International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Distributed Denial-of-Service Attack by SVM and KNN using Hybrid Machine Learning Techniques

Prof. Vinod Desai, Mr. Nikhil Ambiger, Miss. Vinita Khanapuri

Department of Computer Science and Engineering, Angadi Institute of Technology and Management Belagavi, Karnataka, India

Article Info
Volume 7 Issue 4
Page Number: 231-237
Publication Issue :
July-August-2020

Article History
Accepted : 05 Aug 2020
Published : 12 Aug 2020

ABSTRACT

DDoS attacks are primary concern in internet security today. A distributed denial-of-service is a malicious attempt to disrupt normal traffic of a targeted server, service or network by overwhelming the target or its surrounding infrastructure with a flood of internet traffic. DDoS attacks achieve effectiveness by utilizing multiple compromised computer system as sources of attack traffic. DDoS attacks in ICMP protocol is called as ICMP flood attack where the attacker attempts to overwhelm a targeted device with ICMP echo-request packets, causing the target to become inaccessible to normal traffic. This type of attack is also known as Ping(ICMP) Flood attack.Fast detection of the DDoS attack, quick response mechanisms and proper mitigation must be done. An investigation has been performed on DDoS attack and it analyzes the details of its phase using machine learning technique to classify the network status. In this paper, we propose a hybrid KNN-SVM method on classifying, detecting and predicting the DDoS attack.

Keywords : Distributed denial of services (DDoS), Ping(ICMP) attack, Machine learning classifiers, Security, Intrusion detection, Prediction, support vector machine (SVM), k-nearest neighbor (KNN), KNN-SVM

I. INTRODUCTION

Distributed denial of service (DDOS) attacks have drawn extensive attention in the cyberspace during the last few years.in the recent years,the concepts and the techniques of the software defined network(SDN) have been introduced and widely researched. The DDOS attack can threaten the availability of the SDN due to the difference in the archietecture between SDN network and the traditional network. Especially the SDN network and the traditional network be affected by the DDOS attack, in general, the DDOS attack is an attempt to make the resources of a network unavailable for legitimizing users.This paper analyzes current research challenges in DDoS by evaluating machine learning algorithms for detecting and predicting DDoS attack, which includes feature extraction, classification, and clustering. Besides, various hybrid approaches have been employed. It is illustrated that these evaluation results of research challenges are mainly suitable for machine learning technique.

For classifying microarray data, one can use the classical liner discriminate analysis, artificial neural

networks, KNN, as well as some more sophisticated machine learning methodologies including bagging, boosting and kernel methods. Among them, SVM is one of the most powerful supervised learning algorithms in gene expression analysis. SVM has been found generalization ability and useful in handling classification tasks in case of the high dimensionality and sparsity of data points.

This paper is organized as follows. Section 2 provides a related study on an overview of machine learning techniques and briefly describes a number of related techniques for intrusion detection. Section 3 compares related work based on the types of classifier design, the chosen baselines, datasets used for experiments, etc. Conclusion and discussion for future research are given in Section 4.

A. OVERVIEW

Here K-Nearest Neighbor (KNN) classifier and Support Vector Machine (SVM) classifier are applied for one simple reason that the KNN is restrictive whereas SVM does not need a predefined value. If the performance of KNN applied on subset selected using Genetic Algorithm (GA) is better than that of SVM applied on subset selected using Genetic Algorithm (GA) then the performance of SVM has to be further improved. The ultimate goal is to find a better classifier than that are already used in gene data classification.

B. OBJECTIVE

The objective of the project is to apply classifiers on a subset of gene data selected using a different algorithm. The project aims at providing a classifier with:

- High Accuracy
- High Efficiency
- False rate low

II. RELATED WORK

In this section, we analyze some related work, focusing on detection of DDoS attacks. V.Deepa, K.Muthamil Sudar, P.Deepalakshmi in [1] proposed that Software Defined Network (SDN) provides a promising solution over traditional networks by decoupling the control plane and network plane. Since the controller acts as a core part of the SDN environment, there is a serious threat towards the controller in terms of security. A Distributed Denial of Service (DDoS) attack is the most potential attack in SDN environment and proposed the hybrid machine learning model to protect the controller from DDoS attacks. They showed us that this hybridiztion method is more beneficial in terms of accuracy, detection rate and less false alarm rate compared to simple machine learning models.

In [2] Shi Dong and Mudar Sarem proposed that the emerging Software Defined Networkin (SDN)provides a new way to reconsider the defense against DDoS attacks. In this paper, they had propose two methods to detect the DDoS attack in SDN. One method adopts the degree of DDoS attack to identify the DDoS attack. The other method uses the agents using software programs. The third component is the agent hosts which generate a large number of packets towards the victim host. And the fourth component is the target host which is the victim.

Yavuz Canbay and Seref Sagiroglu in [3] proposed that To over come the problem of conventional security systems such as anti-virus and firewall cannot be successful , better and more intelligent Intrusion Detection Systems(IDS) solutions are required. Here they have proposed a hybrid approach to use to detect network attacks. Genetic Algorithm (GA) and K-Nearest Neighbor (KNN) methods were combined to model and detect the attacks. This hybrid system was first applied in intrusion detection field. The system provides advantages such as,

decreasing dependency of full training data set and providing plausible solution for intrusion detection.

In [4] Dr. J.Subash Chandra Bose ,Dr. Ahmed Said Badawy,Dr. Suresh Babu Changalasetty ,Dr. Wade Ghribi ,Dr. Jamel Baili and Mr. Harun Bangali proposes an effective comparison of different algorithms for classification namely Genetic Algorithm, K-Nearest Neighbor (kNN) and Support Vector Machines (SVM) techniques. The goal of the comparison is to compare the effects of the classification rules from data. The algorithm is stimulated by the behavior of classification. The datasets that are considered can be any data involving choice of a fitness function, and an evaluation method depends on the problem metrics given to the simulation environment based on parameter values.

III. PROPOSED WORK

In order to handle this DDoS attack in ICMP protocol, we have proposed a combination of two machine learning based model with Support Vector Machine (SVM) and k-Nearest Neighbors (KNN). In this section, we discuss the details of methods that have been utilized in this work for detection and prediction of DDoS attack. There are k-nearest neighbor (KNN) and support vector machine (SVM) or known as KNN-SVM.

1. Support Vector Machine (SVM)

In this module the gene data are classified using SVM algorithm. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. Viewing input data as two sets of vectors in an ndimensional space, an SVM will construct a separating hyper plane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyper planes are constructed, one on each side of the separating hyper plane, which are "pushed up against" the two data sets. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the neighboring data points of both classes, since in general the larger the margin the lower the generalization error of the classifier. The selected features are given as an input to this module. In the case of support vector machines, suppose the incoming data belong to one of two classes, a data point is viewed as a dimensional vector (a list of p numbers).There is many hyper planes that might classify the data. The hyper plane is selected based on the distance from the hyper plane to the nearest data point. That is to say that the nearest distance between a point in one separated hyper plane and a point in the other separated hyper plane is maximized.

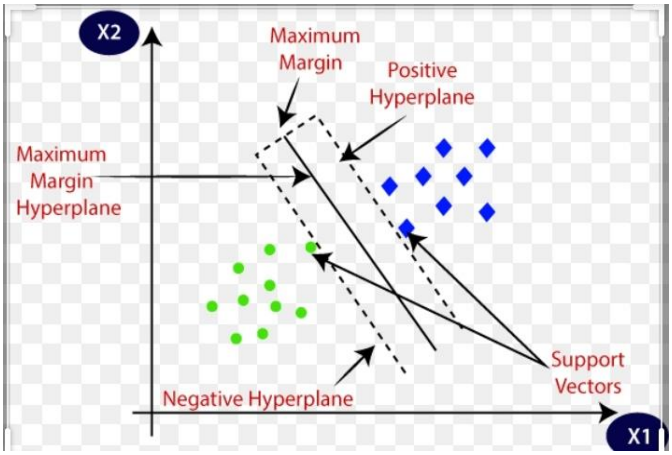

Fig 1.1 : SVM Training Examples

2. K-Nearest Neighbor (KNN)

In this module the gene data are classified using KNN algorithm. K-nearest neighbor is a supervised learning algorithm where the result of new instance query is classified based on majority of K-nearest neighbor category. K Nearest neighbor algorithm used neighborhood classification as the prediction value of the new query instance. The purpose of this algorithm is to classify a new object based on attributes and training samples. The classifiers do not use any model to fit and only based on memory. The selected features are given as an input to this module.

The K (number of nearest neighbors) values are chosen that are closest to the query point. The distance between the query-instance and all the training samples are calculated. The distance are then sorted and nearest neighbors based on the K th minimum distance is determined. The category Y of the nearest neighbors is gathered. The simple majority of the category of nearest neighbors as the prediction value of the query instance is used. Any ties can be broken at random.

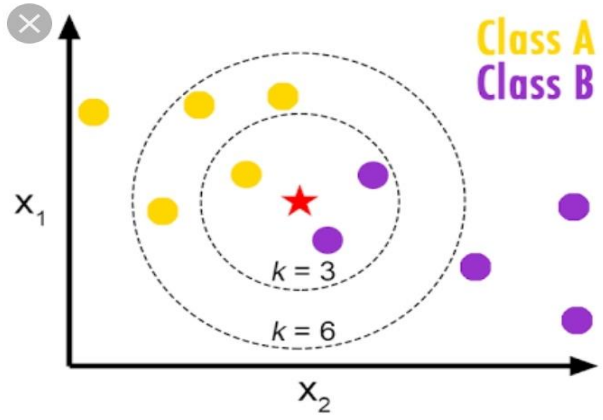

Fig 1.2 : KNN Training Examples

Work Flow Diagram

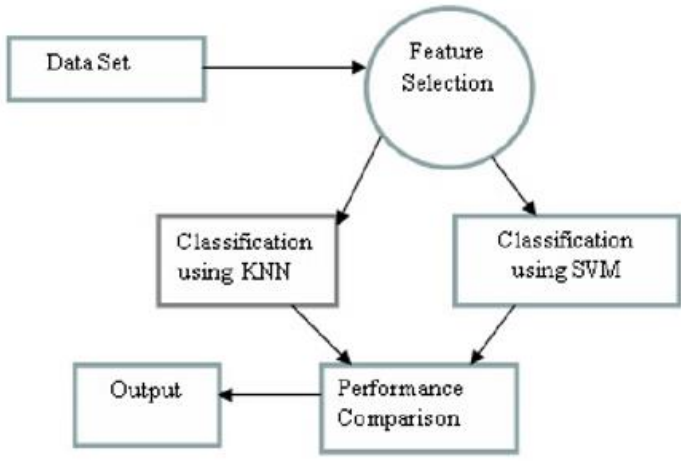

Fig 1.3 Work Flow diagram of proposed system

IV. EXPERIMENTAL SETUP

1. SVM Algorithm

The support vector machine has been chosen because it represents a framework both interesting from a machine learning perspective. A SVM is a linear or non-linear classifier, which is a mathematical function that can distinguish two different kinds of objects. These objects fall into classes, this is not to be mistaken for an implementation.

To work with SVM we use leaner kernel for implementation. In linear algebra and functional analysis, the kernel of a linear operator L is the set of all operands v for which L(v) = 0. That is, if L: V → W, then

$$\ker(L) = \{ v \in V : L(v)=0 \}$$

Where 0 denotes the null vector in W. The kernel of L is a linear subspace of the domain V.

The kernel of a linear operator Rm → Rn is the same as the null space of the corresponding n × m matrix. Sometimes the kernel of a linear operator is referred to as the null space of the operator, and the dimension of the kernel is referred to as the operator's nullity

KNN Algorithm

In pattern recognition or classification, the k-nearest neighbor algorithm is a technique for classifying objects based on closest training examples in the problem space. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification [3]. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor.

The k-NN algorithm can also be adapted for use in estimating continuous variables. One such implementation uses an inverse distance weighted average of the k-nearest multivariate neighbors.

This algorithm functions as follows :

a) Compute Euclidean or Mahalanob is distance from target plot to those that were sampled.

b) Order samples taking for account calculated distances.

c) Choose heuristically optimal k nearest neighbor based on RMSE done by cross validation technique

d) Calculate an inverse distance weighted average with the k-nearest multivariate neighbors

To analyze the performance of proposed system, we have implemented SVM, KNN along with our proposed hybrid algorithm using SVM and KNN stated in respectively to detect DDoS attack. For our simulation, we have created a custom network topology with six switches and 21 host systems as shown in using mini net, an emulation tool, which helps to create a various types of DDoS attacks are studied to select the traffic parameters that change unusually during such attacks. There are eight features extracted from both datasets using information gain rank. Then we rank all the features to identify which one is more relevant. Many machine learning problems can actually enhance their accuracy by applying features selection and extraction. This situation intensively indicates that feature selection is also important for ranking [10]. Information gain is applied to measure the importance of each feature. The information gain of a given attribute X with respect to the class Y is the reduction in uncertainty about the value of Y, after observing values of X. The uncertainty about the value of Y is measured by its entropy defined as

$$H(Y) = -\sum_i P(Y_i) \log_2((P(Y_i))$$

Where P(Yi) is the prior probabilities for all values of Y

Where P(Yi) is the prior probabilities for all values of Y. The uncertainty about the value of Y after observing values of X is given by the conditional entropy of Y given X defined as The performance of the attack detection is displayed by the detection rate (DR) and false alarm rate (FAR); the formulas are calculated as the values:

Accuracy

It is defined as how close our prediction is? For example if we can say our data set contains 10 instances and we found 9 time our prepared model provide correct target values then the accuracy is 90%. Derived using the formula

Accuracy = (correct prediction/ total supplied values) * 100

$$\mathbf{DR} = DD/(DD + DN)$$

In this formula, DD indicates that the attack flow is detected as an attack flow, and DN means that the attack flow is detected as a normal flow.

Efficiency

It is defined as time taken to build model using supplied data. Or we can simply say training time for the data model. It is defined as time required predicting values.

False rate

$$\mathbf{FAR} = FD/(FD + TN)$$

In the formula, FD means that the normal flow is detected as an attack flow, and TN indicates that the normal flow is detected as a normal flow.

III. PERFORMANCE EVALUATION

The performance of proposed hybrid algorithm is analyzed using the parameters such as accuracy, Efficiency and false alarm rate calculated using Equations 1, 2&3respectively. The test shows the efficiency as search time, accuracy, false rate for SVM, KNN, hybrid SVM-KNN. Since SVM is a

supervised machine learning model, it should be trained with labeled data only. And it is very complex to detect the new attack. In case of KNN, it is a unsupervised machine learning model, it can able to detect the new attacks. But in case of KNN, false alarm rate will be high. In order to avoid the drawbacks in SVM and KNN, we have used hybrid model.

Efficiency

Search time, Time taken to evaluate model is defined as search time. Below given values are for single value prediction in second

Table. 1 Table showing the search time of both systems

Data set size	SVM (seconds)	k- NN(seconds)
1000	.0642	.261
500	.0662	.527
200	.0642	.527
100	.0642	.229
50	.0642	.103

In our model, first the traffic is passed through the SVM module and attacks are identified. To detect the new kind of attacks, resultant traffic from SVM module is again passed through the KNN module. Once the attack is detected, particular connection will be closed and rules are updated in flow table. Our proposed hybrid model provides high accuracy, high efficiency, and low false alarm rate of 0.032

Accuracy

Accuracy After evaluation we found the following results of K-NN and SVM given in %.

Table 2. Table showing the accuracy of both systems.

Data set size	SVM	k- NN
1000	82.542	79.225
500	76.279	76.538
200	81.528	86.151
100	80.73	85.245
50	78.282	86.864

IV. EXPERIMENTAL RESULTS

Table 3 : Experimental Results

Datasize	Accuracy(%)	Efficiency(sec)	Falserate(%)
1000	89.08	0.380	0.032
500	91.10	0.321	0.030
200	92.73	0.261	0.020
100	94.87	0.223	0.013
50	97.08	0.103	0.010

Now we can see about the result how the accuracy and efficiency improved and false rate two low.

V. CONCLUSION

In this paper, we have proposed a hybrid machine learning model to detect DDoS attack in SDN environment .We also analyzed our proposed work based on the three performance metrics such as accuracy, detection rate, and false alarm rate. KNN, an unsupervised machine learning algorithm, which works well for detection of attacks compared to SVM algorithm in TCP protocol. But by using our proposed hybrid machine learning model (SVM- KNN), we have achieved more accuracy, detection rate and low false alarm rate compared to simple machine learning model. In future, we will try to implement ensemble machine learning models to detect DDoS attack in data plane by imposing the security rules in the flow table.

VI. REFERENCES

[1]. V. Deepa, K. Muthamil Sudar, P.Deepalakshmi3 et al. " Detection of DDoS Attack on SDN Control plane using Hybrid Machine Learning Techniques" Proceedings of the IEEE (ICSSIT 2018).. IEEE Xplore Part Number: CFP18P17-ART; ISBN:978-1-5386-5873-4

[2]. SHI DONG AND MUDAR SAREM " DDoS Attack Detection Method Based on Improved KNN With the Degree of DDoS Attack in

Software-Defined Networks." ReceivedDecember2,2019,acceptedDecember22 ,2019,dateofpublicationDecember30,2019,dateo fcurrentversionJanuary8,2020.

[3]. Yavuz CANBAY and Seref SAGIROGLU, "A Hybrid Method for Intrusion Detection" In IEEE 14th International Conference on Machine Learning andApplications",2015.

[4]. "Ahmed badawy,surseh babu and jamie ball " A hybrid knn/svm algorithm for classification of data"publication at researchgate.net feb 2019

[5]. Saurav Nanda, Faheem Zafari, CasimerDeCusatis, Eric Wedaa and Baijian Yang, "Predicting Network Attack Patterns in SDN using Machine Learning Approach", In IEEE Conference on Network Virtualization and Software Defined Networks (NFV-SDN),2016.

[6]. Gisung Kim, Seungmin Lee, Sehun Kim "A novel hybrid attack detection method integrating anomaly detection with misuse detection", - journal on Expert Systems with Applications – Online].

[7]. Niyaz, Quamar, Weiqing Sun, and Ahmad Y. Javaid. "A deep learning based DDoS detection system in software-defined networking (SDN)." arXiv preprint arXiv:1611.07400(2016).

[8]. Barki, Lohit, et al. "Detection of distributed denial of service attacks in software defined networks." Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on. IEEE,2016.

[9]. The most popular types of DNS attacks. Online]. Available: https://securitytrails.com/blog/most-popular-types-dns-attacks (visited on 01/06/2020).

[10]. D. Smith, Portmapper is preying on misconfigured servers to amplify attacks, Sep. 2015. Online]. Available: https://blog.radware.com/security/2015/09/port mapper-preying-on-serve rs/ (visited on 01/25/2020).

[11]. Akamai, Attackers using new MS SQL reflection techniques, Feb. 2015. Online]. Available: https://blogs.akamai.com/2015/02/plxsert-warns- of- ms- sql- reflection- attacks.html (visited on 01/08/2020).

[12]. J. M. Alonso, R. Bordon, M. Beltran, and A. Guzman, "LDAP injection techniques," in IEEE Singapore International Conference on Communication Systems, Guangzhou, China, Nov.2008, pp. 980–986.

[13]. Microsoft, MS03-034: Flaw in NetBIOS could lead to information disclosure, Sep. 2019. On-line]. Available: https://support.microsoft.com/en-us/help/824105/ms03-034-flaw-i n-netbios-could-lead-to-information-disclosure (visited on 01/16/2020).Cloudflare, NTP amplification DDoS attack. Online]. Available: https://www.cloudflare

Cite this article as :

Prof. Vinod Desai, Mr. Nikhil Ambiger, Miss. Vinita Khanapuri, " Distributed Denial-of-Service Attack by SVM and KNN using Hybrid Machine Learning Techniques, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 7, Issue 4, pp.231-237, July-August-2020. Available at doi : https://doi.org/10.32628/IJSRSET207463 Journal URL : http://ijsrset.com/IJSRSET207463