Detection of DDoS Attack and Classification Using a Hybrid Approach

Suman Nandi
Computer Science and Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal
suman1nandi1@gmail.com

Santanu Phadikar
Computer Science and Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal
sphadikar@yahoo.com

Koushik Majumder
Computer Science and Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal
koushikwbutcse@gmail.com

2020 Third ISEA Conference on Security and Privacy (ISEA-ISAP) 978-1-7281-6708-4/20/$31.00 ©2020 IEEE 10.1109/ISEA-ISAP49340.2020.234999

Abstract— **In the area of cloud security, detection of DDoS attack is a challenging task such that legitimate users use the cloud resources properly. So in this paper, detection and classification of the attacking packets and normal packets are done by using various machine learning classifiers. We have selected the most relevant features from NSL KDD dataset using five (Information gain, gain ratio, chi-squared, ReliefF, and symmetrical uncertainty) commonly used feature selection methods. Now from the entire selected feature set, the most important features are selected by applying our hybrid feature selection method. Since all the anomalous instances of the dataset do not belong to DDoS category so we have separated only the DDoS packets from the dataset using the selected features. Finally, the dataset has been prepared and named as KDD DDoS dataset by considering the selected DDoS packets and normal packets. This KDD DDoS dataset has been discretized using discretize tool in weka for getting better performance. Finally, this discretize dataset has been applied on some commonly used (Naive Bayes, Bayes Net, Decision Table, J48 and Random Forest) classifiers for determining the detection rate of the classifiers. 10 fold cross validation has been used here for measuring the robustness of the system. To measure the efficiency of our hybrid feature selection method, we have also applied the same set of classifiers on the NSL KDD dataset, where it gives the best anomaly detection rate of 99.72% and average detection rate 98.47% similarly, we have applied the same set of classifiers on NSL DDoS dataset and obtain the average DDoS detection of 99.01% and the best DDoS detection rate of 99.86%. In order to compare the performance of our proposed hybrid method, we have also applied the existing feature selection methods and measured the detection rate using the same set of classifiers. Finally, we have seen that our hybrid approach for detecting the DDoS attack gives the best detection rate compared to some existing methods.**

Keywords— DDoS attack, Cloud computing, Machine learning, Weka tool, Feature selection, Cross-Validation, Classifiers.

I. Introduction

In the area of computing, cloud computing is one of the growing fields. Cloud computing allows its customers to usages the pool of hardware and software resources. The various types of resources such as storage, network, server, applications are virtualized such that several cloud users independently access the resources very easily using "pay as you use" model [1].

Cloud computing is a way such that resources are also scaled up and down according to the user's demand. The organization or any cloud user stored their data on the cloud and access those data at a low cost or free of cost. There are various organizations and various types of users are present, but their data are being stored in some storage area. The user data are stored in multiple locations by the cloud service provider (CSP) using redundant storage techniques. There are various security issues occur related to cloud data storage, also various attacks present in the cloud such that actual users are prevented from getting the cloud services efficiently. One of the major security issues in the cloud environment is the Distributed Denial of Service (DDoS) attack where the services are not available for the actual users [2].

In the Distributed Denial of Service (DDoS) attack, the server gets too many service requests from multiple systems. After getting so many requests the server becomes very busy and cannot respond to any of the service requests. Hence the resources and also network bandwidth becomes unavailable for legitimate cloud users. In DDoS attack, the attackers find out the vulnerable machine within a network and install malicious code, such that the attacking machine performs various malicious operations under the control of the attacker. The attacking machine, disturb the server by flooding the fake packets and make the server busy. So the legitimate users are prevented from getting the services properly from that server [3, 4].

In this paper, the DDoS packets have been detected and classified by using various machine learning classifiers. Several exiting DDoS detection models have been studied in section II. Section III described the collection of the dataset and the pre-processing step. Section IV presented the proposed model that will select the most relevant features from the dataset and also detected the DDoS packets by using several machine learning classifiers. Top relevant features are selected by applying existing feature selection methods as well as our hybrid feature selection methods from the NSL KDD dataset. After that, we have performed instance filtering, and prepared a new dataset and named as KDD DDoS dataset by considering only DDoS packets and normal packets. In section V, the efficiency of our hybrid feature selection method has been measured by applying the five classifiers (Naive Bayes, Bayes Net, Decision Table, J48 and Random Forest) on the NSL KDD dataset and NSL DDoS dataset. Then the performance of our proposed hybrid method has been measured and compared with existing methods using the same set of classifiers.

978-1-7281-6708-4/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.

II. Related Works

Bharot et al. [5] used Hellinger Distance (HD) function in the traffic analysis phase to calculate the difference between baseline request and the incoming request. The value of HD greater than the threshold indicates there is some attack that needs to be isolated. Then in the packet analysis phase, the most relevant and appropriate features are selected from the NSL-KDD dataset. Features are selected and ranked by calculating information gain, gain ratio, and chi-squared test, after ranking all of the selected features the final output is calculated by dividing one-third of the three filter methods. In the request classification, the legitimate and DDoS packets are classified by J48 classifier that classifies the packets with 99.67% accuracy. Now the legitimate request will be given permission to access the cloud resources and DDoS packet are transferred to the intensive care unit, where the unit trying to find the source address of the attacker.

Rawashdeh Et al. [6] implemented a model using an evolutionary neural network that's integrated with the neural network using PSO (Particle Swarm Optimization). To detect the intrusion they enhanced the performance of ANN by using PSO algorithm that determines the optimal weights of connection for the feed forward NN. The PSO maintain swarm of particle where every particle illustrates a probable solution in the entire swarm. In a multi-dimensional space, the position of the particle is modified based on pbest (personal experience), gbest (global experience) and velocity. In the training phase, the fitness function (i.e. error rate) for each particle is calculated. Based on the calculated error rate they also calculate the pbest and gbest. Until the termination criteria are not met the position as well as the velocity of the particles is accordingly updated. Whenever the termination criteria are satisfied, the weight and bias parameters (i.e. gbest) of the NN model is prepared. The proposed hypervisor-based intrusion dataset for an experiment that contains normal packet, UDP flood, and TCP SYN attacks.

The author kumar et al. [7] design a network security model that detects the DDoS attack in the application layer. For collecting the dataset they create a website and maintain a log record of attacking users as well as normal users. When the user access logs from the server then the values of the features are stored in Mysql database and converted it into csv file using Weka. They have calculated two new features, one is DT (from a particular ip address the differences of two successive time of website requests) and bts (indicate similarity as well as dissimilarity in size of byte). Using SMOTE the dataset are resample to avoid overfitting problem and they separate the dataset into 70% of training, 15% of testing and 15% of cross validation set. The instances are classified using naïve bayes technique, which produces 99% accuracy to determine the DDoS request and legitimate request.

Singh et al. [8] proposed an algorithm that selects the features very efficiently by using ensemble methods. They used 7 feature selection methods (Information gain, SVM, chi-squared, gain ratio, correlation ranking, RelieF, Symmetrical Uncertainty) and also the average of feature ranking value is calculated for each method. The threshold value is also calculated by averaging the value of 7 filtering methods. They use CAIDA 2007 dataset that contains 16 features, out of them the feature is selected if its value greater than the threshold, otherwise the feature is being dropped. After selecting the feature they use multiple classifiers using WEKA tool and also shows multilayer perceptron give the best result with 98.3% accuracy.

Sindia et al. [9] proposed a new framework that utilizes the network traffic data from their correlation features. The correlation features depend upon the variance of the entropy that is calculated between the features. The feature representative is formed by computing the variance of entropy. After that the threshold value is calculated from the median of each and every feature. In the training stage, relevant knowledge is imparted to the controller by which it can differentiate request packet and normal request. In the testing phase, the featured representative of the test sample is compared with the knowledge base by calculating the Euclidean distance. Based on this comparison classify the test samples as normal and attack scenarios. The model is developed by using CAIDA 2007 dataset and they also show that the detection rate and time are much better compared to other existing methods.

III. Dataset Collection and Pre-processing

One of the benchmark dataset available publicly is NSL KDD dataset, where user can develop and implement various IDS model by using this dataset [10]. Finding a practical dataset according to our requirement is very crucial and also creating a new dataset is a very expensive and time-consuming process. So we have used NSL KDD dataset [11], which is an advance and inherent version of kdd cup 99 dataset where the size of the NSL KDD dataset is reduced that makes the classifier easy, complete and affordable. Similarly, by using that dataset we only detect some particular types of DDoS attacks that are present on the decision class of that dataset.

The pre-processing step is very much essential for creating an efficient DDoS classification. All of the data in the dataset are not significant that makes the classifier confused such that the rate of false positive became increased. So the pre-processing step eliminates the incomplete, redundant and also missing information in the dataset.

IV. Proposed Model

The DDoS detection is the process of analyzing the network packets such that the abnormal packers are prevented from reaching the destination. At the same time, it is also very much essential for DDoS detector to permit the legitimate packets to reach their destination. So implementing an effecting mechanism is very important to classify the normal and DDoS packets.

In Fig. 1, at first, we have selected the most appropriate features from the NSL KDD dataset by using five feature selection algorithms: information gain, gain ratio, chi-squared, reliefF, and symmetrical uncertainty. Each of the feature selection algorithms selects and ranked the features from the dataset in descending order. From the ranking of each feature

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.

selection algorithm, top fifteen most relevant features are selected and then combined into a single feature set. From the entire feature set, top fourteen features are selected by using our feature selection algorithm. The NSL KDD dataset contains several anomalous instances, so we have performed instance filtering that separate DDoS instances from the anomalous instances using our selected features. Finally, we have prepared a newly updated dataset and named as KDD DDoS that only contains DDoS packets and normal packets. To getting better classification accuracy we have used the discretized filter in WEKA tool that converts actual numeric valued attributed to nominal attributes. Then we have used some commonly used machine learning classifiers (Naive Bayes, Bayes Net, Decision Table, J48 and Random Forest) to identify the detection rate of our hybrid method. K-fold cross-validation has been used here for measuring the robustness of the system. After that, we have calculated the detection rate of our selected features compared to other feature selection algorithms on NSL DDoS dataset and using the same set of classifiers.

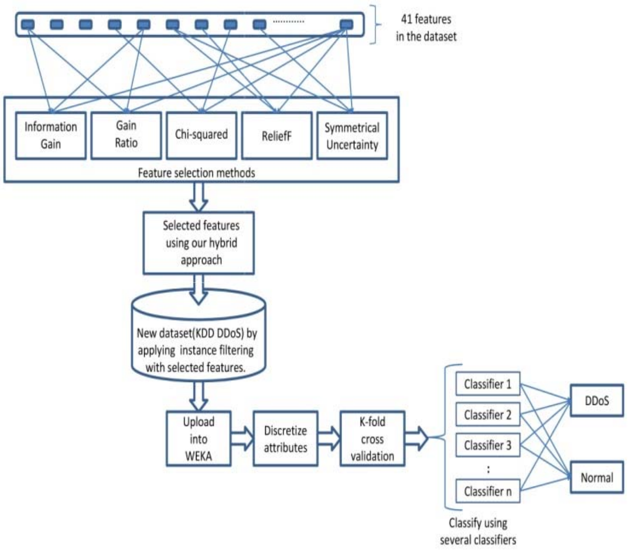

Fig. 1. Our proposed model for detection and classification of DDoS attack

A. Feature Selection

Selecting the most appropriate features from a dataset is the most challenging task. To increase the accuracy of the detection algorithm we need to select the best features very efficiently. The dataset contains many irrelevant and redundant features, so the main goal of feature selection is to remove all the irrelevant also the redundant features.

The NSL KDD dataset contains 41 distinct features, but for our DDoS detection, all of the features are not required. Our main target is to select the most important features for DDoS classifier from the dataset. There are several features selection methods are available in the WEKA tool, where each of the methods selects different subsets of features. So each of the feature selection methods selects different features set according to the ranking of the features.

-**Information gain:** This technique is used to determine the relevant features present in the dataset based on the information theory. Top features from the available dataset are selected to find the defined result with respect to the available class [12]. The highest value of information gain for a particular attribute indicates the best relevant feature and that attribute becomes the root node of the decision tree. The information gain is determined by calculating the entropy of the remaining attributes along with the target attribute [5].

-**Gain ratio:** The major drawback of information gain is when the test has many distinct outcomes i.e. information gain indicates the biasing towards the attributes with much values [13]. The gain ratio is a modified method of information gain; it normalized the outcome of information gain by divided with splitted information. So to make the decision tree, gain ratio chooses an attribute by taking size and number of branches into account [5].

-**Chi-squared:** The chi-squared test is a statistical test, used to calculate self reliability between two attributes. It determines the significant differences between the expected and observed value. Chi-squared measured the independence of any attributes with respect to the decision class where before computing the score of the features, independently assume the chi-squared score of the features and the decision class [14].

-**ReliefF:** The importance of the feature is calculated by identifying the difference of features value between the nearest neighbours. The value of the feature is observed by performing repeatedly sampling to differentiate between nearest miss and nearest hit. The value of the feature score decrease if the nearest hit is done i.e. pair of neighbour instance with identical class, similarly the value of the feature score increase if the nearest miss is done i.e. pair of neighbour instance with another class. The weight of every feature is added as per to its efficiency by using attribute evaluator to differentiate the various classes [15].

-**Symmetrical Uncertainty:** The drawback of information gain is to biasing towards the attributes with much values, this compensates is devised by the symmetrical uncertainty feature selection method by using the symmetrical property of the information gain method. To calculate the goodness of any feature, symmetrical uncertainty is calculated between that feature and the corresponding target [16].

-**Our hybrid approach:** We have seen that most of the researchers select the most relevant features without considering the occurrences as well as the rank of each feature. Our proposed method for selecting features is based on combining the output of the previous five feature selection methods. We have computed the rank of every feature from NSL KDD dataset using those five algorithms and also arranged the features in descending order of their rank. After that top fifteen features are selected from each of the five algorithms that is described in TABLE I. For each of the features, we have computed the total rank by taking the sum of ranks that are calculated by five feature selection algorithms. Then we have computed the one-fifth split of

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.

each computed rank (sum of the ranks) of each the features and assign that value in a variable P_i, where i indicate individual features i.e. $\{feature\ 1, feature\ 2, ..\}$Then we have calculated the occurrences of each feature from the combined feature set and assign the number in a variable N_i. Now the final rank R_i of each individual features is calculated by-

$$R_i = P_i \times \log \frac{N_i}{2} \quad (1)$$

Where i = Individual features present in the dataset.

$P_i = \frac{1}{5} *$ (Summation of rank for feature i, selected by five feature selection algorithms).

N_i = Occurrences of feature i.

R_i = Calculated rank of feature i.

Then we have selected those features which have calculated rank value (R_i) greater than 0. According to the value of R_i, features are shorted in descending order that is shown in TABLE I. So the log value will select those features which occur 3 or more times in the combined feature set. Now from the sorted feature, we have selected the top fourteen features that are most relevant to classify the DDoS and normal packets by any classification algorithm. We have also observed that those selected fourteen features give the best detection rate compared to any other number of features.

Algorithm 1: Feature Selection strategy

```
Input: NSL KDD dataset with 41 distinct features.
Output: Select 14 most important features.
procedure: Ranking features
    Step 1: Applying 5 feature selection methods, we have sorted the top 15 features in descending order, according to their rank.
    Step 2: Combined individuals feature ranks by several feature selection algorithms. So for each feature, we have calculated the sum of ranks that are computed by several feature selection algorithms.
    Step 3: For each feature, we have computed the one-fifth split of each computed rank (output of step 2 for each feature) and assigned the value in a variable $P_i$.
    Step 4: Then we have calculated the occurrences of each feature in the combined feature set and assigned the value in the variable $N_i$.
    Step 5: Now the final rank $R_i$ is calculated for each individual features by-
            $R_i = P_i \times \log \frac{N_i}{2}$ Where $i$ indicates individual features i.e. $\{feature\ 1, feature\ 2, ..\}$
    Step 6: According to the value of $R_i$, features are sorted in descending order and also those features are eliminated which have the calculated rank ($R_i$) value 0 or negative.
    Step 7: We have selected the top 14 important features from the output of step 6.
end procedure
```

After selecting the most relevant features from the NSL KDD dataset, the features are ranked in descending order and summarized in TABLE I. We have selected the top fifteen (15) features from the NSL KDD dataset using the five feature selection algorithms and from there top fourteen features are selected by using our hybrid approach.

TABLE I. TOP RANKING FEATURES BY APPLYING SEVERAL FEATURE SELECTION ALGORITHM

Method Used	Rank of the features (In descending order)	Number of features
Information Gain	5,3,6,4,30,29,33,34,35,38,12,39,25,23,26	15
Gain Ratio	12,26,4,25,39,30,38,6,5,39,3,37,8,33,34	15
Chi-Squared	5,3,6,4,30,29,33,34,35,12,23,38,25,39,26	15
ReliefF	3,29,38,32,33,4,23,36,40,39,34,35,31,30,24	15
Symmetrical Uncertainty	12,4,26,6,39,25,5,30,38,29,3,33,34,35,37	15
Our selected features	4,3,5,6,29,30,12,38,33,39,26,25,34,35	14

B. *Instance Filtering*

Instance filtering is one of the effective ways to make a strong and demandable dataset. After selecting the most relevant features from a dataset the size and the complexity of the dataset is reduces and we get a new normalized dataset. But according to our demand, getting a proper dataset for detecting DDoS attack is very crucial. The NSL KDD dataset contains various types of attacking and normal instances [17]. So instance filtering is very much essential such that we can separate the useful instances from a dataset to make a reliable dataset. We have performed the instances filtering using those fourteen features which are selected by our hybrid algorithm. Using those fourteen selected features, we have separated the DDoS instances from the anomalous instances on NSL KDD dataset [18]. The top selected features (feature number in NSL KDD dataset) and their description are:

- service (3): Type of services used by the network connection (ftp, http…). Depend upon the service type various attacks are done.
- flag (4): Indicate status (normal or error) of the connection. E.g. S0,S1,SF,REJ….
- src_bytes (5): Amount of data byte sends from source to destination. The data byte of attacking packets is sometimes very less (E.g. SYN flood attack) and sometimes high (E.g Ping of Depth attack).
- dst_bytes (6): Amount of data byte sends from destination to source. For any attacking packet the size this feature value is very less.
- Logged_in(12): After successful logged-in, the value of this feature is 1, otherwise 0. So for every normal packet, the value of this feature is 1.
- serror_rate (25): Percentage of connections having SYN error. So the activated flag of that packet is S0, S1 or S2 and the value of this feature is also aggregated with the number of connections to a similar host as the

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.

present in last 2 seconds. For the SYN flood attacking (Like Land, Nepture [19]) packets, flag of the packet is always S0 and the value of this feature is always high.

- srv_serror_rate(26): Percentage of connections having activated flag S0, S1 or S2 and the value of this feature is also aggregated with the number of connections to a similar service (i.e. port number) as the present in last 2 seconds. So in the attacking scenario, the flag is set as S0 and the value of this feature is high.
- same_srv_rate (29): Percentage of connections having the same services. In a DDoS attack, the value of this feature is high (>0), since the attackers are distributed in nature.
- diff_srv_rate (30): Percentage of connections having different services. In DDoS attack, the rate of different services is low for any attacking packets.
- dst_host_srv_count (33): The number of connections that uses the same service and has the same destination port. In a DDoS attack, the value of this feature is always greater than 0.
- dst_host_same_srv_rate (34): Percentage of connections using the same service and having the same destination port. In DDoS packets, the value of this feature is high (>0) because for a particular host the percentage of connection having the same services is high.
- dst_host_diff_srv_rate (35): Percentage of connections having different services on the present host. In DDoS attack, the value of this feature is very low.
- dst_host_serror_rate(38): Percentage of connections having activated flag S0, S1 or S2 and the value of this feature also aggregated with the number of connections that have similar destination addresses. For DDoS attacking packets, the value of this feature is high when the flag is set to S0.
- dst_host_srv_serror_rate (39): Percentage of connections having activated flag S0, S1 or S2 and the value of this feature also aggregated with the number of connections that uses the same service and has the same destination port. So in DDoS attacking scenario, the value of this feature is high when the flag is set to S0.

Finally, the DDoS instances are separated from the whole anomaly instances by using the above fourteen selected features. So after filtering the DDoS instances from the whole anomalous instances, the number of instances in the dataset is decreased. Now finally we got an updated dataset named as KDD DDoS which contains only DDoS instances and the normal instances. KDD DDoS dataset contains 33536 DDoS instances, and 35302 normal instances, where the normal instances are selected randomly from the entire dataset by using modulo function.

C. Discretize Attribute

To meet the requirement of several machine learning classifiers, we have reshaped our dataset. Most of the machine learning classifiers are comfortable to classify with discrete attributes. So in WEKA tool, we have used the discretized filter that converts actual numeric valued attributed to nominal attributes [20]. The advantages of using discretize filer are:

- The learning algorithm became faster and accurate.
- The number of feature values that are continuous will reduce.
- For the expert or any user, discretize of features are easier to understand.

V. Results and Analysis

We have performed the experiment using several machine learning classifiers and find the detection rate of our proposed hybrid method. The results are analyzed by using NSL KDD and KDD DDoS dataset also. Where The NSL KDD dataset contains several types of anomaly packets as well as normal packets and KDD DDoS dataset contains DDoS packets as well as normal packets. The performance of our selected features is measured by using several machine learning classifiers and k fold cross-validation in WEKA tool.

The main goal of any classifier is to analyzing several patterns and finds the comparison between new patterns with the existing patterns. Based on the analysis, the final decision will make such that the instances are classified into DDoS and normal. So after discretized dataset, we have applied some commonly used machine learning classification algorithms [21] and based on the classifiers the detection rate is calculated.

A. K-fold cross-validation

It randomly divides the entire dataset into k nearly equal size folds, where the first fold is the test set and the remaining k-1 folds are the training set. The value of k is chosen for the data samples. If the value of k is too low then the data may be highly biased that reduce the detection rate of the model. So we have chosen the value of k is 10 such that the dataset is grouped into 10 folds that produce low bias. There are 10 iterations is to be done, in each iteration, the testing fold is changed with respect to the remaining 9 training folds [22].

B. Performance Measurement

At first using the five classifiers (Bayes Net, Naive Bayes, Decision Table, J48 and Random Forest), we have calculated the detection rate of NSL KDD dataset that contains several anomaly packets. We have got the best detection rate of 99.72% and an average detection rate of 98.47%, illustrated in TABLE II. The graphical representation is shown in Fig. 2.

TABLE II. DETECTION RATE OF SEVERAL CLASSIFIERS ON NSL KDD DATASET

Classifier Used	Detection Rate
Bayes Net	97.18
Naive Bayes	97.12
Decision Table	99.54
J48	99.72
Random Forest	98.83

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.

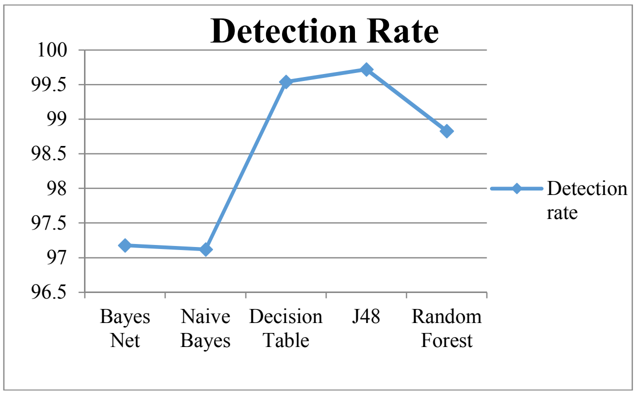

Fig. 2. Anomaly detection rate using several classifiers on NSL KDD dataset.

Now from the NSL KDD dataset, we have performed instance filtering since all the anomaly packets in the NSL KDD do not belong to DDoS categories. We have also applied the same set of classifiers on KDD DDoS dataset, which contains only DDoS packets and normal packets. We have got the best detection rate of 99.86% and an average detection rate of 99.01% that is provided in TABLE III. The graphical representation is shown in Fig. 3.

TABLE III. DETECTION RATE OF SEVERAL CLASSIFIERS ON KDD DDOS DATASET

Classifier used	Detection rate
Bayes Net	97.99
Naive Bayes	97.94
Decision Table	99.48
J48	99.79
Random Forest	99.86

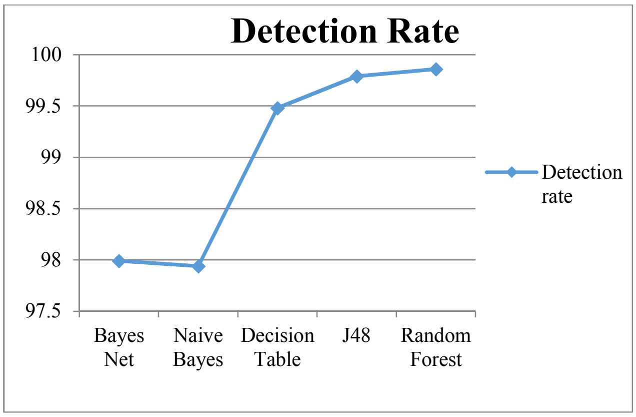

Fig. 3. DDoS detection rate using several classifiers on KDD DDoS dataset.

Now from the KDD DDoS dataset, we have again applied the five (Information Gain, Gain Ratio, Chi-Squared, ReliefF, Symmetrical Uncertainty) feature selection methods and our hybrid feature selection method, illustrated in TABLE IV. According to the rank, the features are selected and arranged in descending order. After that top relevant features are selected for detecting DDoS attack by using several feature selection methods.

TABLE IV. TOP RANKING FEATURES ON KDD DDOS DATASET

Method Used	Rank of the features (In descending order)
Information Gain	5,30,3,29,4,34,6,35,33,23,39,38,25,26,12
Gain Ratio	6,26,5,12,25,30,4,39,38,29,34,37,35,3,31
Chi-Squared	5,30,3,29,4,34,35,6,33,23,25,39,38,26,12
ReliefF	3,29,12,38,2,32,36,4,34,33,23,40,35,39,41
Symmetrical Uncertainty	6,5,30,4,26,25,12,39,38,29,34,35,3,33,23
Our selected features	4,29,3,34,12,38,5,30,35,39,6,23,25,26

Finally, the performance of our proposed hybrid method is computed by using several machine learning classifiers on the KDD DDoS dataset. In order to compare the performance of our hybrid feature selection approach, we have applied the same set of classifiers on the other five feature selection methods as well as our hybrid method, demonstrated in Fig. 4.

TABLE V. COMPARISON OF THE DETECTION RATE OF OUR HYBRID METHOD WITH OTHER FEATURE SELECTION METHODS.

Feature Selection Method	Classifiers Used					Best Detection Rate	Average Detection Rate
	Naïve Bayes	Bayes Net	Decision Table	J48	Random Forest		
Information Gain	97.26	97.31	99.33	99.77	99.85	99.85	98.70
Gain Ratio	95.68	98.17	99.38	99.76	99.85	99.85	98.56
Chi-Squared	97.26	97.31	99.33	99.77	99.85	99.85	98.70
ReliefF	94.47	94.51	98.91	99.60	99.76	99.76	97.45
Symmetrical Uncertainty	97.31	97.26	99.33	99.60	99.85	99.85	98.67
Our Hybrid method	98.00	98.06	99.38	99.81	99.87	99.87	99.02

From TABLE V, we have seen that at each classifier, our hybrid feature selection method gives the best detection rate compared to any other feature selection algorithms. The graphical representation is described in Fig. 4.

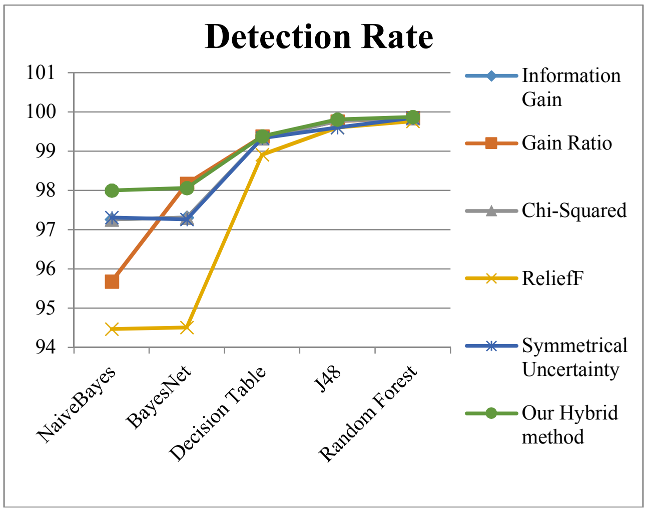

Fig. 4. DDoS Detection rate of our selected feature compared to other feature selection method on KDD DDoS dataset.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.

VI. Conclusion and Future Work

Distributed denial of service (DDoS) attack is one of the major security issues in the cloud where the resources being unavailable for legitimate users. So the detection of DDoS attack is a challenging work such that actual users are not suffering from the unavailability of resources. To detect the DDoS attack from a dataset, the most important thing is to select the appropriate features such that the attacking packets are correctly classified by any classifiers. So the effective feature selection plays a significant role to make an efficient DDoS detector. In this paper, we have used a hybrid approach that selects the top most important features from the entire feature set that are selected and ranked by five feature selection algorithm. Most of the available intrusion detection dataset contains anomalous and normal packets, but not all the anomalous instances do not belong to DDoS category. So we have performed instance filtering such that we can create a new dataset to keep only DDoS instances. We have used several classifiers from where the result shows that our hybrid approach gives the best DDoS detection rate compared to other methods.

In future work, we will try to develop a DDoS detector in a real cloud environment where we can get real traffic and we can also try to construct a prevention scheme to mitigate those real DDoS attacks.

References

[1] T. Radwan, M. A. Azer, and N. Abdelbaki, "Cloud computing security: challenges and future trends," International Journal of Computer Applications in Technology. Vol. 55, no. 2, pp. 158-172, 2017.

[2] O. Osanaiye, K. K. R. Choo, and M. Dlodlo, "Distributed denial of service (DDoS) resilience in cloud: review and conceptual cloud DDoS mitigation framework," Journal of Network and Computer Applications, vol. 67, pp. 147-165, 2016.

[3] M. Yusof, F. Mohd, and M. Drais, "Detection and defense algorithms of different types of ddos attacks," International Journal of Engineering and Technology, vol. 9, no. 5, pp. 410, 2017.

[4] T. Siva and E. S. P. Krishna, "Controlling various network based ADoS attacks in cloud computing environment: by using port hopping technique," Int. J. Eng. Trends Technol, vol. 4, no. 5, pp. 2099-2104, 2013.

[5] N. Bharot, P. Verma, S. Sharma, and V. Suraparaju, "Distributed Denial-of-Service Attack Detection and Mitigation Using Feature Selection and Intensive Care Request Processing Unit," Arabian Journal for Science and Engineering, vol. 43, no. 2, pp. 959-967, 2018.

[6] A. Rawashdeh, M. Alkasassbeh, and M. Al-Hawawreh, "An anomaly-based approach for DDoS attack detection in cloud environment," International Journal of Computer Applications in Technology, vol. 57, no 4, pp. 312-324, 2018.

[7] V. Kumar and H. Sharma, "Detection and Analysis of DDoS Attack at Application Layer Using Naïve Bayes Classifier," Journal of Computer Engineering & Technology, vol. 9, no. 3, pp. 208-217, 2018.

[8] K. J. Singh, K. Johnson, and T. De, "Efficient Classification of DDoS Attacks Using an Ensemble Feature Selection Algorithm," Journal of Intelligent Systems, vol. 29, no. 1, pp. 71-83, 2017.

[9] T. V. Sindia, and J. P. M. Dhas, "SBS-SDN based Solution for Preventing DDoS Attack in Cloud Computing Environment," vol. 12, pp. 3593-3599, 2006.

[10] S. Revathi and A. Malathi, "A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection," International Journal of Engineering Research & Technology (IJERT), vol. 2, no.12, pp. 1848-1853, 2013.

[11] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009. Accessed on: Nov. 2, 2019. [Online]. Available: https://www.unb.ca/cic/datasets/nsl.html

[12] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446-452, 2015.

[13] S. Mukherjee and N. Sharma, "Intrusion detection using naive Bayes classifier with feature reduction," Procedia Technology, vol. 4, pp. 119-128, 2012.

[14] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," Journal of King Saud University-Computer and Information Sciences, vol. 29, no. 4, pp. 462-472, 2017.

[15] H. P. Vinutha and B. Poornima, "An ensemble classifier approach on different feature selection methods for intrusion detection," Information systems design and intelligent applications, Springer, Singapore, pp. 442-451, 2018.

[16] O. Osanaiye, H. Cai, K. K. R. Choo, A. Dehghantanha, Z. Xu, M. Dlodlo, "Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing," EURASIP Journal on Wireless Communications and Networking, vol. 2016, no. 1, pp. 130, 2016.

[17] A. Harbola, J. Harbola, and K. S. Vaisla, "Improved intrusion detection in DDoS applying feature selection using rank & score of attributes in KDD-99 data set," In: 2014 International Conference on Computational Intelligence and Communication Networks, IEEE, 2014, pp. 840-845.

[18] H. Nkiama, S. Z. M. Said, and M. Saidu, "A Subset Feature Elimination Mechanism for Intrusion Detection System," International Journal of Advanced Computer Science and Applications, vol. 7, no. 4, pp. 148-157, 2016.

[19] M. Darwish, A. Ouda, and L. F. Capretz, "Cloud-based DDoS attacks and defences," In: International Conference on Information Society (i-Society 2013), IEEE, 2013, pp. 67-71.

[20] A. Rajalakshmi, R. Vinodhini, and K. F. Bibi, "Data Discretization Technique Using WEKA Tool," International Journal of Science, Engineering and Computer Technology, vol. 6, no. 8, pp. 293, 2016.

[21] O. Osanaiye, K. K. R. Choo, and M. Dlodlo, "Analysing feature selection and classification techniques for DDoS detection in cloud," in Proceedings of Southern Africa Telecommunication, 2016.

[22] M. Alkasassbeh, "An empirical evaluation for the intrusion detection features based on machine learning and feature selection methods," arXiv preprint arXiv:1712.09623, 2017.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 22:45:08 UTC from IEEE Xplore. Restrictions apply.