Volume 3, Issue 5, May 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Determining feature set of DOS attacks

Ms Pooja Bhoria
M Tech Scholar, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra India.

Dr. Kanwal Garg
Assistant Professor, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra India.

Abstract--Denial of service attacks are large scale co operative attacks on the networking structure. It disables servers/ victims from providing services to its clients. These attacks adversely affect the network badly. Therefore they must be detected on time. IDS using classification plays a significant role in detecting such intrusions but it takes significant classification time due to large number of features. It reduces its efficiency. So in order to improve efficiency or to reduce classification time, researcher provides relevant set of features for detection of DOS attacks. For this purpose, researcher is using NSL KDD dataset and analysis is performed on orange canvas V2.6.1 data mining tool.

Keywords: DOS (Denial of Service) attack, LAND (local area network denial) attack, POD (ping of death) attack.

I. Introduction

Denial of service attacks on computer infrastructure is becoming a serious issue now days. It adversely affect computer network by launching large scale co operative attacks on victim machines called zombies [2]. It results in disabling of services from providing services to its clients which makes it a critical issue for protecting the systems against such threats. Intrusion detection is interesting approach that could be helpful toward it. This Intrusion detection task is performed by installing intrusion detection system (IDS) at system nodes. From these nodes, IDS continuously observes the network traffic passing through the node with the help of sensors [1],[2]. These sensors are capable of collecting patterns either from network streams or from host machine or from any particular applications like web [3]. Whenever it encounters attacks patterns (i.e. patterns that violets security principles like integrity, confidentiality and availability) it generates notification through alarm, e mails [2],[4]. These intrusion detection attacks can be alienated into two groups i.e. anomaly and misuse detection attacks. 1) Misuse detection attacks: These attacks identify the attack patterns by equating patterns obtained from sensors with the stored patterns in data storage [3]. Whenever match occurs, it's an intrusion. It is also known as known attack strategies or signature based intrusion detection methods [4]. 2) Anomaly detection attacks: anomaly or unknown attack stores normal behaviour patterns in data storage. Patterns obtained from sensor are matched with the patterns stored. Whenever deviation occurs, it is an attack pattern [5]. To detect all the DOS attack patterns covered under these categories, researcher determines relevant features set for efficient intrusion detection of attacks. This detection is performed by appling the selected feature set to the C4.5 decision tree classification algorithm. This decision tree algorithm is selected because it provides best classification accuracy among supervised learning algorithms. To frame all the objectives stated above, this paper is divided into six sections. Section one comprises introduction. Section two portrays IDS dataset Description and Tool Information. Section three discusses feature selection and its implementation to C4.5. Section four describes experimental results. In section five, researcher finally concludes the paper and section six comprises of references.

II. Ids Dataset Description And Tool Information

NSL KDD dataset is offline network data based on KDD 99 dataset. It provides benchmark for intrusion detection occurring in the network. This dataset has about 4,90,000 single connection records with no redundancy. Each connection record has 41 attributes and one class attributes. Class attribute labels connection as normal or attack with exactly one specific attack type. But for analysis researcher is using 20% of NSL KDD dataset i.e. about 22,495 records with class attribute as normal or DOS attack with exactly one specific attack type. This specific type includes:

1) Back attack: In this type of dataset, attacker forges the origin IP Address by placing it with IP Packet. Due to this, victim can't determine attacker. As soon as the attacker receives the packets, they start responding to unknown attacker. Researcher's selected dataset includes 196 instances of this attack. [9]
2) LAND attack: also called local area network denial attack. In this attack, forged packets with the host IP address are sent to host computer. It causes host computer to be busy in replying to these packets. Researcher's selected dataset includes 1 instance of this attack. [9]
3) Ping of death attack: In this type of attack, attacker attacks the victim machine by sending huge packets of size larger than 64,000 bytes which results in crashing of system. Researcher analysed dataset includes 38 instances of this pod attack. [9]
4) Smurf attack: For attempting this attack, attacker uses forged IP address. Attacker spreads ICMP packets with IP address of victim machine to network. Whenever victim machine receives these packets, they send echo reply

messages as response. It causes extra traffic and prevent victim from providing services. Analysed dataset includes 529 instances of attack. [9]

5) Neptune attack: In this attack, attacker sends session establishment packet with forged source IP address. The victim machine then uses its resources and waits for session conformation. During this time slice, victim machine becomes unavailable to legitimate traffic. Analysed dataset includes 8282 instances of attack. [9]

To detect all these normal and DOS attack patterns, Analysis is performed on Orange Canvas version 2.6.1 data mining tool. This data mining tool provides unified benchmark for researchers for analysis. Along with that, this tool provides better user interface to users to create appropriate workflow schema according to our requirements.

III. Feature Selection And Implementation To C4.5 Decision Tree Classification Algorithms

Analysed NSL KDD dataset includes 41 features. These features include numeric, symbolic, continuous attributes as depicted in figure 1. These all features can further be broadly categorises into three categories: 1) Basic features: This category includes all the attributes that can be extracted from TCP/IP connection [5]. 2) Traffic features: This category of features includes time based traffic features and host based traffic features. Time based features describe connections within time slice of past two seconds [5]. Host based features depicts attributes using the window of 100 connections to same host [5]. 3) Content based features: This type of features look for suspicious behaviour in the data such as number of failed login attempts [5].

BASIC FEATURES	duration, protocol_type, service, flag, src_bytes, dst_byte, land, wrong_fragment, urgent
CONTENT FEATURES	hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_root, num_file_creation, num_shells, num_access_files, num_outbound_cmds, is_host_login, is_guest_login
TRAFFIC FEATURES (TIME BASED + HOST BASED)	count srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate, dst_host_srv_rerror_rate

FIGURE 1: FEATURES IN NSL KDD DATASET

In order to improve the performance of IDS, following feature set is selected for analysis. First feature set includes all the features of NSL KDD dataset. Then after combinations of basic features and time based features are used for determination of normal and DOS attacks patterns. This second feature set includes 28 attributes that are used for classification.

FEATURE SET 1	duration, protocol_type, service, flag, src_bytes, dst_bytes, land, wrong_fragment, urgent, hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_root, num_file_creations, num_shells, num_access_files, num_outbound_cmds, is_host_login, is_guest_login, count, srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate, dst_host_srv_rerror_rate
FEATURE SET 2	duration, protocol_type, service, flag, src_bytes, dst_bytes, land, wrong_fragment, count, srv_count, serror_rate, srv_serror_rate, count, srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate, diff_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate, dst_host_srv_rerror_rate
FEATURE SET 3	protocol_type, src_bytes, dst_host_same_srv_rate, count, diff_srv_rate, dst_host_serror_rate, dst_host_diff_srv_rate, wrong_fragment

FIGURE 2: FEATURE SETS FOR PERFORMANC IMPROVEMENT

And within third feature set, it includes 8 features as stated in figure 2. Among these features selected, features such as protocol_type, src bytes, dst_host_same_srv_rate, count, diff_srv_rate, dst_host_serrror_rate provides better classification of normal patterns. For land DOS attack, attributes i.e. count, diff_srv_rate, dst_host_serror_rate, dst_host_diff_srv_rate performs efficiently. For back attack, features such as dst_host_same_srv_rate, protocol type, src_bytes provides better classification. For Neptune DOS attack, feature such as count, diff_srv_rate, dst_host_serror_rate, src_bytes, dst_host_diff_srv_rate performs much better than feature set of 41 attributes. For POD DOS attacks, protocol type, src_bytes and wrong fragment features are used. For Smurf attack, collection of features which includes protocol type, src_bytes, wrong fragment are used. And for the efficient detection of teardrop DOS attack, feature set includes protocol type, src bytes.

IV. Analysis And Interpretation

Analysis work is performed by appling NSL KDD dataset to C4.5 decision tree classification algorithms on Orange canvas data mining tool. This tool is installed on the system having Intel core 2 duo processor 2.0 G HZ processor, and 1 GB of RAM. During analysis, researcher performs 6 folds cross validation on NSL KDD benchmark intrusion detection

dataset. Standard parameters includes classification time, accuracy are used for measuring the performance. When all the 41 attributes are considered for decision tree classification then it needs classification time of 54.37 seconds. This classification provides accuracy of 99.89% which determines how correctly it identifies the class of patterns. Table 1 tabulates the experimental results of classification.

	CLASSIFICATION ACCURACY	CLASSIFICATION TIME
FEATURE SET 1	99.89%	54.37 Sec
FEATURE SET 2	99.91%	31.31 Sec
FEATURE SET 3	99.93%	8.53 Sec

TABLE 1: EXPERIMENTAL RESULTS WITH ATTRIBUTES SELECTION

This classification accuracy can better be illustrated by confusion matrix. Confusion matrix gives detailed overview of how much instances of each class are correctly classified and incorrectly classified. Table 2 depicts confusion matrix when 41 attributes are considered. It concludes that C4.5 Decision tree Classification algorithm classifies 13441 instances out of 13449 correctly but 9 instances are incorrectly classified as attack. Likewise all the patterns are identified, some are correctly classified while some are incorrectly classified as depicted in table 2.

	normal	back	land	neptune	pod	smurf	teardrop	
normal	13441	1	0	7	0	0	0	13449
back	4	192	0	0	0	0	0	196
land	1	0	0	0	0	0	0	1
neptune	5	0	0	8277	0	0	0	8282
pod	0	0	0	0	36	2	0	38
Smurf	0	0	0	0	0	529	0	529
teardrop	1	0	0	0	0	0	187	187
	13452	193	0	8284	36	531	187	22683

TABLE 2: CONFUSION MATRIX WITH 41 ATTRIBUTES

These incorrectly identified instances results in high false alarm rate which is the combination of false positive rate and false negative rate. Higher will be the false alarm rate, worse will be the accuracy so it must be as much less as possible. This accuracy and less classification time are achieved by reducing the number of attributes and produce a relevant set of features. When collection of all basic features and time based features i.e. feature set having collection of 28 features is applied to C4.5 decision tree classification, it provides much better accuracy of 99.91%. This achieved classification accuracy is much better than feature set having 41 attributes. Confusion matrix for features set having 28 attributes is tabulated in table 3. Along with accuracy, this feature set provides better classification time of 31.31 seconds but in actual it's too large for classification. If classification time is that much large then definitely it might be possible that it can't detect all attacks due to large traffic in the network. It adversely affects the efficiency.

	normal	**back**	**land**	**neptune**	**pod**	**smurf**	**teardrop**	
normal	13440	2	0	7	0	0	0	13449
back	2	194	0	0	0	0	0	196
land	1	0	0	0	0	0	0	1
neptune	5	0	0	8277	0	0	0	8282
pod	0	0	0	0	36	2	0	38
Smurf	0	0	0	0	0	529	0	529
teardrop	1	0	0	0	0	0	187	188
	13449	196	0	8284	36	531	187	22683

TABLE 3: CONFUSION MATRIX WITH 28 ATTRIBUTES

So in order to achieve better accuracy and classification time, relevant features set of total 8 features are used with 1 as Meta feature. This feature set includes features i.e. protocol type, src bytes, dst_host_same_srv_rate, count, diff_srv_rate, dst_host_serror_rate, dst_host_diff_srv_rate, wrong fragment and attack type as meta attribute. When this feature set is applied for classification then it is concluded that it takes classification time of 8.53 seconds which is very much less than other feature sets selected. Along with better classification time, this feature set provides much better accuracy of 99.93%. Confusion matrix depicts this achieved accuracy with much better illustration and is shown in table 4.

	normal	**back**	**land**	**neptune**	**pod**	**smurf**	**teardrop**	
normal	13442	2	0	0	0	0	1	13449
back	1	195	0	0	0	0	0	196
land	1	0	0	0	0	0	0	1
neptune	5	0	0	8277	0	0	0	8282
pod	0	0	0	0	36	2	0	38
smurf	0	0	0	0	0	529	0	529
teardrop	1	0	0	0	0	0	187	188
	13450	197	0	8281	36	531	188	22683

TABLE 4: CONFUSION MATRIX WITH 8 ATTRIBUTES

V. Conclusion

In this paper, researcher statistically analysed NSL KDD dataset with 6 folds cross validation. The analyses concluded that when more relevant feature set is applied for classification then it improves performance of IDS. This improved performance is achieved due to increased classification accuracy and reduced classification time. This increased classification accuracy provides much better detection capacity of IDS towards attacks and hence results in safe networking to users. Classification time is also reduced by selecting more efficient feature set. This reduced classification time increases the capacity of patterns categorised. It means that lesser the classification time of single pattern, more will be the patterns determined by IDS. Concluding the combination of accuracy and classification time makes IDS more contributing towards efficiency and reliability.

Acknowledgments

Author like to thanks to thesis guide Dr. Kanwal Garg for his deep efforts and support towards the development of this research work. Also, Author would like to thanks MIT Lincoln laboratory for providing such a valuable dataset for research in the field of intrusion detection.

References

[1] Sherish Johri [2012], "Novel Method for Intrusion Detection using Data Mining", Dated 10-01-2013, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), ISSN: 2277-128X, Volume 2, Issue 4.

[2] Kanwal Garg, Rshma Chawla [2011], "Detection of DDOS attacks using Data Mining", Dated 07-01-2013, International Journal of Computing and Business Research (IJCBR), ISSN(Online):2229-6166, Volume 2, Issue 1.

[3] Deepak Rathore, Anurag Jain [2012], "Design Hybrid Method for Intrusion Detection using Ensemble Cluster Classification and SOM network", Dated: 03-01-2013, International Journal of Advanced Computer Research (IJACR), ISSN: 2277:7970, Volume 2, Number 3, Issue 5.

[4] H. Patel & J. Sarvakar [2011], "Analysis of Data Mining Algorithm in Intrusion Detection", Dated 02-09-2012, International Journal of Emerging Technology and Advanced Engineering (IJETAE), ISSN 2250-2459, Volume 1, Issue 2, U.V.

[5] Radhika Goal, Anjali Sardana, & Ramesh C. Joshi[2011], "Parallel Misuse and Anomaly Detection Model", Dated 12-12-2012, International Journal of Network Security (IJNS), PP. 211-222, Volume 14, Number 4.

[6] Mr Anurag Adhare, Prof Arvind Bhagat Patil [2012], "Mitigating Denial Of Service attack using Genetic Approach", Dated: 02-02-2013, IOSR Journal Of Engineering, Volume 2(3), PP 468-472.

[7] Alpa Reshamwala, Dr. Sunita Mahajan, "prediction of DOS attack sequences", Dated 12-02-2013.

[8] J. Koshal, Monark Bag [2012], "Cascading of C4.5 Decision Tree and Support Vector Machine for Rule Based Intrusion Detection System", Dated 10-01-2013, Pages 8-20, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India.

[9] Farhad Soleimanaian Gharchcho Pogh, Neda Jabbari, Zonab Ghaffari Azar [2012], " Evaluation of Fuzzy K Means Clustering Algorithm in Intrusion Detection Systems", Dated: 14-02-2013, International Journal of Scientific and Technology Research, ISSN: 2277-8616, Volume 1, Issue 11.

[10] Rupali Datti, Shilpa Lakhina[2012], "Performance Comparison of Feature Reduction Techniques For Intrusion Detection Systems", Dated: 10-02-2012, International Journal of Computer Science and Technology, IJCST, ISSN : 0976-8491, Volume 3, Issue 1.

[11] Shilpi Lakhina, Sini Joseph and Bhupendra Verma [2010], "Feature Reduction using Principal Component Analysis for effective Anomaly Based Intrusion Detection on NSL KDD", Dated: 19-02-2013, International Journal of Engineering Science and Technology (IJEST), ISSN: 1790-1799, Volume 2(6).

[12] Hafiz Muhammad Imran, Azween Bin Abdullah, Muhammad Hussian, Seflappan Palaniappan, Iftikhar Ahmed [2012], "Intrusion Detection Based on Optimum Features Subset and Efficient Dataset Description", Dated: 04-12-2012, International Journal of Engineering and Innovative technology (IJEIT), ISSN : 2277-3754,Volume 2, Issue 6.

[13] Rupalli Datti, Bhupendra Verma [2010], "Feature Reduction for Intrusion Detection using Linear Discriminant Analysis" Dated: 04-12-2012, International Journal of Computer Science and Engineering (IJCSE), ISSN: 1072-1078, Volume 2, Issue No 4.

[14] Maher Salim, Ulrich Buehler [2012], "Mining techniques in Network Security to Enhance Intrusion Detection Systems", Dated: 02-01-2013, International Journal of Network Security and its Applications (IJNSA), Volume 4, Issue no 6.

AUTHOR PROFILE

Ms Pooja Bhoria presently pursuing M Tech in CSE in Department of Computer Science And Applications, Kurukshetra University, Kurukshetra. My area of research is databases, data mining, and data warehouse in which I tried to incorporate data mining with security applications i.e. IDS.

Dr Kanwal Garg presently working as Assistant Professor in Department Of Computer Science And Application, Kurukshetra University Kurukshetra. Owe the credit of more than 50 research papers published in international & national journals, conference & seminar.. His area of expertise is Data Bases, Data Mining, & warehousing.

