www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 3 March 2017, Page No. 20504-20509
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i3.19

Cluster Based DDoS Detection Method in Data Mining

Ms.R.Keerthika[1], Dr.C.Nalini[2] ,Ms.P.Suganthi[3] ,Ms.S.Abinaya[4]
[1]Assistant Professor, Department of Information Technology, Karpagam College of Engineering, Coimbatore.
[1]E-mail id:keerthikait@gmail.com.
[2]Professor,Department of Information Technology, Kongu Engineering College, Perundurai.
[2]E-mail id:nalinikec@gmail.com.
[3,4]Student, Department of Information Technology, Karpagam College of Engineering, Coimbatore.

***Abstract*—Distributed denial of service attacks (DDoS) are a growing threat to business worldwide. By adopting new purpose built solutions designed specifically to detect and defeat DDoS attacks, businesses can keep their business operation running smoothly. Data mining algorithm is used presents a DDoS attack detection model. Experimental result source that DDoS attacks can be detected efficiently and swiftly.**

KEYWORDS- Distributed Denial of Service (DDoS), data mining, Clustering, Hidden Markov model .

I.INTRODUCTION

Interconnected systems, such as Web servers, database servers, cloud computing servers etc, are now under threads from network attackers. It means, Denial of Service (DoS) attacks cause serious impact on these computing systems. In this paper, we present a DoS attack detection system that uses Multivariate Correlation Analysis for accurate network traffic characterization by extracting the geometrical correlations between network traffic features. These makes our solution capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic only. It shows that our system open forms two other previously developed state of-the-art approaches in terms of detection accuracy. In this paper, we propose a statistical based approach to detect rogue access points using a (HMM) Hidden Markov Model applied to passively measure packet header data collected at a gateway router. Our approach utilizes variations in packet interval time to differentiate between authorized access points and rouge access points. We designed and developed our Hidden Markov Model by analyzing Denial of Service attacks and the traffic characteristics of Wireless Local Area Networks based on 802.11.

Experimental validations demonstrate the effectiveness of our approach. Our trained Hidden Markov Model can detect the presence of a Rogue Access Point promptly within one second with extreme accuracy (very low false positive and false negative ratios are obtained)[2]. The success of our approach lies in the fact that it leverages knowledge about the behaviour of the traffic characteristics of 802.11 based WLANs and properties of Denial of Service attacks. Our proposed IDS adopts an anomaly detection approach and it profiles the CRN system parameters through a learning phase. So, our proposal is also able to detect new types of attacks. As an example, we present the case of detection of a jamming attack, which was not known to the IDS beforehand. The proposed IDS is evaluated through computer based simulations, and the simulation results clearly indicate the effectiveness of our proposal.

A.RELATED WORK

As advances in networking technology help to connect the distant corners of the globe and as the Internet continues to expand its influence as a medium for communications and commerce, the threat from spammers, attackers and criminal enterprises has also grown accordingly. It is the prevalence of such threats that has made intrusion

detection systems - the cyberspace's equivalent to the burglar alarm - join ranks with firewalls as one of the fundamental technologies for network security. However, today's commercially available intrusion detection systems are predominantly signature based intrusion detection systems that are designed to detect known attacks by utilizing the signatures of those attacks. Denial of Service (DoS) attacks constitute one of the major threats and among the hardest security problems in today's Internet. With little or no advance warning, a DDoS attack can easily exhaust the computing and communication resources of its victim within a short period of time[1]. Other major contributions include the high rate of detection and very low rate of false alarms obtained by flow analysis using Self Organizing Maps.

B.SCOPE OF THIS WORK

Once the network traffic detection module detect the traffic value is abnormal, the packet protocol status detection module will be started immediately to detect these packets in order to make sure if the current abnormal network traffic value results from DDoS attacks, at the same time the traffic threshold model could dynamically update according to the result of the packet protocol status detection module. It could adapt to the traffic variety caused by the increase of users or the change in application. The achievement of this module will greatly reduce the data which needs to be detected and detect DDoS attack in real time.

The main contributions of the current paper are as follows.

1. DDoS attacks are easier than ever to execute, drain more resources and budget to mitigate and place more revenue at risk than ever before. Read this eBook to find out how to mitigate against common DDoS attacks.
2. This method can effectively differentiate between normal and attack traffic.
3. The linear complexity of the method makes its real time detection practical.
4. This method can detect even very subtle attacks only slightly different from the normal behaviors.
5. The attacker sends a large number of packets from zombies to a server, to prevent the server from conducting normal business operations.

II.HIDDEN MARKOV MODEL

HMM (Hidden Markov model) can describe most practical stochastic signals, including non-stationary and the non-Markova. It has been widely applied in many areas such as mobility tracking in wireless networks, This document play a vital role in the development of life cycle (SDLC)as it describes the complete requirement of the system. It means for use by developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process. Spiral Model was defined by Barry Boehm in his 1988 article, "A spiral Model of Software Development and Enhancement. This model was not the first model to discuss iterative development, but it was the first model to explain why the iteration models. As originally envisioned, the iterations were typically 6 months to 2 years long.

Each phase starts with a design goal and ends with a client reviewing the progress thus far. Analysis and engineering efforts are applied at each phase of the project, with an eye toward the end goal of the project.The steps for Spiral Model can be generalized as follows: The new system requirements are defined in as much details as possible. This usually involves interviewing a number of users representing all the external or internal users and other aspects of the existing system. A preliminary design is created for the new system. A first prototype of the new system is constructed from the preliminary design. This is usually a scaled-down system, and represents an approximation of the characteristics of the final product The final system is constructed, based on the refined prototype. The final system is thoroughly evaluated and tested. Routine maintenance is carried on a continuing basis to prevent large scale failures and to minimize down time.

The particular algorithm procedure is as follows:

DOI: 10.18535/ijecs/v6i3.19

Step1 Select at random K initial cluster centerK_1, K_2,..., K_k in m time window

Step2 Calculate the distance between each network traffic data x_i and initial cluster center through D_j

$= \min\{\| x K_i - _v \|\}$, the sample point that is the nearest to cluster center would be assigned to the cluster whose center is K_v

Step3 Move every K_w to its cluster center and recalculate the cluster center according to new data added in cluster. Then calculate the deviation including sample value in each cluster domain through formula:

$D = \sum[\min r_{=1,...,k} d\ x\ K(\ i,\ r\ )\ ]2.\quad i_{=}1$ [1]

Step4 The repetitive execution of step3 and step4 until the convergence of D value and all the cluster center will not move. After that the cluster center is the traffic mean value in different time window.

III. EXPERIMENTALSETUP

HsMM training:

a)Use the outputs of ICA module as the model training data set to estimate the parameters of HsMM.

b)Compute the entropy of the training data set and the threshold.

The monitoring phase includes the following steps:

1) Compute the difference matrix between the testing AM and the average matrix obtained in the training phase by the PCA.

2)Using the eight matrix, compute the feature dataset of the testing AM.

3)Using the de-mixing matrix, compute the independent signals.

4)The independent signals are inputted to the HsMM; entropies of the testing dataset are computed.

5) Output the result based on the threshold of entropy that was determined in the training phase based on the entropy distribution of the training data set.

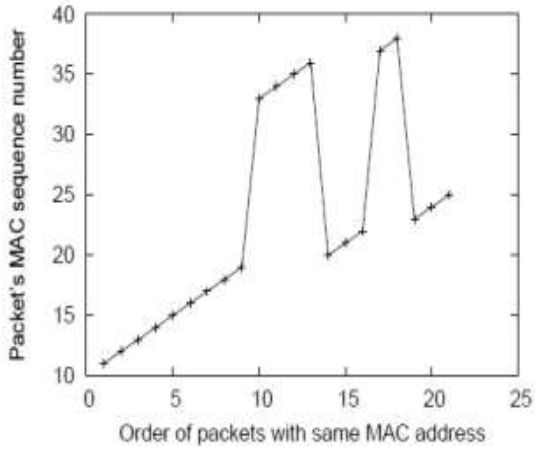

Fig 1: Detecting packets from mac-address

Detecting the more than one of successive packets from a one mac address (as shown in figure 1) can be used to identify the freeloaders. Session management page which associated with secure and a non-persistent cookie which contains the cryptographically random session id can be used to prevent session hijacking. Further, the page will be reloaded in a certain period and verify the session id in the cookie.

Constant Rate Attack:

Constant rate attack, the simplest attack technique, is typical among known DDoS attacks. We do not arrange the attack sources to simultaneously launch constant rate App-DDoS attacks and to generate requests at full rate, so that they cannot be easily identified through attack intensity. We use to denote the parameters of the constant rate attack. The notation is listed. Three parameters (i.e.,) are set randomly by each attack node before it launches the attacks. It shows the entropy varying with the time, where curve represents the normal flash crowd's entropy and curve represents the entropy of flash crowd mixed with constant rate App-DDoS attacks in zone B. Therefore, it is easy to find out that there exist attacks in the period B.

DOI: 10.18535/ijecs/v6i3.19

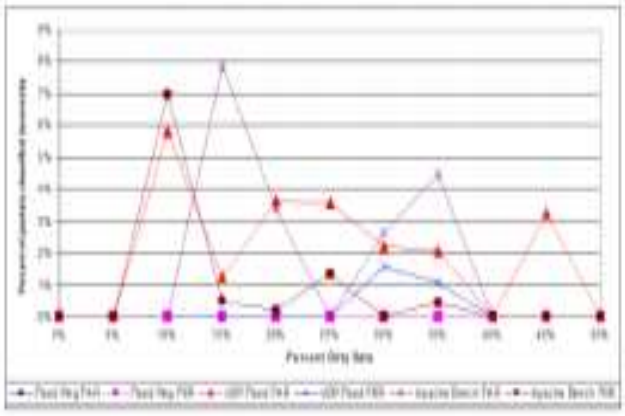

Fig 2 :False acceptance (FAR) and false rejection rates (FRR) for input data of different quality

The above-mentioned result (Figure 2) was obtained with a clean dataset but later the research group tested the system with some dirty data which has a mix of benign samples with malicious traffic. This mix depicts the real world scenario. Even though the clean data had almost 100% accuracy in classification[1], with the presence of dirty data there was an error rate below 10%. When noise was deliberately added to traffic a small percentage was classified wrongly (e.g. about 6% of the malicious packets are wrongly accepted if the baseline contains 10% attack traffic and the attack traffic includes 10% legitimate requests)

Increasing Rate Attack:

An abrupt change in traffic volume is an important signal to initiate anomaly detection. The attacker may use the gradually increasing rate. The state change in the victim network could be so gradual that services degrade slowly over a long period, delaying detection of the attack. We use to denote the parameters of the increasing rate attack. Five variables (i.e., are set randomly by each attack node before it launches the attacks. It shows the entropy changing with the time, where curve represents the normal flash crowd's entropy and curve represents entropy of the traffic that mixes normal flash crowd with increasing rate App-DDoS attacks[16], where the attacks start gradually in zone B and end gradually in zone D. As it shows, the entropy can be used to discover the attacks in the early beginning. Stochastic Pulsing Attacks: Instead of constantly injecting traffic flows with huge rates into the network, pulsing attacks, which are also called shrew attacks, are much more difficult to be detected.

	1 min			5 min		
Attack tools	Detection rate	Error rate	Residual rate	Detection rate	Error rate	Residual rate
SYN Flood	100%	0%	0%	100%	0%	0%
Stacheldraht	96.53%	1.05%	2.42%	99.53%	0.54%	0.13%
Trinoo	98.92%	0.41%	0.67%	99.69%	0.13%	0.18%
TFN2K	97.20%	2.45%	0.35%	98.65%	1.12%	0.23%

Table1 :DDoS attack detection result in different duration

From the analysis of DDoS attacks (Table 1) in this experiment, it is found that this system has a high detection efficiency, the detection rate of it could reach more than 97%. moreover, with the increase in the duration of DDoS attacks, higher is the attack detection rate of this system. The function test result of this system shows that it could meet the daily detection needs well

IV.CONCLUSION

Creating defences for attacks requires monitoring dynamic network activities in order to obtain timely and signification information. While most current effort focuses on detecting Net-DDoS attacks with stable background traffic, we proposed detection architecture in this paper aiming at monitoring Web traffic in order to reveal dynamic shifts in normal burst traffic, which might signal onset of App-DDoS attacks during the flash crowd event. Our method reveals early attacks merely depending on the document popularity obtained from the server log.

The proposed method is based on PCA, ICA, and HsMM. We conducted the experiment with different App-DDoS attack modes (i.e., constant rate attacks, increasing rate attacks and stochastic pulsing attack) during a flash crowd event collected from a real trace. Our simulation results show that the system could capture the shift of Web traffic caused by attacks under the flash crowd and the entropy of the observed data fitting to the HsMM can be used as the measure of abnormality. In our experiments, when the detection threshold of entropy is set 5.3, the DR is 90% and the FPR is 1%. It also demonstrates that the proposed architecture is expected to be practical in monitoring App DDoS attacks and in

triggering more dedicated detection on victim network. To handle the seriousness of DDoS attacks mimic or occur during the flash crowd event of a popular Website a new approach proposed to be implemented.

REFERENCES

[1]. R.Keerthika and Dr.C.Nalini,"Retrieving Datasets from nearest neighbor search using Spatial Queries"International Journal of Innovative Research and Technology,Volume 1 Issue 10 March 2015 ISSN No:2349-6002.

[2]. R.Keerthika and Dr.C.Nalini,"Image Retrieval Based on Content Search Mechanism", International Journal of Innovation Research in Computer and Communication EngineeringVolume 2 Issue 2 Febuary 2014 ISSN No:2320-9801.

[3]. M.Jayakameswaraiah and S.Ramakrishna,"Implementation of Improved ID3 Decision Tree Algorithm"Sri Venkateswara University, Tirupati, India,2014.

[4]. I. H. Witten, E. Frank, "Data Mining Practical Machine Learning Tools and Techniques", San Francisco: Morgan Kaufmann Publishers. China Machine Press, second edition ISBN 0-12-088407-0,560 pp, 2005.

[5]. D. Jiang, Information Theory and Coding [M]: Science and Technology of China University Press, 2001.

[6]. S. F. Chen, Z. Q. Chen, "An Artificial intelligence in knowledge engineering [M]". Nanjing: Nanjing University Press, 1997.

[7]. M. Zhu, "Data Mining [M]". Hefei: China University of Science and Technology Press Page No (67-72), 2002.

[8]. A. P. Engelbrecht., "A new pruning heuristic based on variance analysis of sensitivity information [J]". IEEE Trans on Neural Networks, Volume-12 Issue-06, Page No (1386-1399), November 2001.

[9]. N. Kwad, C. H. Choi, "Input feature selection for classification problem [J]", IEEE Trans on NeuralNetworks, Volume-13 Issue-01, Page No (143- 159), 2002.

[10]. X. J. Li, P. Wang, "Rule extraction based on data dimensionality reduction using RBF neural networks". ICON IP2001 Proceedings, 8th International Conference on Neural Information Processing [C]. Shanghai, China, Page No (149- 153), 2001.

[11]. S. L. Han, H. Zhang, H. P. Zhou, "correlation function based on decision tree classification algorithm for computer application", November 2000.

[12]. S. Y. Zhang, Z. Y. Zhu, "Study on decision tree algorithm based on autocorrelation function". Systems Engineering and Electronic Volume-27 Issue-07 Jul. 2005.

[13]. Bharati.M, Ramageri,"Data Mining Techniques and Applications", Indian journal of Computer Science and Engineering, Volume-01, Issue-04, Page NO (301-305), 2010.

[14]. Kalpesh Adhatrao, Aditya Gaykar, AmirajDhawan, RohitJha and Vipul Honrao,"Predicting,"Students Performance Using ID3 and C4.5 classification Algorithms", International journal Data mining and knowledge management process,Volume-03,Issue05,September 2013.

[15] Yufei Tao and Cheng Sheng,"Fast Nearest Neighbor Search with Keywords," IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 4, 2014

[16] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in the geographic web search engines. in the Proc. of ACM Management of Data (SIGMOD), pages 277–288, 2006.

[17] B. Chazelle, J. Kilian, R. Rubinfeld, and the A. Tal. the bloomier filter: efficient data structure for static support lookup tables. In the Proc. of the Annual ACM-SIAM Symposium on the Discrete Algorithms (SODA), pages 30–39, 2004.

[18] S. Agrawal, S. Chaudhuri, and the G. Das. Dbxplorer: the system for keyword-based search over relational databases. in the Proc. of International Conference on the Data Engineering (ICDE), pages 5–16, 2002.

[19] I. Kamel and C. Faloutsos. Hilbert R-tree: An improved r-tree using fractals. in the Proc. of Very Large Data Bases (VLDB), pages 500–509, 1994.

[16] N. Beckmann, H. Kriegel, R. Schneider, and the B. Seeger. the R*-tree: efficient and the robust access method for points and rectangles. in the Proc. of ACM Management of the Data (SIGMOD), pages 322–331, 1990.

[20] E. Chu, A. Doan, and J. Naughton. The Combining keyword search and forms for ad hoc querying of databases. In Proc. of ACM Management of Data (SIGMOD), 2009.
[21] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudar-shan. Keyword searching and the browsing in the databases using banks. in the Proc. of International Conference on the Data Engineering (ICDE), pages 431–440, 2002.

Ms.R.Keerthika Completed Bachelor's Degree at SSM College of Engineering, India in the year 2007 and Master's Degree at Anna University of Technology Coimbatore, India in the year 2011.Registered Ph.D in 2012 at Anna University Chennai,India. She is having 8 years of experience in teaching in engineering, Currently she is working as Assistant Professor, Department of Information Technology at Karpagam College of Engineering.

Her area of interest includes Data Mining, Image Processing, Wireless sensor Networks. She is a member in professional societies like ISTE, ACM, CSTA, IACSIT, ICGST, ISOC and IRED.

Dr.C.Nalini Completed her Master's Degree at Bharathiyar University, India in the year 2000. Completed her Ph.D in 2011 at Anna University Chennai. She is having 21 years of experience in teaching in engineering, currently she is working as a Professor, Department of Information Technology at Kongu Engineering College.

Her area of interest includes Data Mining, Network security, Database Management System. She is a member in professional societies like ISTE, CSI.