ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

A Study of DDoS Attacks Detection Using Supervised Machine Learning and a Comparative Cross-Validation

Wedad Alawad[1], Mohamed Zohdy2[2], Debatosh Debnath[3]

PhD candidate, Department of Computer Science and Engineering, Oakland University, Michigan, USA [1]

Professor, Department of Electrical and Computer Engineering, Oakland University, Michigan, USA2[2]

Associate Professor, Department of Computer Science and Engineering, Oakland University, Michigan, USA2[3]

ABSTRACT: Despite the benefits that encourage organizations to move toward the cloud, security issues are strong barriers. Distributed denial of service (DDoS) attack can target cloud computing environments and compromise the availability of cloud-based services. Thus, offensive techniques are highly recommended to detect DDoS and decrease the possibility of their success. One of the techniques used to detect such attacks is machine-learning. In this paper, the performance and detection accuracies of three supervised machine learning classifiers are compared:Naive Bayes, Decision Tree, and Linear Discriminate Analysis. The impact of the training sample size onclassifier accuracy is investigated as well. Furthermore, a novel accuracy estimation method, F-Hold Cross-Validation, is proposed and compared to the K-Fold Cross-Validation method to assess it. The results show that F-Hold Cross-Validation is time-efficient and its estimated values are acceptable.

KEYWORDS: Machine learning; security; cloud computing; Cross-Validation; supervised classifier

I. INTRODUCTION

Cloud computing, a popular service platform, provides user services in a new and feasible manner by virtualizing various resources and supplying them to customers based on their demands. One of these services is cloud-based storage; this service allows users to store their data in cloud data centers and eliminates the need for users to store data on their own computers [1]. For example, Simple Storage Service (S3) allows users to collect, store, and analyse huge amounts of data in the cloud [2]. Attractive features of cloud computing include scalability, fault-tolerance, elasticity, and pay-as-you-use. Furthermore, decreasing the cost of owning and maintaining physical networks and devices and reducing the need for additional work spaces are two benefits that give organizations, especially small ones, the confidence to move to a cloud environment [3].

Despite these appealing characteristics, cloud computing adoption is plagued by security issues [4]. The results of a questionnaire that studied the security of cloud computing show that 88.46% of college students are wary of using cloud computing services due to security issues [3]. Some of these security concerns have been discussed in [5]. Among them are cross-tenant side channel attacks, such as stealing secrets and denial-of-service DoS attacks.DoS attacks prevent legitimate users from getting services by making resources unavailable; the resources are flooded by a huge number of false requests in order to consume them. The performance of the whole system can be downgraded as well [6].

An advanced version of DoS attacksis DDoS attacks which are launched by several sources targeting the same victim. To launch DDoS attacks, the attacker first uses some scanning techniques to compromise a network of vulnerable nodes called a botnet. Then the attacker sends the DDoS attack command to a botnet and forces it to launch the attack [7]. In addition to physical bots, DDoS attacks are also launched on commodity clouds by renting many virtual machines and using them as VM bots to attack the outside world [8]. Compromised and controlled IoT devices can function as a botnet as well [9].In short, DDoS attacks are very easy to launch but extremely difficult to trace back to the real attackers [10].Fig. 1 illustrates how DDoS attacks are launched.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

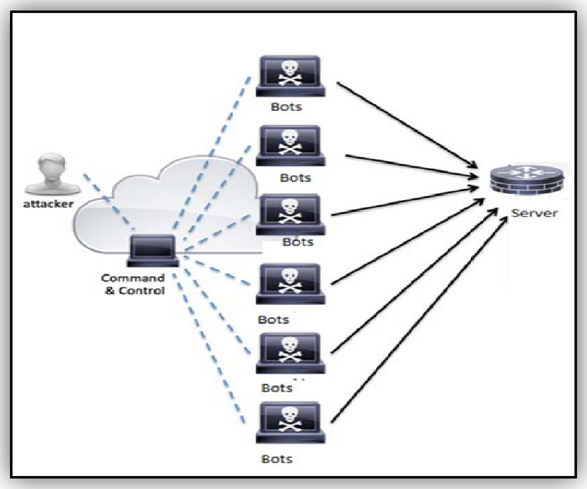

Fig. 1. Process of launching DDoS attacks [11]

DDoS is adangerous attackthat target the availabilities of network resources and services. According to a McAfee Lab report [12], DDoS are the second most frequent attacks. More than one-third of attacks in the world are DDoS attacks [10]. Furthermore, there was a 129% increase in DDoS attacks in the first quarter of 2016 as compared to the second quarter of 2015 and a 73% increase in attack size in 2016 versus 2015[13] [14]. Additionally, more than one third of government, education, and enterprise organizations were targeted by DDoS attacks in 2015. One quarter of these organizations have experienced more than ten DDoS attacks per month. Moreover, the attacks exceeded the Internet capacity of half of these organizations [15].

Even though cloud environment providers, such as Amazon EC2, have a huge pool of resources that make it unlikely to launch a successful DDoS against the cloud, cloud customers still can suffer from DDoS attacks. Cloud customers usually have two resource allocation plans: i)short-term, on-demand allocation andii) long-term allocation, in which the maximum contracted resources are made available to the customer. In the first case, the customer is exposed to Economic Denial of Sustainability (EDoS) attacks because more resources will be provided to cover the increased resource demand. In the second case, DDoS attacks could be successful because all limited allocated resources could be consumed [16].

To detect and mitigate DDoS attacks in the cloud, many strategies from different defence approaches have been presented. One promising detection approach is machine-learning-based. Getting the help of a machine's intelligence enhances analysis and detection accuracy as well as decreases detection delay.

The rest of the paper is organized as follows: Section II discusses the work related to DDoS mitigation methods, and Section III presents the research contributions and the experiments environment. Section IV discusses the classifiers' performances and detection accuracies. Section V illustrates the impact of train sample size on the model accuracy and Section VI discusses a novel cross-validation concept. Finally,the conclusion and future work are presented in Section VII.

II. RELATED WORK

Numerous studies have been done in the area of DDoS attack defense. One of these studies has investigated the capability of firewalls to mitigate DDoS attacks in the cloud [17]. This empirical study concluded that both software-based and hardware-based firewalls are not enough to defend against DDoS. Thus, more DDoS mitigation strategies are required. Some of them have been presented in a previous paper [17]. Another strategy that can be applied in the cloud environment to beat DDoS attacks that target individual cloud computing environment consumers has been presented in [16]. This strategy depends on allocating resources dynamically. When DDoS attacks are detected, customers are given additional intrusion prevention servers (IPS) to mitigate the attack. These extra resources are returned to the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

available resource pool when the attack ends. One more method that has been used to detect attacks is entropy-based. In [18] entropies of some selected meaningful traffic features are measured to detect DDoS attacks. Furthermore, Snort,which is a signature-based detection method, is effective to detect known attacks, but it is less so when it comes to a new attack because the signature was unknown when the attack happened [19].

In addition to the aforementioned defense strategies, anomaly-based methods are considered strong approaches to detect DDoS attacks.In [20], many statistic-based detection algorithms have been studied to detect SYN flooding DoS attacks.Some data mining-based DDoS detection approaches have been explored in [21] as well.The performance of several supervised and unsupervised machine learning algorithms in detecting DDoS attacks are evaluated in [10]. Furthermore, the use of semi-supervised algorithms to enhance the classifier's intrusion detection performance are discussed in [22]. Authors in [10] have proposed a machine learning-based DDoS attack defense mechanism that is based on analyzing the gathered information from servers' hypervisors and virtual machines. Their method is applied close to the attacker location in the cloud environment. In fact, Neural Networks algorithms are used in several DDoS detection mechanisms. In [23], a hybrid neural network technique that archives high accuracy in detecting DDoS attacks was proposed. A Multi-Layer Perceptron Neural Network was also selected as a base for attack detection methods in [24] [25]. Furthermore, NIDS, which is an attack classification method and uses a 2-layered feed-forward neural network,is presented in [26] and has been deemed accurate. In addition to the mentioned algorithms, a Radial-basis-function Neural Network is the core of other DDoS detection mechanisms [27] [28]. Moreover, time Delay Neural Networks have been used to defend against DDoS attacks as well [29]. Furthermore,Self-Organizing Feature Map (SOFM) algorithmscan be applied to enhance attacks classificationaccuracy [30] [31].

Like Neural networks, Naive Bayes algorithms are also used to present accurate defense techniques against network attacks [32]. Furthermore, decision trees are used in many methods to detect attacks. ENDER is a mechanism that applies a decision tree algorithm to detect HX-DoS attacks that combine HTTP and XML messages to target cloud services [33]. Besides utilizing one supervised machine learning classifier to provide network attack defense mechanisms, multi classifiers are combined in one attack recognition method to enhance detection accuracy [34] [35].

Various studies have evaluated different machine learning classifiers based on their performance in detecting DDoS attacks. Some of them have compared classifiers that belong to many machine learning algorithm types, while other research focused on classifiers located under one machine learning algorithm type. The NSL-KDD dataset was used to compare C4.5, Naive Bayes, Multilayer Perceptron, SVM and PART classifier models in [36] and BayesNet, Logistic, IBk, JRip, PART, J48, RandomForest, RandomTree and REPTree in [37]. Additionally, the KDD99 dataset was used to evaluate neural networks and decision trees in [38] and RBP, SVM, K-Nearest Neighbor, Decision Tree, and K-Means techniques in [35]. Furthermore, the CAIDA dataset was used to compare Naive Bayes, C4.5, SVM, KNN, K-means and Fuzzy c-means in [39]. In addition to CAIDA, the DARPA scenario specific dataset and CAIDA Conficker datasets were used to evaluate Naive Bayes, Multi-Layer Perceptron, IBK, RBF network, Bayesnet, J48, Bagging+Random Forest, Voting, Random Forest, and Adaboost+Random Forest in [40]. Moreover, authors in [41] evaluated Multilayer Perceptron, Random Forest, and Naive Bayes by using their own generated data. From the perspective of the same class, BP neural network and LVQ neural network have been evaluated in [42]. Another group of studies have assessed ensemble methods that combine different machine learning classifiers either from the same class or different classes. Unlike [34] that evaluated ensembles of only neuro-fuzzy classifiers, [43] compares ensembles of GA with SVM and GA with ANN.In our paper, three datasets are used to evaluate Linear Discriminant Analysis (LDA), Naive Bayes (NB), and Decision Tree(DT) in their ability to detect DDoS attacks.Moreover, a comprehensive study of existing DDoS attack defense mechanisms has been done,and the authors advocate for the creation of comprehensive, collaborative, and distributed defense mechanisms [44].

III. CONTRIBUTIONS AND EXPERIMENTAL SETUP

A. *Contributions*

In this paper we offer the following contributions:

- The abilities of Naive Bayes, Decision Tree, and Linear Discriminate Analysis in detecting DDoS attacks were investigatedempirically using the Anaconda platform and the Scikit-learn machine learning library. Their detection precisions and alsotheir trainingand testing times were examined.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

- The effect of the training dataset size on the classifiers performanceswasstudied by using the whole KDD99 dataset and only 2.6% of the KDD99 dataset.
- A new resampling method (F-Hold Cross-Validation)was proposed and compared with (K-Fold Cross-Validation).

B. *Experimental Setup*

The details of the system, datasets, and machine learning algorithms that have been usedto achieve our contributions are as follows:

- Operating System: Windows 10 Enterprise, 64-bit
- Processor: Intel (R) Core (TM) i7 -7700 CPU @3.60GHz
- RAM memory: 32.0 GB
- Data science platform: Anaconda Distribution [45]
- Python IDE: Spyder [46]
- Machine Learning Library: Scikit-learn [47]
- Supervised machine learning classifiers: DecisionTreeClassifier, D LinearDiscriminantAnalysis (LDA) and GaussianNB
- Datasets:(Table 1 contains the details of datasets used)

Table 1. Datasets used in our study

Dataset name	Training Dataset		Testing dataset		Target	Dataset collection date
	Size	No. samples	Size	No. samples		
Kddcup99 [48]	743M	4898431	133 MB	311029	Different types of DDoS attacks	1999
NSL-KDD [49]	18.2 MB	125973	3.21 MB	22544	Normal or Anomaly	2009
2.6% of Kddcup99	19.3M	125973	133 MB	311029	Different types of DDoS attacks	1999

IV. EXAMINATION OF SUPERVISED MACHINE LEARNING ALGORITHMS

Many research papers have analyzed the performance of different machine learning algorithms in detecting DDoS. However,no previous paper to our knowledge has compared all three selected classifiers on both KDD99 and NSL-KDD datasets. In most of the comparison studies, the classifiers are compared after tuning their parameters and selecting features. This research takes into account whether these approaches are not optimal for some classifiers. In other words,does the evaluation of only optimized classifiers give a fair picture of the classifiers performance? Thus, in this study, using three datasets, we compared the classifiers in three stages.

For each dataset, the algorithms are compared three times. First, they are compared using the default algorithms' parameters and all dataset features. Second, they are compared after selecting the optimal features. Third, they are compared after selecting the optimal features and tuning parameters. The algorithm comparison matrices that have been used are precision, which is the ratio of true positive instances among all positive instances, and model training and testing computation times.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

A. *Algorithm comparison results*

Stage 1: When the default algorithms' parameters and all datasets features were used

Table 2. Using default algorithms' parameters and all datasets features to compare algorithms

	Part of KDD Dataset			KDD99 Dataset			NSL-KDD Dataset		
	Precision	Training Time	Testing Time	Precision	Training Time	Testing Time	precision	Training Time	Testing Time
Decision Tree	0.71846	0.28979	0.11360	0.898595	22.68869	0.09545	0.778877	0.811543	0.007240
Naive Bayes	0.53913	0.19333	0.66109	0.767031	10.32593	2.37736	0.450319	0.211356	0.017478
LDA	0.67101	0.45685	0.07194	0.814249	25.03510	0.13328	0.734697	0.505615	0.005164

Stage 2: When the default algorithms' parameters and selected features were used

Table 3.Using default algorithms' parameters and selected features to compare algorithms

	Part of KDD Dataset			KDD99 Dataset			NSL-KDD Dataset		
	Precision	Training Time	Testing Time	Precision	Training Time	Testing Time	precision	Training Time	Testing Time
Decision Tree	0.718078	0.181601	0.01924	0.897543	9.2234854	0.038119	0.797685	0.2310613	0.0032517
Naive Bayes	0.728842	0.115567	0.09687	0.843847	7.3921531	0.349810	0.450364	0.1435950	0.0039483
LDA	0.791048	0.249586	0.03511	0.795562	15.626319	0.075959	0.770316	0.2807176	0.0005637

Stage 3: When selected parameters and features were used

Table 4.Using selected algorithms' parameters and datasetsfeatures to compare algorithms

	Part of KDD Dataset			KDD99 Dataset			NSL-KDD Dataset		
	Precision	Training Time	Testing Time	Precision	Training Time	Testing Time	precision	Training Time	Testing Time
Decision Tree	0.718120	0.196144	0.02221	0.898235	8.6456474	0.037390	0.815339	0.1872117	0.0026353
Naive Bayes	0.728842	0.120144	0.09767	0.843847	7.5373749	0.361617	0.450364	0.1319760	0.0016077
LDA	0.614811	0.289852	0.03254	0.801559	17.419728	0.072209	0.771159	0.3120263	0.0005555

B. *Comparisons Results Discussion*

Overall, results show an obvious decrease in classifiers' training and testing times when tuning their parameters and selecting appropriate features. Furthermore, these operations improve the detection precision of these classifiers in most cases. Among all three classifiers, the parameter optimized Decision Tree with the optimum set of features is the most suitable classifier to detect attacks with a reasonable training time. Its precision reaches to 0.898235, and it needs 8.65

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

second training time for the very big dataset KDD99. Even without optimizing the Decision Tree classifier's parameters and selecting features, it still gives very satisfactory precision (0.898595) but it takes more training and testing times, 22.689 seconds and 0.09545 seconds respectively.

Even though Naive Bayes gives low precision in all dataset in the first stage, the appropriate selection of features can enhance its precision incredibly. Without feature selection, the precisionfor part of KDD dataset and KDD99 dataset is 0.53913 and 0.767031 respectively. After the feature selection process, their accuracies jumped to 0.728842 and 0.843847 respectively. However, its accuracy for NSL-KDD is still low even after tuning its parameters and selecting features (0.450).

Still, LDA gives acceptable precisions compared to Naive Bayes on the NSL-KDD dataset and when default features and parameters are applied. On NSL-KDD, the precision of LDA is 0.771159 and Naive Bayes precision is just 0.450364.

In general, the results illustrate that the Decision Tree gives the best precision for both the KDD99 and the NSL-KDD datasets in all stages. However, Naive Bayes and LDA overcome the Decision Tree when only 2.6% of KD99 dataset is used. Additionally, classifiers are fitmore accurately when the whole KDD99 dataset is used. For the two other datasets, the results differ based on the classifier used. From the training and test times, the processes of feature selection and parameter tuning decrease these times in a noticeable manner while enhancing detection accuracy in most cases.

V. The Effect of Training Dataset Size on classifier Performance

In this paper, we studied the impact of training dataset size on the detection accuracy of three machine learning classifiers. The algorithms were trained on the whole KDD training dataset and on just 2.6% of the same KDD training dataset. Then the trained models were tested on the same unseen data. The results show that when more data was used, the more accuracy obtained. In all stages, the precisions of all three classifiers on the 2.6% of the KDD dataset are lower than the precisions when those classifiers are applied on the whole KDD datasets. See Fig. 2, Fig. 3, and Fig. 4.The difference between their accuracies reaches to almost 0.23 as in the case of the Naive Bayes' application because the precision on the part of KDD is 0.53913 whereas its precision on the whole dataset is 0.767031 as shown inTable 2.

Fig. 2.The precision of Decision Tree classifier when applied on KDD99 Dataset and 2.6% of it

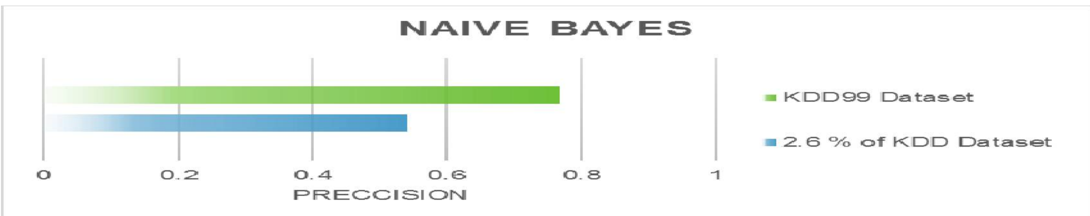

Fig. 3. The precision of Naive Bayes classifier when applied on KDD99 Dataset and 2.6% of it

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Fig. 4. The precision of Linear Discriminant Analysis classifier when applied on KDD99 Dataset and 2.6% of it

VI. F-HOLD CROSS-VALIDATION

A. *F-Hold Cross-Validation Concept*

Fitting the model to the training data beyond required level leads to poor performance when testing the model on unseen data. Thus, it is important to ensure that the model can generalize well [50]. Well-known techniques to avoid over-fitting are resampling methods, such as K-Fold Cross-Validation [51]. In K-Fold Cross-Validation, the training data is divided into K folds to be used in K training and validation iterations. K-1 chunks are used for training,and the remaining portion is used for testing. The test segment differs in each iteration [52].In this paper a novel cross-validation method, F-Hold Cross-Validation, has been proposed and compared to K-Fold Cross-Validation.

F-Hold Cross-Validation follows the same procedure of K-Fold Cross-Validationexcept that in each iteration the data is divided into threesections (Train, Test, and Hold). The classifier is trained using the Train and evaluated using the Test. The Hold part is not used in each iteration. The theory behind F-Hold Cross-Validation is thatnoise data might be part of the Hold chucks.Therefore, when Hold chunks are not used in the training process, the model may generalize well.

B. *F-Hold Cross-Validation Vs. K-Fold Cross-Validation*

F-Hold Cross-Validation has been implemented and compared to K-Fold Cross-Validation from the perspectives of estimated accuracy and cross-validation computation time. In the evaluation process, K and F values have been set to 5,8,10, and 15, and parameter tuning and feature selection have been done before starting the evaluation process. The estimated precisions and cross-validationcomputation times of K-Fold and F-Hold Cross-Validation methods on different K and F values are illustrated in Fig.5 to Fig.13.

A comparison study of three accuracy estimation methods, hold out, bootstrap, and cross-validation, using C4.5 and NB was achieved in [53]. Because the results of that research stated that 10-Fold Cross-Validation is recommended for model selection, we will focus first on the accuracy estimates of 10-Fold Cross-Validationand 10-Hold Cross-Validation in our experiment and then compare the other values. As Fig.5 to Fig. 13 display, both 10-Fold and 10-Hold Cross-Validation methods give almost close estimate values and the difference does not exceed 0.1 as in Fig. 8, whereas F-Hold Cross-Validationovercomes K-Fold Cross-Validation computation time and the difference reaches to 18 seconds as in Fig. 10. In general, there is a small discernable differencein accuracy estimates between K-Fold Cross-Validationand F-Hold Cross-Validation when K and F values are 8, 10 and 15. The difference does not exceed 0.07 as in the case of K=8. See Fig.8. However, the computation times of F-Hold Cross-Validation are always less than the computation times of K-Fold Cross-Validation. SeeFig.5 to Fig. 13. Therefore, we can use 10-Hold Cross-Validation to decrease the computation time while still getting close accuracy estimate values.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

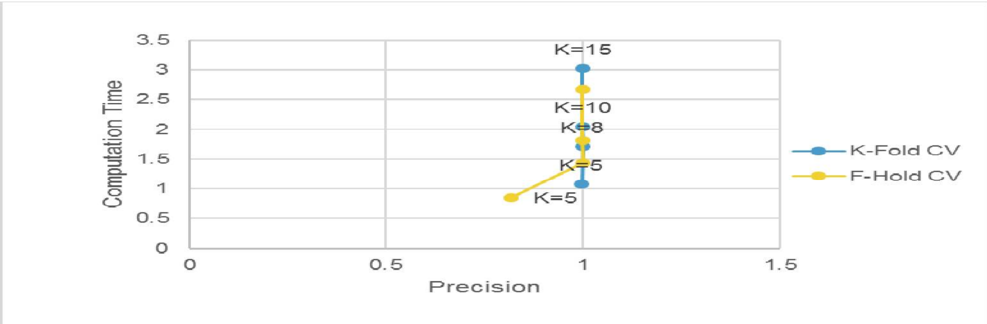

Fig. 5. Computation times andestimated precisions of K-Fold and F-Hold when DT applies on 2.6% of KDD

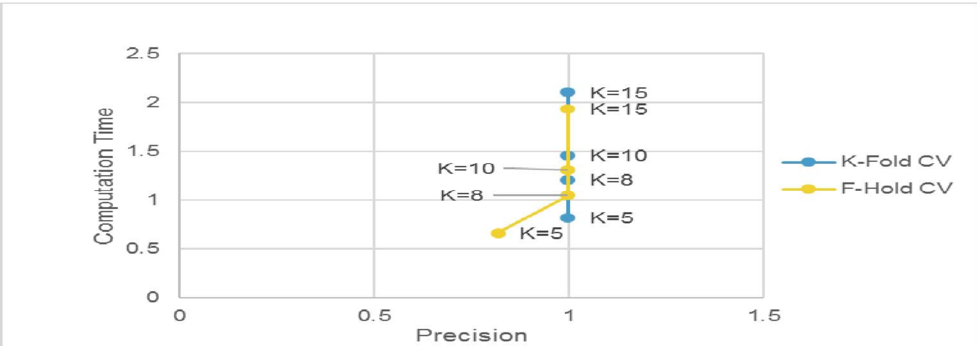

Fig. 6.Computation times and estimated precisions of K-Fold and F-Hold when NB applies on 2.6% of KDD

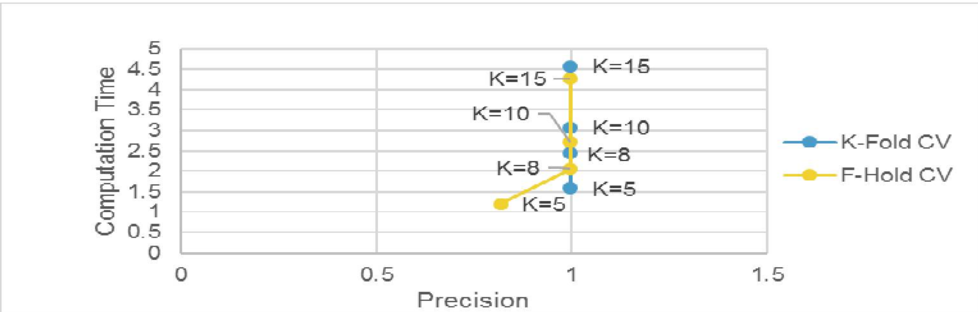

Fig. 7.Computation times and estimated precisions of K-Fold and F-Hold when LDA applies on 2.6% of KDD

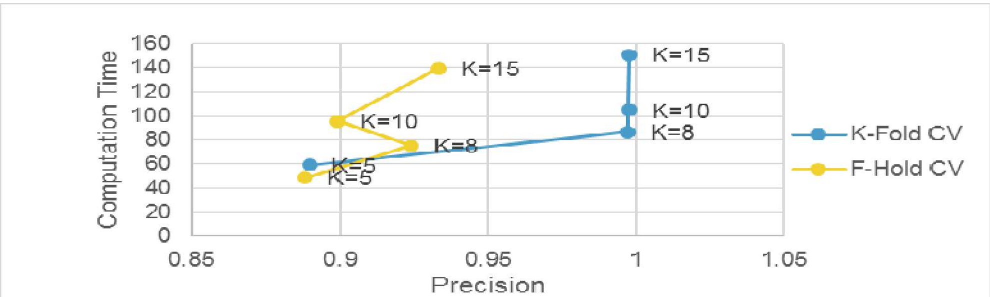

Fig. 8. Computation times and estimated precisions of K-Fold and F-Hold when DT applies on KDD

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

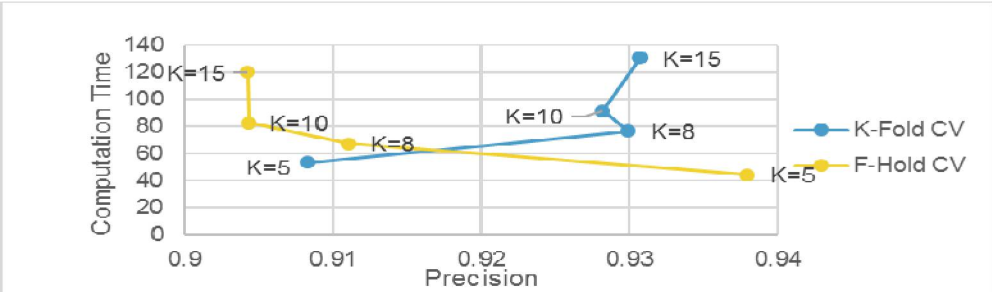

Fig. 9. Computation times and estimated precisions of K-Fold and F-Hold when NB applies on KDD

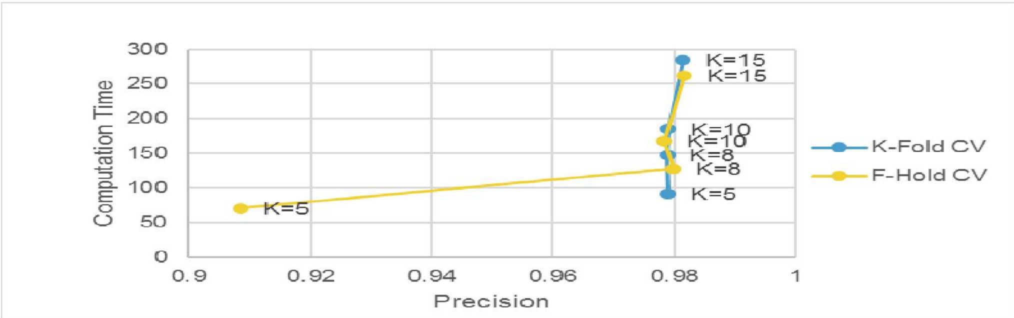

Fig. 10. Computation times and estimated precisions of K-Fold and F-Hold when LDA applies on KDD

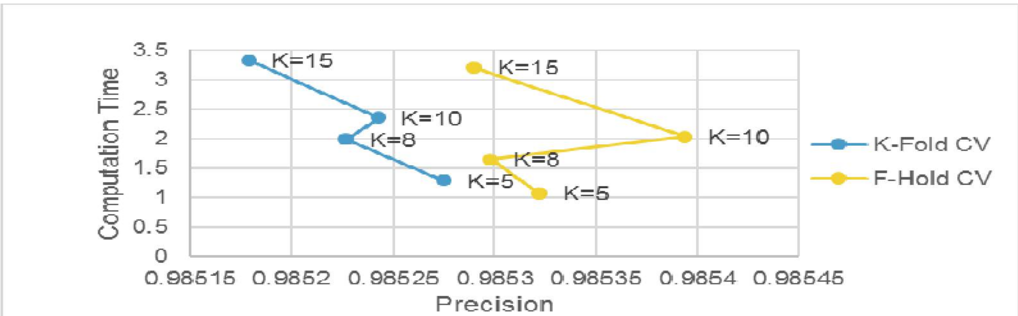

Fig. 11. Computation times and estimated precisions of K-Fold and F-Hold when DT applies on NSL-KDD

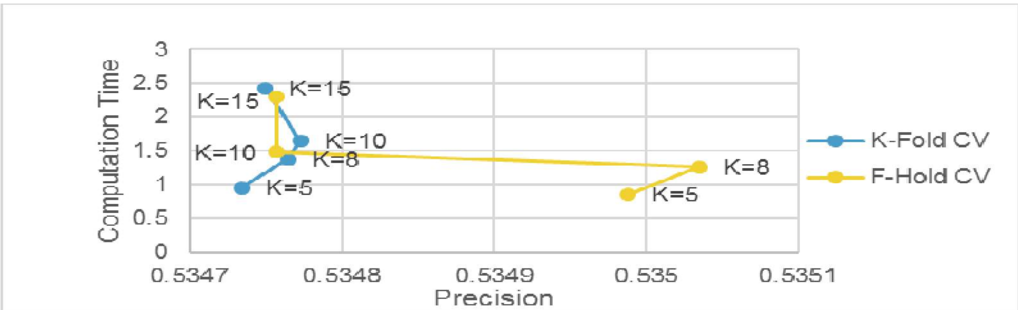

Fig. 12. Computation times and estimated precisions of K-Fold and F-Hold when NB applies on NSL-KDD

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

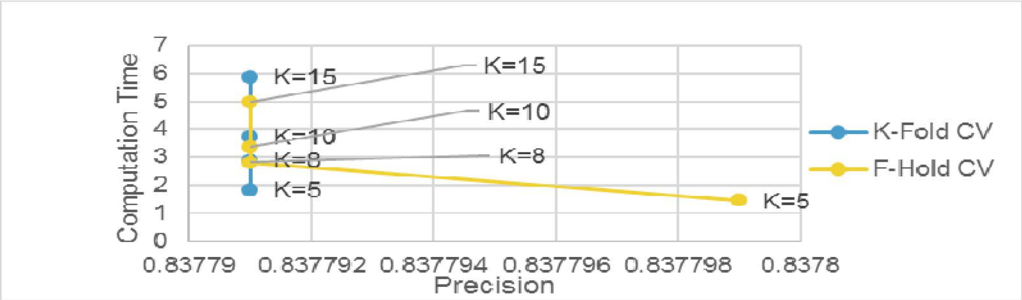

Fig. 13. Computation times and estimated precisions of K-Fold and F-Hold when LDA applies on NSL-KDD

VII. Conclusion

Machine learning-based DDoS attack detection methods formulated through the supervised algorithms are considered effective DDoS attack defence methods. Thus, applying them in the cloud is a promising solution for potential compromisesin cloud services. The results of the comparison study of three supervised machine learning classifiers, Naive Bayes, Decision Tree, and Linear Discriminant Analysissuggests that the Decision Tree classifier provides better defence against DDoS attacksthan the two other classifiers. Additionally, this research illustrates that big train datasets (Approximately four million instances) can fit the classifier more perfectly than small datasets (Approximately 130,000 instances). Furthermore, the F-Hold Cross-Validation is proposed as a time-efficient model selection method. Further analysis of F-Hold Cross-Validation will be done in the future, and it will be applied on a different research area as well.

REFERENCES

1. Zheng Yan, Xueyun Li, Mingjun Wang, and Athanasios Vasilakos. "Flexible data access control based on trust and reputation in cloud computing." IEEE Transactions on Cloud Computing, 2017.
2. Amazon Web Services, Inc. Amazon Simple Storage Service (S3) — Cloud Storage — AWS. [online] Available at: https://aws.amazon.com/s3/ [Accessed 17 Sep. 2017].
3. Mehul Nanda, Aakarsh Tyagi, Karan Saxena, and Neeru Chauhan. "Hindrances in the security of Cloud Computing." 7th International Conference of Cloud System and Big Data Engineering, pp. 193-198. IEEE, 2016.
4. Mohamed Almorsy, John Grundy, and Ingo Müller. "An analysis of the cloud computing security problem." arXiv preprint arXiv:1609.01107, 2016.
5. Jay Aikat, Aditya Akella, Jeffrey S. Chase, Ari Juels, Michael K. Reiter, Thomas Ristenpart, Vyas Sekar, and Michael Swift. "Rethinking Security in the Era of Cloud Computing." IEEE Security & Privacy 15, no. 3, pp. 60-69, 2017.
6. Kai Zhao, and Lina Ge. "A survey on the internet of things security." 9th International Conference on Computational Intelligence and Security (CIS), pp. 663-667. IEEE, 2013.
7. Jesus Gonzalez, David Terrazas, and Witold Kinsner. "Zero-crossing analysis of Lévy walks for real-time feature extraction: Composite signal analysis for strengthening the IoT against DDoS attacks." 15th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), IEEE, pp. 143-153, 2016.
8. Rui Miao, Rahul Potharaju, Minlan Yu, and Navendu Jain. "The dark menace: Characterizing network-based attacks in the cloud." Proc of the 2015 ACM Conference on Internet Measurement Conference, ACM, pp. 169-182, 2015.
9. Gang Gan, Zeyong Lu, and Jun Jiang. "Internet of things security analysis." International Conference on Internet Technology and Applications (iTAP), IEEE, pp. 1-4, 2011.
10. Zecheng He, Tianwei Zhang, and Ruby B. Lee. "Machine Learning Based DDoS Attack Detection from Source Side in Cloud." 4th International Conference on Cyber Security and Cloud Computing (CSCloud), IEEE, pp. 114-120, 2017.
11. G. Aline Sophia, and Meera Gandhi. "Stealthy DDoS detecting mechanism for cloud resilience system." International Conference on Information Communication and Embedded Systems (ICICES), IEEE, pp. 1-5, 2017.
12. Anonymous "McAfee Labs threats report," McAfee Inc., Santa Clara, CA.Available: https:// Www.Mcafee.Com/Us/Resources/Reports/Rp-Quarterly-Threats-Mar-2017.Pdf, pp. 1-49, 2017.
13. Arbornetworks.com. 2016 DDoS Attack Statistics | Arbor Networks®. [online] Available at: https://www.arbornetworks.com/arbor-networks-releases-global-ddos-attack-data-for-1h-2016 [Accessed 17 Sep. 2017].
14. Cloud and IoT Threats Predictions. [ebook] McAfee Labs. Available at: https:// www.mcafee.com/us/resources/misc/infographic-cloud-iot-predictions-2017.pdf [Accessed 17 Sep. 2017].
15. Worldwide infrastructure security report. [ebook] arbornetworks.com: Arbor Networks. Available at: https://www.arbornetworks.com/images/documents/WISR2016_EN_Web.pdf [Accessed 17 Sep. 2017].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

16. Shui Yu, Yonghong Tian, Song Guo, and Dapeng Oliver Wu. "Can we beat DDoS attacks in clouds?" IEEE Transactions on Parallel and Distributed Systems 25, no. 9, pp. 2245-2254, 2014.
17. Awatef Balobaid, Wedad Alawad, and Hanan Aljasim. "A study on the impacts of DoS and DDoS attacks on cloud and mitigation techniques." International Conference on Computing, Analytics and Security Trends (CAST), IEEE, pp. 416-421, 2016.
18. Sidharth Sharma, Santosh Kumar Sahu, and Sanjay Kumar Jena. "On selection of attributes for entropy based detection of DDoS." International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, pp. 1096-1100, 2015.
19. Anna Buczak, and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection." IEEE Communications Surveys & Tutorials 18, no. 2, pp. 1153-1176, 2016.
20. Vasilios Siris, and Fotini Papagalou. "Application of anomaly detection algorithms for detecting SYN flooding attacks." Computer communications 29, no. 9, pp. 1433-1442, 2006.
21. Kanwal Garg, and Rshma Chawla. "Detection of DDoS attacks using data mining." International Journal of Computing and Business Research (IJCBR) 2, no. 1, pp. 2226-6166, 2011.
22. Rana Ashfaq, Aamir Raza, Xi-Zhao Wang, Joshua Zhexue Huang, Haider Abbas, and YuLin He. "Fuzziness based semi-supervised learning approach for intrusion detection system." Information Sciences 378, pp. 484-497, 2017.
23. Wei Pan, and Weihua Li. "A hybrid neural network approach to the classification of novel attacks for intrusion detection." In International Symposium on Parallel and Distributed Processing and Applications, Springer, Berlin, Heidelberg, pp. 564-575, 2005.
24. Mohammad Reza Norouzian, and Sobhan Merati. "Classifying attacks in a network intrusion detection system based on artificial neural networks." 13th International Conference on Advanced Communication Technology (ICACT), IEEE, pp. 868-873, 2011.
25. Mehdi Barati, Azizol Abdullah, Nur Izura Udzir, Ramlan Mahmod, and Norwati Mustapha. "Distributed Denial of Service detection using hybrid machine learning technique." International Symposium on Biometrics and Security Technologies (ISBAST), IEEE, pp. 268-273, 2014.
26. Fariba Haddadi, Sara Khanchi, Mehran Shetabi, and Vali Derhami. "Intrusion detection and attack classification using feed-forward neural network." 2nd International Conference on Computer and Network Technology (ICCNT), IEEE, pp. 262-266, 2010.
27. Reyhaneh Karimazad, and Ahmad Faraahi. "An anomaly-based method for DDoS attacks detection using RBF neural networks." Proc of the International Conference on Network and Electronics Engineering, pp. 16-18. 2011.
28. Dimitris Gavrilis, and Evangelos Dermatas. "Real-time detection of distributed denial-ofservice attacks using RBF networks and statistical features." Computer Networks 48, no. 2, pp. 235-245, 2005.
29. Chang-Lung Tsai, Allen Y. Chang, and Ming-Szu Huang. "Early warning system for DDoS attacking based on multilayer deployment of time delay neural network." 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), IEEE, pp. 704-707, 2010.
30. Tony Bazzi, Jasser Jasser, and Mohamed Zohdy. "Affine Arithmetic Self Organizing Map." matrix 5, no. 04, pp.359-365, 2016.
31. Jasser Jasser, Tony Bazzi, and Mohamed Zohdy. "On Modified Self Organizing Feature Maps." Transactions on Machine Learning and Artificial Intelligence 4, no. 6, pp.71-77, 2017.
32. Fatma Gumus, C. Okan Sakar, Zeki Erdem, and Olcay Kursun. "Online Naive Bayes classification for network intrusion detection." International Conference onAdvances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM, pp. 670-674, 2014.
33. Ashley Chonka, and Jemal Abawajy. "Detecting and mitigating HX-DoS attacks against cloud web services." 15th International Conference on Network-Based Information Systems (NBiS), IEEE, pp. 429-434, 2012.
34. P. Arun Raj Kumar, and S. Selvakumar. "Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems." Computer Communications 36, no. 3, pp. 303-319, 2013.
35. P. Arun Raj Kumar, and S. Selvakumar. "Distributed denial of service attack detection using an ensemble of neural classifier." Computer Communications 34, no.11, pp. 1328-1341, 2011.
36. Md Tanzim Khorshed, ABM Shawkat Ali, and Saleh A. Wasimi. "Monitoring insiders activities in cloud computing using rule based learning." 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE, pp. 757-764, 2011.
37. Sumouli Choudhury, and Anirban Bhowal. "Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection." International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), IEEE, pp. 89-95, 2015.
38. Yacine Bouzida, and Frederic Cuppens. "Neural networks vs. decision trees for intrusion detection." In IEEE/IST Workshop on Monitoring, Attack Detection and Mitigation (MonAM), vol. 28, pp. 29, 2006.
39. Manjula Suresh, and R. Anitha. "Evaluating machine learning algorithms for detecting DDoS attacks." Advances in Network Security and Applications, pp.441-452, 2011.
40. RR Rejimol Robinson, and Ciza Thomas. "Ranking of machine learning algorithms based on the performance in classifying DDoS attacks." Recent Advances in Intelligent Computational Systems (RAICS), IEEE, pp. 185-190, 2015.
41. Mouhammd Alkasassbeh, Ghazi Al-Naymat, Ahmad BA Hassanat, and Mohammad Almseidin. "Detecting Distributed Denial of Service Attacks Using Data Mining Techniques." International Journal of Advanced Computer Science and Applications 7, no. 1, 2016.
42. Jin Li, Yong Liu, and Lin Gu. "DDoS attack detection based on neural network." 2nd International Symposium on Aware Computing (ISAC), IEEE, pp. 196-199, 2010.
43. Amin Dastanpour, Suhaimi Ibrahim, Reza Mashinchi, and Ali Selamat. "Comparison of genetic algorithm optimization on artificial neural network and support vector machine in intrusion detection system." IEEE Conference on Open Systems (ICOS), IEEE, pp. 72-77, 2014.
44. Saman Taghavi Zargar, James Joshi, and David Tipper. "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks." IEEE communications surveys & tutorials 15, no. 4, pp. 2046-2069,2013.
45. Anaconda Documentation. [online] Available at: https://docs.anaconda.com, [Accessed 13 Dec. 2017].
46. Spyder - Documentation. [online] Available at:https://pythonhosted.org/spyder/, [Accessed 13 Dec. 2017].
47. Scikit-learn Machine Learning in Python. [online] Available at: http://scikit-learn.org/stable/, [Accessed 13 Dec. 2017].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

48. KDD Cup 1999 Data. [online] Available at: http://kdd.ics.uci.edu/databases/kddcup99/ kddcup99.html, [Accessed 13 Dec. 2017].
49. NSL-KDD dataset. [online] Available at:http://www.unb.ca/cic/datasets/nsl.html, [Accessed 13 Dec. 2017].
50. Tom Dietterich. "Overfitting and undercomputing in machine learning." ACM computing surveys (CSUR) 27, no. 3, pp. 326-327,1995.
51. Davide Anguita, Alessandro Ghio, Sandro Ridella, and Dario Sterpi. "K-Fold Cross Validation for Error Rate Estimate in Support Vector Machines." In DMIN, pp. 291-297, 2009.
52. Payam Refaeilzadeh, Lei Tang, and Huan Liu. "Cross-validation." In Encyclopedia ofdatabase systems, Springer US, pp. 532-538, 2009.
53. Ron Kohavi. "A study of cross-validation and bootstrap for accuracy estimation and model selection." In IJCAI, vol. 14, no. 2, pp. 1137-1145, 1995.