

Performance Analysis Of PCA And Artificial Intelligence Techniques For The Detection Of Application Layer Attacks

Prof. P.Krishna Kumar¹, Prof. Dr.K.Vijayalakshmi², Prof. Dr.R.Bharathi³

¹ *ASP/CSE ,PET Engineering College,Vallioor,Tamilnadu,India*

² *Prof&HOD/CSE, Ramco Institute of technology, Rajapalayam,Tamilnadu,India*

³ *AP/ECE,University College of Engineering Nagercoil,India*

{ ponkrishkumar@yahoo.com, vijaya@mepcoeng.ac.in,ret nabharathi15@gmail.com}

Abstract

With the rapid growth in on-line banking and on-line shopping which are integrated with Internet, the attacks on such services has also increased. This paper deals with the method of detecting the application layer attacks launched against websites. The real time experimental work is carried out using the data collected from the log file of a web server. A comparative performance analysis is done between the Principal Component Analysis (PCA) technique and the artificial intelligence techniques. The technique based on PCA is tested on several data sets and it accurately identifies 94.4% of APP-DDOS attacks whereas the existing method based on HSMM discriminates only 80.4% of DDOS attacks. The results also show that the artificial intelligence technique based on Fuzzy CMeans and Neural Network(FCM-NN) outperform PCA with a detection rate of 100 % and with detection accuracy of 100%.

Keywords: Principal Component Analysis, Application layer attacks, Artificial Intelligence techniques, Fuzzy CMeans, clustering, Fuzzy, Neural Network.

1. Introduction

The rapid increase in Internet connectivity and accessibility has provided a way for malicious users all over the world to involve in illegal activities. A 'Denial of Service' attack may be defined as an explicit attempt by attackers to prevent the authorized access to system resources for the legitimate clients. A Distributed Denial of service (DDoS) attack is a multiplication of DoS attack in which an attacker makes use of multiple computers in a coordinated manner to perform an attack. Many researchers

have proposed and implemented efficient techniques to detect the attacks on network layer [3]. Since the attacks on the network layer fails, the attackers have shifted their focus to application-layer attacks. Application layer DDoS attacks (App-DDoS) exploit the vulnerabilities at the application layer rather than at the network layer. Another special phenomenon on application layer is referred as Flash crowds [2]. Flash crowds are a sudden, large surge in traffic to a particular web site. The most important fact about the flash crowd incident is the fact that even the web site administrators are not prepared most of the time to handle the special scenario of spontaneous bulk visits from people all over the world to any specific web page or web site.

This paper compares the performance between intrusion detection methods using sequence-order independent web browsing behavior of users based on PCA [1] with artificial intelligence techniques. The detection framework uses sequence order independent attributes rather than the sequence order of web page to construct behavior based attribute vector matrix. To model the browsing pattern, PCA is used which also reduces the spare data. Depending on the time interval attribute the browsing pattern is clustered using k-means clustering algorithm. Removal of spare data will improve the clustering speed. Threshold values of the clusters are used to find the deviation in behavior of the user as a normal or attacker. Principal Component Analysis (PCA) is used to model the web browsing behavior. The behavior pattern derived is clustered using K-means algorithm. The proposed method requires only 0.368 seconds at an average to perform the detection process. This rate is 0.126 seconds lesser than the existing HSMM method. The method also accurately identifies 94.4% of attacks in the application layer.

The paper then examines the performance of Artificial Intelligence based techniques for intrusion detection using the same dataset. The research work has made substantial contribution to the detection and mitigation of DDOS attacks and has also presented efficient techniques for automated detection of intrusions and anomalies in an open network environment. The remaining part of the paper is organized as follows: Section 2 involves the works related to probable solutions for attack prevention and intrusion detection in the Application layer. Section 3 involves the description of the proposed method. Section 4 involves the performance analysis of the proposed work. The paper is concluded in Section 5.

2. Literature Survey

Previous research works were mainly focused on detecting the Network layer attacks. Some of the techniques proposed to detect the Application layer attacks are discussed here. A novel approach called Double Guard for detecting intrusions in multitier web applications was proposed by Meixing et al (2012) [17]. Shui Yu et al (2012) [9] used flow correlation coefficient for discriminating DDoS attacks from flash crowds. However in order to disguise the flow finger prints, bot writers may include many attack packet generation functions and make each bot randomly choose one function to generate the attack packets. Gautam Thatte et al (2011)[18] proposed a bivariate parametric detection mechanism for anomaly detection in aggregate traffic. The

major limitation is that it addresses only low rate DDoS attacks. Hongxin Hu et al (2012) [19] implemented a visualization based fire wall policy analysis tool using rule based segmentation.

Patrick P. Tsang et al. (2011) [20] proposed a scheme called NYMBLE for blocking misbehaving users in anonymizing networks which provided multiple dynamic thresholds for resource usage indication. However the attackers can make use of this anonymizing network to launch a DDoS attack. Yi Xie et al (2009) [6] used hidden semi markov model for anomaly detection to monitor the application layer DDoS attack. ShuiYu et al. (2007)[21] proposed trace back of DDoS attacks using entropy variations. This method treats flash crowds as a DDoS attack and it results in false positive alarm. Ranjan et al (2004)[23] used statistical methods to detect characteristics of HTTP sessions. Yen et al (2005)[13] defended the application layer DDoS with constraint random requests by the statistical methods. Jung et al (2002) has applied two properties to distinguish the DDoS from flash crowds. Yuan et al (2005) used the cross correlation analysis to capture the traffic patterns and to decide where and when a DDoS attack possibly arises. Yi Xie et al (2005)[4] used Hidden semi-Markov model to detect application layer DDoS attacks for popular websites. Georgios et al proposed a model using three aspects of human behavior a) request dynamics b) request semantics and c) ability to process visual clues to differentiate DDoS bots from human users . Other existing defense methods may be those based on man-machine interaction, e.g. puzzles, passwords, and the CAPTCHAs. However Kandula et al (2004) and Ranjan et al. (2006)[23] have pointed out that those schemes are not effective for the DDoS attack detection because they may annoy users and introduce additional service delays.

S. Kandula et al (2005)[22] have designed a probabilistic authentication mechanism using CAPTCHAs acronym for “Completely Automated Public Turing Test to tell computers and Humans Apart”) to protect a web cluster from DDoS attack. Several DDoS defence mechanism has been designed by mining the behavior of a webuser. S. Burkle et al (2005) have developed a probabilistic model which uses a zipf-like distribution to model to page jump-probability. J. Velasquez et al (2003)[14] use web user’s usage patterns from the click streams data set and page content. However this method is not very suitable for on-line detection because it requires intensive computation for page content processing and data mining. The work proposed by D. Dhyani et al (2003) is based on Markov model. The Markov chains are used to model the URL access patterns that are observed during a web log.

3. Proposed Work

3.1 Principal Component Analysis Based Anomaly Detection

PCA is utilized in this work as an exploratory multivariate analysis technique. Web browsing behaviors are taken as an alternative of web page request sequence [8]. Sequence-order-independent technique is used for representation of web browsing behaviors. To model the web browsing behavior, Principal Component Analysis (PCA) is used. The behavior pattern derived is clustered using k-means algorithm. Browsing behaviors that are clustered is analyzed by threshold values to differentiate

the normal from the anomaly detection. Multivariate analysis is concerned with data sets that have more than one response variable for each observational unit. The data sets can be summarized by data matrices X with n rows and p columns, the rows representing the observations, and the columns the variables.

To detect the anomaly behavior [10], the clusters that are framed earlier are used and a threshold value is computed for those clusters. For a single user, the PCA b_i at the time σ_i is multiplied with the total number of users in the cluster T_{CUser} . This product value of a user is divided by the constant value 2. This process is carried out for all the users in the clusters. The Summation of all the values obtained for each user in a cluster as above is computed. The summation value that we obtained is the threshold for a cluster. Similarly, for all the clusters the values are computed. The division is carried out to obtain optimal value for threshold. The threshold is computed from the equation (1)

$$\text{Threshold} = \sum_{x \in T_{CUser}} \frac{\sigma t(b_i, 1) * T_{CUser}}{2} \quad (1)$$

User's b_i value is compared with the threshold. If b_i of a user is lesser or equal to the threshold, then it is concluded that the user is normal user. Otherwise, the user is an anomaly and his/her requests are dropped by the detector. Fig.1 shows the flow diagram for PCA based Approach.

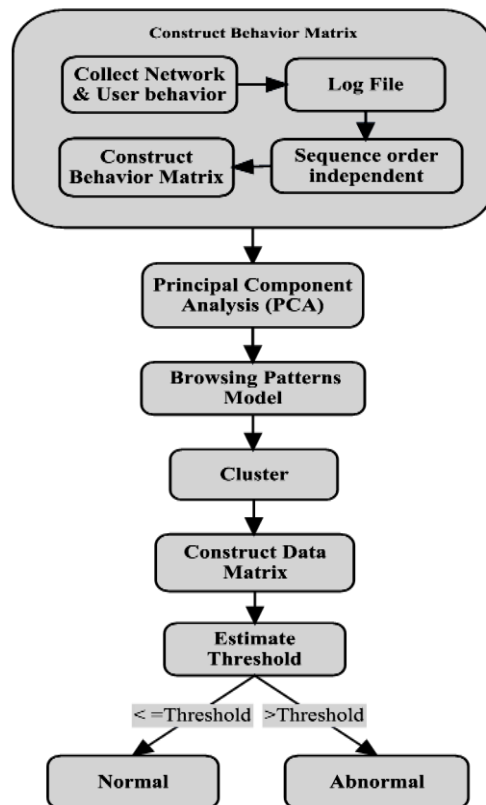


Fig.1 Flow Diagram for PCA based Approach

3.2 Fuzzy C Means Neural Network

The Fuzzy C Means Neural network detection process[16] is detailed below in fig 2. Fuzzy C-means (FCM) is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. It provides a method that shows how to group data points that populate some multidimensional space into a specific number of different clusters. The datasets are classified into two clusters. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade.

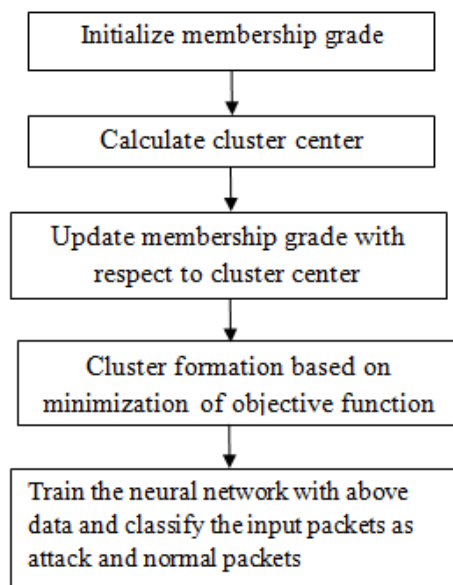


Fig.2 Steps in FCM-ANN

The FCM clustered output with two clusters namely attack and normal is shown in fig.3.

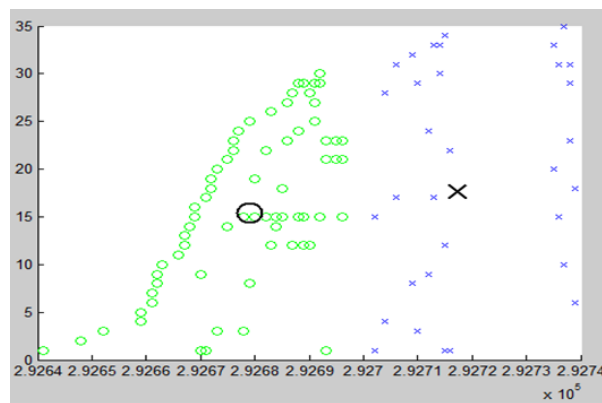


Fig. 3 FCM Clustered Output

The neural network is trained with 11 different training functions for 5 different types of neural networks namely patternet, feed forwardnet, cascade forwardnet, time delaynet and fitnet. The confusion matrix shows the percentages of correct and incorrect classifications. Correct classifications are referred as true positive and false negative. Incorrect classifications are true negative and false positive. The Fuzzy Inference System used is Mamdani's system. Two input membership functions with membership values of Low, Medium and High are defined. The two input functions considered are the number of webpages browsed by each user and the total number of web pages browsed by all users considered by our system.

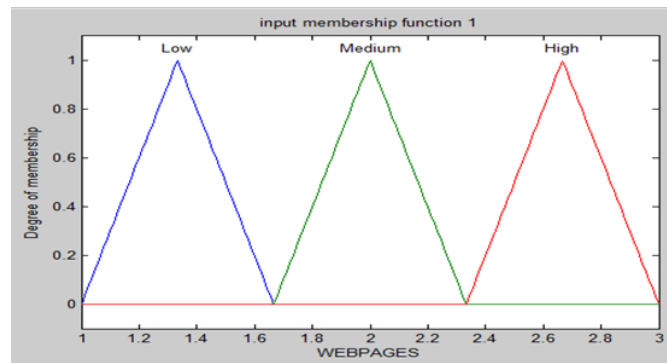


Fig.4 Membership function defined for input1

Fig. 4 shows the input membership function for the webpages browsed by an individual user. Three membership values namely LOW, MEDIUM and HIGH are defined. Fig.5 shows the membership values for the second input namely the total count of users who accessed a particular web page.

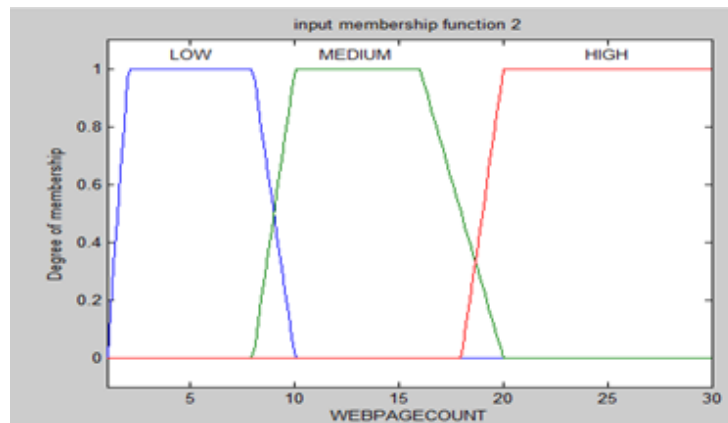


Fig.5 Membership function defined for input2

The output has two values namely attack and normal in the range of 0 to 1. The FIS has 9 rules and the defuzzification method used is centroid. Fig.6 shows the overall Fuzzy rule view.

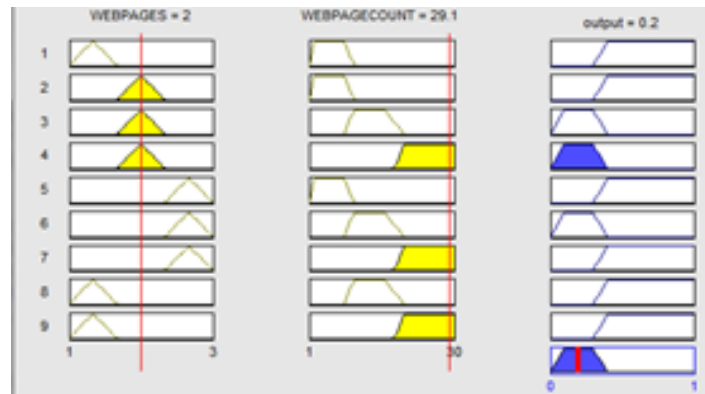


Fig.6 Fuzzy Rule View

4. Performance Analysis

The detection rate is given by the formula . Fig.7 shows the comparison of detection rate in % for the four techniques namely PCA,Neural Network approach, FCM-NN and Fuzzy when the number of users in the system considered are 50,100,150 and 200.

$$\text{Detection rate} = \frac{\text{Total detected attacks}}{\text{Total attacks}} * 100$$

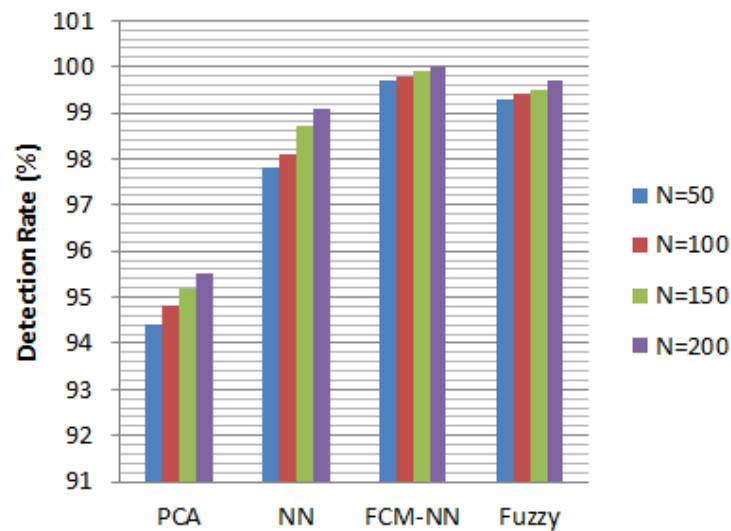


Fig 7. Technique Vs Detection rate (%)

From the comparative analysis of the four methods shown above, it is observed that the FCM-NN has a highest attack detection rate of 100% when the number of users considered in our system is 200.Fig.8 shows the RoC curve of the proposed FCM-ANN technique.

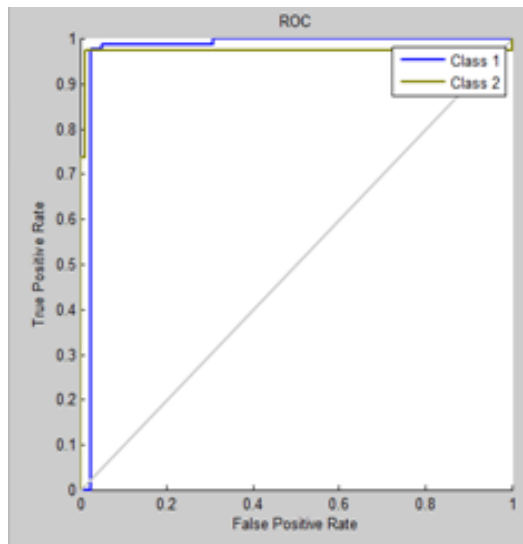


Fig.8 ROC of FCM-ANN

5. Conclusion

This paper focused on detecting Application layer DDoS attacks. We proposed a new model based on PCA that utilizes sequence-order-independent attributes rather than the web page sequence order. The proposed method is practical and efficient for detecting application layer DDoS attacks in real environments. To reliably validate our model, we generated three types of DDoS attacks. Our method detects attacks with an average detection rate of 94.4% and. In addition, the proposed FCM-NN model has the capability of detecting attacks with an average detection ratio of 100%. The proposed research work can be updated in future by training the FCM –NN algorithm to detect the dynamically launched webpage attacks online .

6. References

- [1] Bharathi, R. and Sukanesh, R. "A PCA based frame work for detection of Application layer DDos attacks" in WSEAS Ttransactions on Information science and Applications, Issue 12, Vol.9, pp.389-398,Dec 2012, E-ISSN 2224-3402.
- [2] Bharathi, R. and Sukanesh, R. "Flexible Approach to Perceive DDoS Attack from Flash Crowds in Dynamic Miscellaneous Traffic" in European Journal of Scientific Research ISSN:1450-216X Vol.72 No.3 (2012), pp. 338-347 © Euro Journals Publishing, Inc. 2012.
- [3] Bharathi, R. and Sukanesh, R. "Implementation of a Secured system with Roaming Server and Roaming Ports" in International Journal on Computer Science and Engineering (IJCSE) ISSN: 0975-3397, Vol. 3 No. 5, pp. 1781-1786, , May 2011.

- [4] Yi Xie and Shu-Zheng Yu, "Monitoring the Application -Layer DDoS attacks For Popular Websites", IEEE / ACM Transactions on Networking, Vol.17, No.1, pp.15-25, 2009.
- [5] Yi. Xie and S. Yu, "Measuring the normality of web proxies behaviour based on locality principles", Network and Parallel Computing, 2008.
- [6] Yi. Xie and Shun-Zheng.Yu, "A large scale hidden semi markov model for anomaly detection on user browsing behaviors", IEEE/ACM transaction on networking, February, 2009.
- [7] Jie.Yu, Chengfang Fang, Liming U and Zhoujun Li, "A lightweight mechanism to mitigate application layer DDoS attacks", springer, 2009.
- [8] S.Lee, G. Kim and S. Kim, "Sequence order independent network profiling for detecting application layer DDoS attacks", wireless communication and networking, 2011.
- [9] S. Yu, W. Jia, S. Guo, Y. Xiang and F. Tang, "Discriminating DDoS attacks from flash crowds using flow correlation coefficient", IEEE transaction on parallel and distributed systems, June, 2012.
- [10] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia Fernandez, and E. Vazquez, "Anomaly based network intrusion detection: Techniques, systems and challenges", Computers and security, 2009.
- [11] Yi. Xie, S. Tang, Y. Xiang, and J. Hu, "Resisting web proxy based HTTP attacks by temporal and spatial locality behavior", IEEE transaction on parallel and distributed systems, July, 2013.
- [12] M. A. Rassam, *et al.*, "A survey of intrusion detection schemes in wireless sensor networks," American Journal of Applied Sciences, vol. 9, p. 1636, 2012.
- [13] Yen, W. and Lee, M.F. "Defending application DDoS with constraint random request attacks", In Proc. Asia-Pacific Conf. Commun., Perth, Western Australia, pp. 620-624, 2005.
- [14] Velasquez, J., Yasuda, H. and Aoki, T. "Combining the web content and usage mining to understand the visitor behavior in a web site," in Proc. 3rd IEEE Int. conf. Data Mining (ICDM'03), pp. 669-672, 2003.
- [15] Bridges S, Vaughn R, "Fuzzy data mining and genetic algorithms applied to intrusion detection", Proceedings Twenty Third National Information Security Conference, Maryland, USA, October, 2000.
- [16] Lin Hongxi, "Forecasting Safety Stock of Supply Chain Using Artificial Neural Network", Journal of Minjiang University, Vol. 25 No. 2, 2004, pp.72-75.
- [17] Meixing., Stavrou, A. and Byung, B. "Double Guard: Detecting Intrusions in Multitier Web Applications", IEEE Trans. on Dependable and secure computing, Vol.9, No.4, pp.512-525, 2012.
- [18] GautamThatte, UrbashiMitra, and John Heidemann, Senior Member, IEEE, "Parametric Methods for Anomaly Detection in Aggregate Traffic "IEEE/ACM Transactions On Networking, Vol. 19, No. 2, April 2011

- [19] HongXin Hu, Gail-JoonAhn and KelanKulkarni,"Detecting and Resolving Firewall Policy Anomalies", IEEE Trans. on Dependable and secure computing Vol 9,N0.3,June 2012,pp318-331.
- [20] Patrick P. Tsang, ApuKapadia, Cory Cornelius, and Sean W. Smith," Nymble: Blocking Misbehaving Users in Anonymizing Networks", IEEE Transactions On Dependable And Secure Computing, Vol. 8, No. 2, March-April 2011,pp.256-269.
- [21] Shui Yu et al "Trace back of DDoS attacks using entropy variations "*IEEE Transactions On Depend able And Secure Computing*, Vol. 22, No. 3, March 2011 ,pp412-424
- [22] S. Kandula, D. Katabi, M. Jacob, and A. Berger, "Botz-4-Sale: Surviving Organized DDoS Attacks that Mimic Flash Crowds (Awarded Best Student Paper)," Proc. Second Symp.Networked Systems Design and Implementation (NSDI '05),2005.
- [23] Ranjan, S., Swaminathan, R., Uysal, M. and Knightly, E. "DDoS-resilient scheduling to counter application layer attacks under imperfect detection", In Proc. IEEE INFOCOM,2006 [Online]. Available:<http://www.ece.rice.edu/networks/papers/dos-sched.pdf>