Adaptive feature selection for denial of services (DoS) attack

Conference Paper · November 2017

DOI: 10.1109/AINS.2017.8270429

CITATIONS
29

READS
548

5 authors, including:

Rizaain Yusof
Universiti Putra Malaysia
4 PUBLICATIONS **90** CITATIONS
SEE PROFILE

Nur Izura Udzir
Universiti Putra Malaysia
220 PUBLICATIONS **1,994** CITATIONS
SEE PROFILE

Ali Selamat
Universiti Teknologi Malaysia
517 PUBLICATIONS **6,112** CITATIONS
SEE PROFILE

Hazlina Hamdan
Universiti Putra Malaysia
18 PUBLICATIONS **89** CITATIONS
SEE PROFILE

Adaptive Feature Selection for Denial of Services (DoS) Attack

Ahmad Riza'ain Yusof
School of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Serdang, Selangor, Malaysia.
Email: rizaain@gmail.com

Nur Izura Udzir
School of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Serdang, Selangor, Malaysia.
Email: izura@upm.edu.my

Ali Selamat
Faculty of Computing
Universiti Teknologi Malaysia (UTM)
Johor Bahru, Johor, Malaysia.
Email: aselamat@utm.my

Hazlina Hamdan
School of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Serdang, Selangor, Malaysia.
Email: hazlina@upm.edu.my

Mohd Taufik Abdullah
School of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
Serdang, Selangor, Malaysia.
Email: taufik@upm.edu.my

***Abstract*—Adaptive detection is the learning ability to detect any changes in patterns in intrusion detection systems. In this paper, we propose combining two techniques in feature selection algorithm, namely consistency subset evaluation (CSE) and DDoS characteristic features (DCF) to identify and select the most important and relevant features related DDoS attacks. The proposed technique is trained and tested using the NSL-KDD 2009 dataset and compared with the traditional features selection method such as Information Gain, Gain Ratio, Chi-squared and Correlated features selection (CFS). The result shows that the combined CSE with DCF model overcomes the drawback of traditional feature selection technique such as avoid over-fitting, long training time and improved efficiency of detections. The adaptive model based on this technique can reduce computational complexity to analyze the data when attack occurs.**

***Index Terms*—NSL-KDD, Features Selection, Intrusion Detection, Machine learning**

I. Introduction

Intrusion detection system (IDS) emerged rapidly conjunction with the evolving Internet technology and became part of important tools and widely used for ensuring network security [1]. IDS categorized in two types, namely signature-based and anomaly-based. Signature-based, also known as misuse detection, can differentiate between normal or attack by comparing known signatures of discovered vulnerabilities.

The anomaly-based approach usually deals with statistical analysis and pattern recognition problems [2]. Any changes or deviation from normal behavior in network traffic is a sign of potential attempt in this kind of model. In the Internet evolution era, our network traffic becomes more complicated and dynamic. Furthermore, it can lead to the major difficulty to discover boundaries between normal and anomaly behavior in network traffic.

In spite of numerous research and industrial attempts to design DDoS protection, DDoS attacks are becoming a significant threat. The increasing scale, complexity, and mutation in attack patterns add to the complexity of the problem [3]. The study has produced a number of techniques to fight DDoS strikes at various network locations.

A DDoS detection system gathers and audits network information to examine if there is any intrusion attempt. However, the detection system faces large amounts of network data to be processed as the network extends greatly [4]. This problem also is known as the curse of dimensionality which causes high computational complexity, high overhead in processing and slow processing or classification time.

To overcome the curse of dimensionality problem feature selection technique has been introducing. Features selection (FS) has been widely utilized in fields, such as object recognition, network security, data mining and also classification. Moreover, FS is a part of dimensional reduction, by going through the process of selecting the most effective subsets of features that constitute the whole dataset. We can define it as good feature subsets contain features highly correlated with the classification, yet uncorrelated with each other [5]. They are many reasons why we use feature selection, but the general goal is to minimize the number of features, reduce over-fitting and improve the generalization of models. Feature selection is also used to assess better understanding of the features and their relationship to the response variables.

In this paper, we combined two feature selection method, DDoS characteristic-based features (DCF) and the Consistency-based Subset Evaluation (CSE) to select the significant features in NSL-KDD dataset. The simple majority vote is used to merge the combination of two output from that feature selection method.

This paper is organized as follows: Section 2 provides related work, While Section 3 describes the technique used in feature selection. Section 4 explains the datasets used in experiments. Experimental results are presented in Section 5. Finally, the conclusion is discussed in Section 6.

II. Related Study

Nowadays, a lot of works on feature extraction and selection for decision rules, an intelligent algorithm anomaly intrusion detection were proposed. Lin et al. [6] proposed an algorithm to find best-selected features in data set by combining three machine learning techniques such as Support Vector Machine (SVM), Decision Tree (DT) and simulated annealing (SA). They found that the proposed combining algorithm achieves a better result compared using SVM, DT and SA alone.

Hee et al. [7] observed a feature selection technique using Attribute Ratio (AR) which is calculated by mean and frequency of features. They found that the accuracy of AR method is higher than other feature selection methods such as Information Gain (IG), Correlation-based Feature Selection (CFS) and Gain Ratio(GR) using NSL-KDD dataset. However, they only used J48 decision tree as a classifier and 10-fold cross validation for the testing to estimate accuracy.

Back in 2012, a hybrid approach, which consists of transferring nominal features and normalizing numeric ones, all to the same scale was introduced [8]. They also prepared several dataset based on the hybrid approach and then evaluated using several classifiers. This evaluation shows that these enhancements innervate the IDS.

By developing a reliable and robust IDS framework especially in DDoS attack, feature selection and extraction are taken into account as a critical task for saving computational cost as well as to root out data patterns. The feature selection is used to search and select the most significant features from the original dataset. There are a lot of techniques that can be used for the feature selection such as Chi-Square and Symmetrical Uncertainty [9], Correlation-based feature selection (CFS), Consistency based subset evaluation (CSE) and Principal component analysis (PCA) [10] and ensemble-based multi-filter feature selection (EMFFS) method that combines the output of information gain (IG), gain ratio, chi-squared and ReliefF [11].

There was also a study which tries to detect a DDoS attack by its pattern [12] by selecting some features such as srcID, srcPort, dstIP, dstPort, protocol, flowSize and numOfPacket from the DDoS dataset. Then, they measure the magnitude of the average number of packet per flow in the specific time interval. From the calculation if the packet number is low and contained a large amount of flow, it is assumed that this is DDoS SYN Flood attack.

III. Description of Dataset

The NSL KDD dataset [13] is a modified and refined KDD Cup 99 dataset where redundant records in train set are removed and duplicate records in the test set are eliminated. The number of records in the train and test sets is also reduced to a reasonable amount which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Table I shows the number of instances in the dataset, including the normal and particular type of attack namely Denial of Service (DoS), Probe, Remote to Local Attack (R2L) and User to Root Attack (U2R).

TABLE I
Number of instance in NSL-KDD data set

Type of Dataset	Normal	DoS	Probe	R2L	U2R	Total Record
KDD Train+	67343	45927	11656	52	995	125973
KDD Test+	9711	7458	2421	200	2754	22544

IV. Proposed Method

In this section, we discuss the details of methods that have been utilized in this work for features selection on intrusion detection dataset. Fig. 1 shows the proposed method work flow which is used to extract and select the best features that represent a DDoS attack. Firstly, the feature selection were performed on the full feature training dataset, with a total of 42 features, before parallel applying DDoS characteristic-based features (DCF) and Consistency-based Subset Evaluation (CSE). The output from these two feature selection is combined using the simple majority vote technique in order to pick the most suitable features based on a selected threshold.

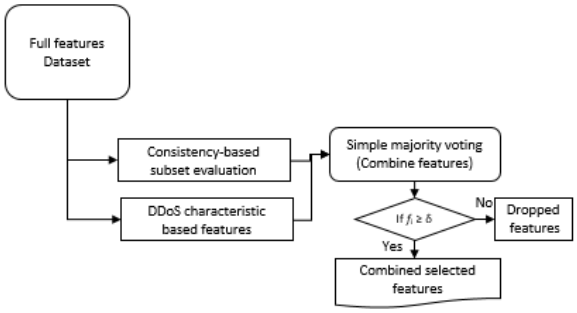

Fig. 1. The proposed feature selection process

A. DDoS characteristic-based features (DCF)

In DDoS characteristic-based features (DCF)[14], five features is selected namely DestIP, DestPort, SrcIP, SrcPort, packet length and ACK. These five features can exploit the abnormalities during DDoS attack.

B. Consistency-based Subset Evaluation (CSE)

This feature selection technique is proposed by Dash et al. [15], which tends to find optimal and evaluate a subset of relevant features that are consistent with each other. The consistency of a subset can be determined by representing a combination of features value with a class given pattern label. All features of given pattern should represent the same class. If there exist at least two instances that differs from the class, the

pattern will become inconsistent. We calculate Inconsistency Count (IC) by:

$$IC(p) = n_p - c_p \tag{1}$$

where c_p is the number of instances of the majority class of the n_p instances and n_p is the number of instances of the pattern p.

$$IR(S) = \frac{\sum_p IC(p)}{\sum_p n_p} \tag{2}$$

Inconsistency Ratio (IR) is used to calculate the overall consistency of a subset S. We can get IR value by counting the sum of all inconsistency over all the patterns of the feature subset divided by the total number of instances in the dataset.

C. Extreme learning machine (ELM)

Zhu et al. [16] and Huang et al.[17] work on a new technique called Extreme Learning Machine (ELM), which is different from traditional training single-hidden layer feed forward neural network (SLFN). In ELM, they randomly choose their parameters in the hidden layer and not tuned. For example, they randomly choose input weight and hidden neuron biases. According to [1] ELM tends to have good generalization performance and is very easy to implement.

V. Experiments and Results

The NSL-KDD dataset is used in the experiments as the attack component. This simulation on all datasets is carried out using WEKA 3.7 and MATLAB R2016a running on a machine with an Intel(R) Core(TM) i7-4790, 3.60GHz CPU and 16GB RAM. The main advantage of the NSL-KDD dataset is that it does not include any redundant and duplicate examples so the classifiers implemented on this dataset is not biased to the repeated records in the train set.

We have closely analyzed four feature selection techniques, and have conducted experiments on our training data and derived the best feature set using these three search techniques. Table II shows the result of the number of features selected using union technique to combined result from DCF and CSE and Table III shows the output of selected feature set according to their feature selection method, the number of the feature selected and the name of the selected features.

TABLE II
Union results using DDoS Characteristic-based features (DCF) and Consistency-based Subset Evaluation (CSE)

Feature Selection	Number of Feature Selected	Selected Features
DCF	11	2,3,4,5,6,7,8,14,23,30,36
CSE	10	1,3,4,5,14,23,32,34,35,37
DCF ∪ CSE	17	1,2,3,4,5,6,7,8,10,14,23,29,30,32,33,36

TABLE III
Feature Selection using other methods

Feature Selection	Number of Feature Selected	Selected Features
Info Gain	16	5,3,6,4,30,29,33,34,35,38,12,39,25,23,26,37
Gain Ratio	15	12,26,4,25,39,30,38,6,5,29,3,37,8,33,34
Chi-squared	17	5,3,6,4,30,29,33,34,35,12,23,38,25,39,26,37,32
CFS	14	29,33,34,12,39,38,25,4,26,23,32,3,28,41
Proposed Model	17	1,2,3,4,5,6,7,8,10,14,23,29,30,32,33,36

A. Performance Evaluation Criteria

The performance of an IDS is measured in terms of detection rate, accuracy, and false alarm rate. In general, IDS requires low false alarm rate, high accuracy, and high detection rate. Four criteria are chosen for evaluating the performance of an IDS classifier: accuracy, F1score, precision and test time. Test time represents actual CPU time taken by the IDS to classify all tests. We can calculate the performance of IDS using the following formula:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \times 100\% \tag{3}$$

$$F1 - score = \frac{(2 * TP)}{(2 * TP) + FP + FN)} \times 100\% \tag{4}$$

$$Precision = \frac{(TP)}{(TP + FP)} \times 100\% \tag{5}$$

Table IV shows the categories of data behavior in intrusion detection for binary category classes (normal and anomalous) in terms of true positive, true negative, false positive and false negative.

TABLE IV
Confusion Matrix

	Actual Class	
Predicted Class	Anomaly	Normal
Anomalous	TP	FN
Normal	FP	TN

Table V represents results comparing the proposed method with Information Gain, Gain Ratio, Chi-squared and Correlation-based feature selection (CFS) using Extreme Learning Machine classifiers. A reliable IDS consists of high accuracy and low detection time. Based on the result, it is shown that our proposed feature selection method gives overall good result compared to other techniques. Although Chi-squared gives the highest accuracy, it does not do well in terms of detection time compared to our proposed method.

TABLE V
THE EVALUATION RESULTS COMPARISON WITH OTHER FEATURE SELECTION METHOD USING EXTREME LEARNING MACHINE

Parameters	Proposed method	Info Gain	Gain Ratio	Chi-squared	CFS
Accuracy	0.917	0.839	0.921	0.962	0.881
time(s)	1.469	2.078	81.797	2.036	6.750

VI. CONCLUSION

In this paper, we have presented a feature selection algorithm using machine learning algorithms for effective intrusion detection, which combines consistency subset evaluation (CSE) and DDoS characteristic features (DCF). The NSL-KDD dataset is used as the attack data and based on some feature selection algorithms such as consistency based subset evaluation and DDoS characteristic-based features (DCF) in order to find the most relevant features. The result of the experiment demonstrates that our proposed model has better accuracy and performance compared to other techniques.

ACKNOWLEDGMENT

This material is based upon work supported by the Ministry of Higher Education Malaysia under Grant No. FRGS 08-01-15-1721FR.

REFERENCES

[1] C. Cheng, W. P. Tay, and G.-B. Huang, "Extreme learning machines for intrusion detection," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. Institute of Electrical & Electronics Engineers (IEEE), Jun 2012.

[2] N. Gao, L. Gao, Q. Gao, and H. Wang, "An intrusion detection model based on deep belief networks," in *2014 Second International Conference on Advanced Cloud and Big Data*. Institute of Electrical & Electronics Engineers (IEEE), November 2014.

[3] A. R. Yusof, N. I. Udzir, and A. Selamat, "An evaluation on KNN-SVM algorithm for detection and prediction of DDoS attack," in *Trends in Applied Knowledge-Based Systems and Data Science*. Springer Nature, 2016, pp. 95–102.

[4] M. H. Aghdam and P. Kabiri, "Feature selection for intrusion detection system using ant colony optimization," *International Journal of Network Security*, vol. 2, no. 3, pp. 420–432, May 2016.

[5] M. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, University of Waikato, 1999.

[6] S.-W. Lin, K.-C. Ying, C.-Y. Lee, and Z.-J. Lee, "An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection," *Applied Soft Computing*, vol. 12, no. 10, pp. 3285–3290, 2012.

[7] H. S. Chae, B. O. Jo, S. H. Choi, and T. K. Park, "Feature selection for intrusion detection using nsl-kdd," *Recent Advances in Computer Science*, p. 184187, 2013.

[8] M. Salem and U. Buehler, "Mining techniques in network security to enhance intrusion detection systems," *International Journal of Network Security and Its Applications*, vol. 4, no. 6, 2012.

[9] E. Balkanli, A. N. Zincir-Heywood, and M. I. Heywood, "Feature selection for robust backscatter DDoS detection," in *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*. Institute of Electrical and Electronics Engineers (IEEE), October 2015.

[10] P. Narang, J. M. Reddy, and C. Hota, "Feature selection for detection of peer-to-peer botnet traffic," in *Proceedings of the 6th ACM India Computing Convention on - Compute*. Association for Computing Machinery (ACM), 2013.

[11] O. Osanaiye, H. Cai, K.-K. R. Choo, A. Dehghantanha, Z. Xu, and M. Dlodlo, "Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing," *J Wireless Com Network*, vol. 2016, no. 1, May 2016.

[12] A. Sanmorino and S. Yazid, "DDoS attack detection method and mitigation using pattern of the flow," in *2013 International Conference of Information and Communication Technology (ICoICT)*. Institute of Electrical and Electronics Engineers (IEEE), March 2013.

[13] W. L. Mahbod Tavallaee, Ebrahim Bagheri and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA)*, 2009.

[14] D. W. et al., *Research on the Detection of Distributed Denial of Service Attacks Based on the Characteristics of IP Flow*. Springer, Berlin, Heidelberg, 2008, vol. 5245, ch. Network and Parallel Computing. NPC, pp. 86–93.

[15] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, no. 1-2, pp. 155–176, 2003.

[16] Q.-Y. Zhu, A. Qin, P. Suganthan, and G.-B. Huang, "Evolutionary extreme learning machine," *Pattern Recognition*, vol. 38, no. 10, pp. 1759–1763, October 2005.

[17] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, December 2006.