back Propagation with a Sigmoid activation function. The statistical features such as packet headers instances, source IP addresses, Identification number, and sequence numbers with both port numbers are considered as the input pattern to the model. The method can detect DDoS attack that is carried out using TCP, UDP and ICMP.

Kalkan et al. [6] proposed a technique call "Score for Core" to find out the DDoS attack packets and filtered them. This technique computes a score for the incoming traffic by comparing the normal traffic with the current traffic. Based on the computed score of the incoming traffic the attack or normal traffic was identified. The profile of the normal traffic is computed by observing the packet traffic during a normal period and similarly for attack traffic during attack period. For the detection of attack, they take into account the features such as IP address, port number, type of protocol, size of the packet, value of time to live (TTL) and TCP flag.

Gavaskar et al. [7] proposed an efficient method for detecting and mitigating against TCP SYN flood attacks using three counters algorithm. The method is able to detect spoofed IP packets up to 80%. Three counting filters are used to record the related information such as recording the first SYN packets of every connection, recording the SYN packets, whose connections have completed the three-way handshake and storing the other SYN packets

Xiao et al. [8] proposed DDoS attack detection in data center based on correlation. The technique observed the correlation information of traffic in the data center. They deploy a mechanism based on K-nearest neighbour with correlation (CKNN) and r-polling model to decrease the problem raised by the size of the training dataset. They found out that CKNN model outperforms KNN model in classifying network traffic even with a high noise signal.

Pengfule et al. [9] proposed an adaptive threshold algorithm which is capable of detecting SYN flood attack in short time for large scale network. However the method deployed has a slow detection, fast recovery mechanism. In a dual-stack firewall the attack detection and defense algorithms are implemented, the validity and performance are tested. When the firewall is under attack, the proposed algorithm improves the system efficiency substantially with minimum memory and CPU overhead.

Sivabalan and Radcliffe [10] proposed a technique that computes user's signatures by deploying Completely Automated Public Turing test to tell Computer and Human Apart (CAPTCHAs) or Are You A Human (AYAH) page. The algorithm creates a signature for every user and decides whether that user

is suspicious. Once the signature is generated, the AYAH result checks whether the signature is for an normal or attack user.

3 Proposed Methodology

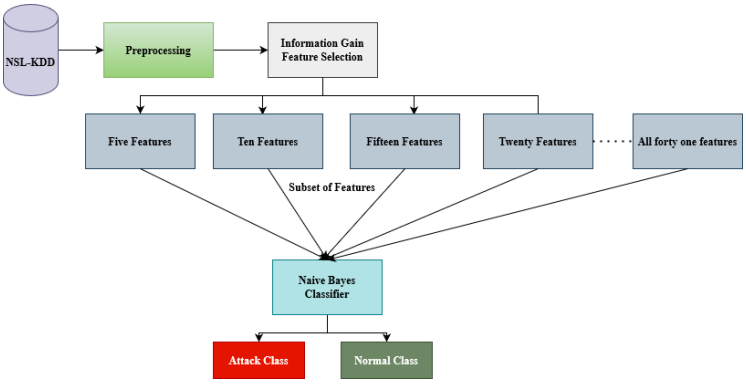

Fig. 2. Proposed method for classification

In the proposed method, we want to illustrate that not all features are important for effectively classifying the class categories. Too many features not only consume time in classification but also generate irrelevant class categories with lower accuracy.

Table 1. NSL-KDD dataset set of features

F.No	Data Feature	F.No	Data Feature	F.No	Data Feature
1	Duration	15	Su-attempted	29	Same-srv-rate
2	Protocol-type	16	Num-root	30	Diff-srv-rate
3	Service	17	Num-file-creations	31	Srv-diff-host-rate
4	Flag	18	Num-shells	32	Dst-host-count
5	Src-bytes	19	Num-access-files	33	Dst-host-srv-count
6	Dst-bytes	20	Numoutboundcmds	34	Dst-host-ame-srv-rate
7	Land	21	Is-host-login	35	Dst-host-diff-srv-rate
8	Wrong-fragment	22	Is-guest-login	36	Dst-host-same-src-port-rate
9	Urgent	23	Count	37	Dst-host-srv-diff-host-rate
10	Hot	24	Srv-count	38	Dst-host-serror-rate
11	Num-failed-logins	24	Serror-rate	39	Dst-host-srv-serror-rate
12	Logged-in	25	Srv-serror-rate	40	Dst-host-rerror-rate
13	Numcompromised	27	Rerror-rate	41	Dst-host-srv-rerror-rate
14	Root-shell	28	Srv-rerror-rate		

Fig.2 illustrates the flow diagram of the proposed method for classification of class categories. The NSL-KDD dataset is pre processed to remove certain fields with no significance such as Active _mean, Active_std, Active_Max, Active_Min, Idle_mean, Idle_std, Idle_Max and Idle_Min. After preprocessing, the features in the dataset will be like the one shown in Table 1 with F.No indicating the feature number. Table 1 show the 41 features which are left after preprocessing.

After the pre processing step, the refined dataset is now fed in a ranking algorithm (Information gain) to list down the features according to the ranks. The ranks determine the significance of the features in the detection of the attacks.

Information gain [11] is a ranking algorithm based on information. Information gain uses the concept of entropy value to calculate the distribution. The computation of entropy for a variable L is defined by equation (1).

$$P(L) = -\sum_i E(l_i) log_2 E(l_i) \quad (1)$$

where, $E(l_i)$ gives the prior probabilities of L. Entropy of L is computed with M as given by equation (2):

$$P(L/M) = -\sum_i E(l_i/m_i)\, log_2 E(l_i/m_i) \quad (2)$$

where, $E(l_i/m_i)$ is the posterior probability of L given M. The information gain is now given by equation (3).

$$\text{Information Gain} = P(L) - P(L/M) \quad (3)$$

After the deployment of the ranking algorithm, the features according to the rank are given in the Table 2 starting from most significant to least one.

Table 2 provides the rank of the features with feature number 2 as first rank and feature number 22 as last rank. To check the effectiveness of these features in classification, we tried a hit and trail method by forming sets of features. We now form multiple sets of features with first five features (2,3,4,5,6) from Table 2 as the first set. The five features that are in the top five ranks are indicated by a rectangular box with the name Five Features in Fig.2.

Table 2. Features listed according to ranks

Ranking Algorithm	Selected Feature number
Information Gain	2, 3, 4, 5, 6,10,17, 23, 24, 30, 32, 33, 1, 7, 34, 37, 8, 9, 39, 41, 11, 25, 27, 29, 31, 12, 13, 14, 35, 38, 40, 36, 28, 26, 21, 15, 16, 20, 19, 18, 22

The first ten features (2, 3, 4, 5, 6,10,17, 23, 24, 30) as second set, the first fifteen features (2, 3, 4, 5, 6,10,17, 23, 24, 30, 32, 33, 1, 7, 34) as third set, the first twenty features (2, 3, 4, 5, 6,10,17, 23, 24, 30, 32, 33, 1, 7, 34, 37, 8,9,39,41) as fourth set, first twenty five features (2, 3, 4, 5, 6,10,17, 23, 24, 30, 32, 33, 1, 7, 34, 37, 8,9,39,41, 11,25,27,29,31) as fifth set, the first thirty features (2, 3, 4, 5, 6,10,17, 23, 24, 30, 32, 33, 1, 7, 34, 37, 8,9,39,41, 11,25,27,29,31,12,13,14,35,38) as sixth set, the first thirty five features (2,3,4,5,6,10,17,23,24,30,32,33,1,7,34,37,8,9,39,41,11,25,27,29,31,12,13,14,35,38,40,36,28 ,26,21) as seventh set and the all the features (2,3,4,5,6,10,17,23,24,30,32,33,1,7,34,37,8,9,39,41,11, 25,27,29,31,12,13,14,35,38,40,36,28,26,21,15, 16,20,19,18,22) as the eight set.

Once we identify the set of features, it's time to feed the features set one after another to a classification model to compare the performance parameters. In the paper, we deploy Naïve Bayes classifier [12] as the classification model.

Naïve Bayes classifier works on the idea of conditional probability built using Bayes theorem and is given by equation (4). It is a classification model which is simple to understand and deployed for quick learning.

$$P\left(\frac{y}{X}\right) = \frac{P\left(\frac{X}{y}\right)P(y)}{P(X)} \quad (4)$$

where, the variable y is the class variable either attack or normal and X represents the features given by equation (5).

$$X = \{x1, x2, x3, \ldots\ldots, xn\} \quad (5)$$

where, x1, x2,....., xn represents the features present in the dataset. In the paper, the first set of features is represented by those five features that have the highest ranks in feature selection.

4 Experimental Results and Discussion

The results of the Naïve Bayes classifier in detecting the class category depending on various set of features are depicted in Table 3. It is noticed from Table 3 that the first set of features provides the highest accuracy with minimum computation time, minimum false positive rate and maximum detection rate.

Table 3. Performance parameter of different set of features

Feature Set	Accuracy (%)	Time (Seconds)	False Positive Rate	Detection Rate (%)
1st Set	99.6	1.6	0.65	98.5
2nd Set	97.33	3.4	1.8	96.3
3rd Set	96.03	4.8	2.6	95
4th Set	95.74	6.4	3.3	94.7
5th Set	95.44	7.8	4.4	94
6th Set	94.6	8.8	5.2	93.6
7th Set	92.57	9.2	7.3	90.4
8th Set	90.33	11.3	8.2	88.6

We now plot the Receiver Operating Characteristic (ROC) curve of the Naïve Bayes classifier with first set of feature and second set of feature as shown in Fig 3 and Fig.4 respectively to have an overview of the performance parameter.

Fig. 3. ROC Curve of the Naïve Bayes Classifier with First Set

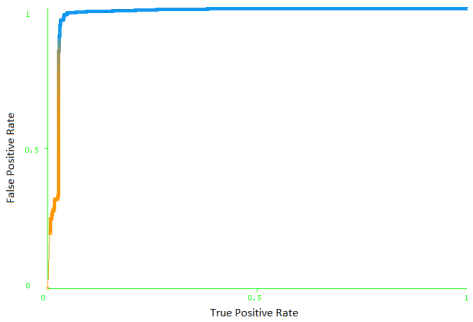

Fig .4. ROC Curve of the Naïve Bayes Classifier with Second Set

Naïve Bayes model is compared with other conventional classifiers such as K-Nearest Neighbour (KNN) [13], Decision Tree [14], Logistic Regression [15] and Support Vector Machine (SVM) [16] using first set of features. Naïve Bayes classifier model outperforms the other conventional model for the given NSL-KDD dataset as illustrated in Fig.5. Naïve Bayes have the highest accuracy and the lowest false positive rate as compared to the conventional classification models.

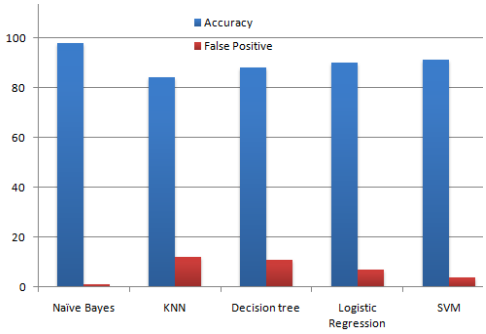

Fig. 5. Comparison of Naïve Bayes with other Classifiers

5 Conclusion

In the paper, we have achieved 99.6% accuracy, 1.6 seconds of computation time and false positive of 0.65 in detection of DDoS attacks using the information gain feature selection with top five features and Naïve Bayes classifier. It is observed that the accuracy decreases and the computation time raises if we increase the number of features. This also concludes that not all features present in the dataset have much significance in detecting the class categories but rather consume computation time and resources. It will be best way to reduce the number of features by making sure that we do not omit the effective features. In towards the future direction, we will investigate the deployment of new classification algorithm for detection and consequently improving the performance parameters.

References

[1] Singh, K. J., Thongam, K. and De, T. (2018). Detection and differentiation of application layer DDoS attack from flash events using fuzzy-GA computation. *IET Information Security*, 12: 502-512.

[2] McGregory, S. (2013). Preparing for the next DDoS attack. *Network Security*, 5: 5-6.

[3] Singh, K. J. and De, T. (2017). MLP-GA based algorithm to detect application layer DDoS attack. *Journal of Information Security and Applications*, 36: 145-153.

[4] NSL-KDD dataset. Available: http://nsl.cs.unb.ca/nsl-kdd/. Last accessed on 23/12/2020.

[5] Saied, A. et al. (2016). Detection of known and unknown DDoS attacks using Artificial Neural Networks. *Neurocomputing*, 172: 385- 393.

[6] Kalkan, K. and Alagoz, F. (2016). A distributed filtering mechanism against DDoS attacks. *Score for Core Computer Networks*, 108: 199-209.

[7] Gavaskar, S., Surendiran, R. and Ramaraj, E. (2010). Three Counter Defense Mechanism for TCP SYN Flooding Attacks. *International Journal of Computer Applications*, 6: 12-15.

[8] Xiao, P. et al. (2015). Detecting DDoS attacks against data center with correlation analysis. *Computer Communications*, 67: 66-74.

[9] Pengfule, D. et al. (2016). Detection and Defense of SYN Flood Attacks Based on Dual Stack Network Firewall. In *IEEE International Conference on Data Science in Cyberspace (DSC)*.

[10] Sivabalan, S. and Radcliffe, P. J. (2013). A novel framework to detect and block DDoS attack at the application layer. In *Proc. IEEE TENCON Spring Conference*.

[11] Sadri, A., Ren, Y. and Salim, F. D. (2017). Information gain-based metric for recognizing transitions in human activities. *Pervasive and Mobile Computing*, 38: 92-109.

[12] Sotiris, K. (2014). Integrating global and local application of Naive Bayes classifier. *International Arab Journal of Information Technology*, 11: 300-307.

[13] Yihua, L. and Rao, V. V. (2002). Use of K-Nearest Neighbor classifier for intrusion detection. *Computers & Security*, 21: 439-448.

[14] Yang, Y. and Chen, W. (2016). Taiga: performance optimization of the C4.5 decision tree construction algorithm. *Tsinghua Science and Technology*, 21: 415-425.

[15] Książek, W., Gandor, M. and Pławiak, P. (2021). Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. *Computers in Biology and Medicine*, 134.

[16] Fernando, R. et al. (2016). Comparison between Bayesian network classifiers and SVMs for semantic localization. *Expert Systems with Applications*, 64: 434-443.