Journal of Computer Science 8 (6): 983-986, 2012
ISSN 1549-3636

Application of Adaptive Neuro-Fuzzy Inference System for Information Secuirty

Altyeb Altaher, Ammar Almomani and Sureswaran Ramadass
National Advanced IPv6 Centre, Universiti Sains, Malaysia

Abstract: Problem statement: Computer networks are expanding at very fast rate and the number of network users is increasing day by day, for full utilization of networks it need to be secured against many threats including malware, which is harmful software with the capability to damage data and systems. Fuzzy rule based classification systems considered as an active research area in recent years, due to their unique capability of classifying. **Approach:** This study presents a neural fuzzy classifier based on Adaptive Neuro-Fuzzy Inference System (ANFIS) for malware detection. Firstly, the malware exe files was analyzed and the most important API calls were selected and used as training and testing datasets, using the training data set the ANFIS classifier learned how to detect the malware in the test dataset. **Results and Conclusion:** The performances of the Neuro fuzzy classifier were evaluated based on the performance of training and accuracy of classification, the results show that the proposed Neuro fuzzy classifier can detect the malware exe files effectively.

Key words: Adaptive Neuro-Fuzzy Inference System (ANFIS), fuzzy logic, malware detection

INTRODUCTION

Computer networks are expanding at a very fast rate and the number of network users is increasing day by day, for full utilization of networks it need to be secured against many threats including malware, which is harmful software with the capability to damage data and systems. The detection of malware and intruders becomes an important part of any modern network for guaranteeing the security issue of information system (Kim *et al*., 2011; Zhou *et al*., 2010; Beg *et al*., 2010; Altaher *et al*., 2011).

Technical reports from detection vendors increasingly warn about new malware and monitor the increased number of infected computer systems. McAfee released its Threat Report for the Fourth Quarter of 2011 which indicated that the number of malware increased continuously. In the Q4 2011 report, McAfee Labs detected approximately 9,300 new malicious sites every day, up from 6,500 per day in Q3. McAfee currently counts more than 700,000 active malicious URLs in its database McAfee Labs, 2011.

Neural networks and Neuro fuzzy techniques have been effectively used in various fields of science, e.g., Detection system, classification, prediction, intelligent systems and decision making.

Application of adaptive neural-fuzzy inference system for identification of malware portable executable files: This study presents a neural fuzzy classifier based on Adaptive Neuro-Fuzzy Inference system (Jang, 1993) for malware detection.Firstly, the malware exe files was analyzed and the most important API calls were selected and used as training and testing datasets, using the training data set the ANFIS classifier learned how to detect the malware in the test dataset.

Malware exe file analysis and feature extraction: The malware exe files were analyzed to extract the Application Programming Interface (API) as a feature to differentiate between the normal files and the malware files, then the API features were ranked to determine the most effective features which can reflect the behavior of the malware files. We used Information Gain Ratio method (IGR) algorithms which work based on the extraction of similarities between sets of e-mails and then gives the highest weight to the most effective features based on the class of Phishing and ham e-mails belonging to IGR (Mori, 2002), as explained in the following Eq 1.

$$\text{Gain\_r}(X,C) = \frac{\text{gain}(X,C)}{\text{spllit\_inf o}(C)} \quad (1)$$

where, gain_r (X, C) represents the gain ratio of the feature X frequency in class C Eq. 2:

Corresponding Author: Altyeb Altaher, National Advanced IPv6 Centre Universiti Sains Malaysia, Malaysia

$$\text{spllit}_{\text{inf o(C)}} = -\sum_i \left(\frac{|C_i|}{C} \right) \log |C_i| / |C| \quad (2)$$

where, Ci and |Ci| denote the frequency of features X in class C, the i-th sub-class of C and the number of features in Ci, respectively.

All the features selected to be used by the classifier were ranked using the information gain ratio method. The more the information gain is, the more helpful a feature will be in the differentiation between the malware and normal files.

Artificial Neuro-fuzzy inference system:

Fuzzy inference system: Fuzzy Inference Systems (FIS) are efficient techniques for studying the behavior of nonlinear systems using fuzzy logic rules. ANFIS is a Neuro-fuzzy system that uses the learning techniques of neural networks, with the efficiency of fuzzy inference systems (Esposito *et al*., 2000). ANFIS uses a hybrid learning algorithm to specify parameters of Sugeno-type fuzzy inference systems. It uses the least-squares method with the backpropagation gradient descent method train FIS membership function parameters simulate a given training data set. ANFIS can be called using optional parameters to validate the model.

ANFIS Architecture: ANFIS structure is similar to the neural network structure based on the Takagi Sugeno model, as illustrated in Fig. 1.

According to the Sugeno fuzzy model, rule sets are as follows:

- If x is A_1 and y is B_1 then $f_1 = p_1x+q_1y+r_1$
- If x is A_2 and y is B_2 then $f_2 = p_2x+q_2y+r_2$

Layer 1: Layer 1 is an input and falsification layer. Every node i in this layer is an adaptive node with a node function Eq. 3 and 4:

$$O_{1,i} = \mu_{Ai}(x) \text{ , for i = 1, 2} \quad (3)$$

$$O_{1,i} = \mu_{Bi-2}(y) \text{ , for i = 3, 4} \quad (4)$$

Layer 2: Layer 2 is the rule layer. Each node in this layer computes the impact of each rule through multiplication Eq. 5:

$$O_{2,i} = w_i = \mu_{Ai}(x) \cdot \mu_{Bi}(y), \ i = 1,2 \quad (5)$$

Layer 3: Layer 3 is normalization layer. Each neuron in this layer computes the normalized effect of a given rule Eq. 6:

$$O_{3,i} = \overline{w} = \frac{w_i}{w_1 + w_2}, \ i = 1,2 \quad (6)$$

Layer 4: Parameters in this layer are considered as consequent parameters Eq. 7:

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i), \ i = 1,2 \quad (7)$$

Layer 5: This layer is designed to calculate the sum of the output of all incoming signal Eq. 8:

$$O_{5,i} = \sum_i \overline{w} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (8)$$

Experiment and results: We used MATLAB version 7.10, for the implementation of the adaptive fuzzy inference system. In our test, we used datasets consist of 288 normal executable files and 416 malware executable files, the dataset downloaded from Nexginrc, 2010 and divided into two datasets, training and testing. The ANFIS classifier was trained using training dataset. Figure 2 shows that that training error and testing error were decaying as the number of epochs increased. The Membership functions were generated by ANFIS classifier as in Fig. 3.

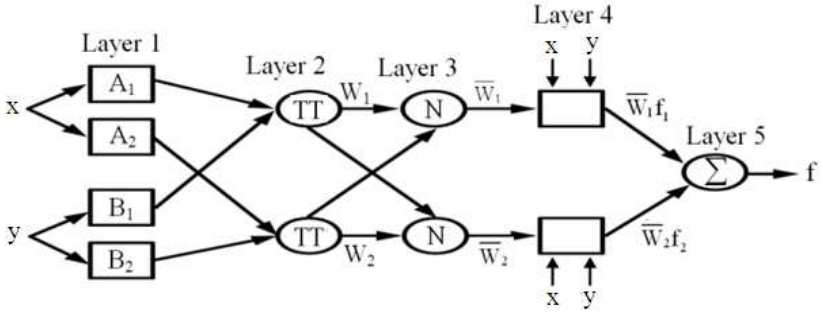

Fig. 1: ANFIS structure (type-3 ANFIS) (Jang, 1993)

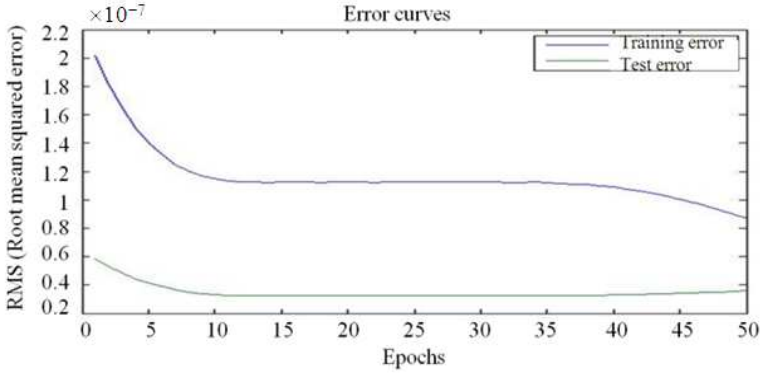

Fig. 2: The training and testing error of the developed ANFIS classifier

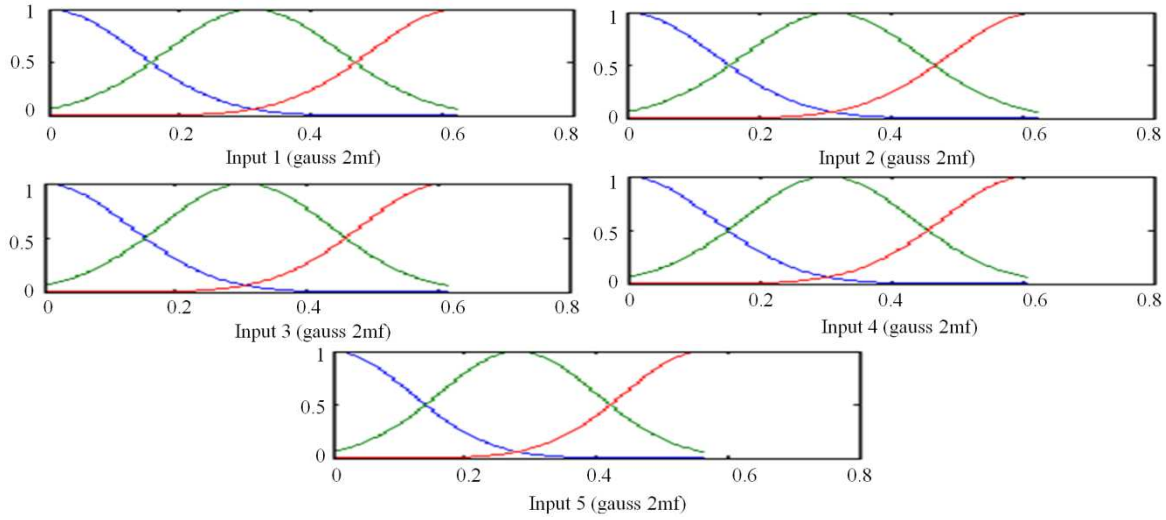

Fig. 3: The training membership functions generated by the developed ANFIS classifier

Based on the obtained results from Fig. 2, the Adaptive Neural Fuzzy Inference System (ANFIS) proved its capability to detect the malware executable files by decreasing the rate of testing and training errors while increasing the level of accuracy. The structure of ANFIS extends like for some to the structure of a neural network, which use the mapping of the input and output functions and related parameters can be used to interpret the input/output map. The parameters associated with the input membership functions will change through the learning process. Figure 3 shows the training membership functions generated by the developed ANFIS classifier.

CONCLUSION

The objective of this study was to develop an ANFIS classifier for malware exe file identification. It was observed that the ANFIS classifier learned how to detect the malware in the test dataset. The performances of the Neuro fuzzy classifier were evaluated based on the performance of training and accuracy of classification, the results show that the proposed Neuro fuzzy classifier can detect the malware exe files effectively.

REFERENCES

Altaher, A., S. Ramadass and A Ali, 2011. Computer virus detection using features ranking and machine learning. Australian J. Basic Applied Sci., 5: 1482-1486.

Beg, S., U. Naru, M. Ashraf and S. Moshin, 2010. Feasibility of intrusion detection system with high performance computing: A survey. Int. J. Adv. Comput. SCI., 1: 26-35.

Esposito, A., E.C. Ezin and C.A. Reyes-Garcia, 2000. Designing a fast Neuro-fuzzy system for speech noise cancellation. Lecture Notes Comput. Sci.

Jang, J.S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cyber., 23: 665-685. DOI: 10.1109/21.256541

Kim, W., O.R. Jeong, C. Kim and J. So, 2011. The dark side of the internet: Attacks, costs and responses. Inform. Syst., 36: 675-705. DOI: 10.1016/j. Is. 2010.11.003

Mori, T., 2002. Information gain ratio as term weight: The case of summarization of IR results. Proceedings of the 19th International Conference on Computational Linguistics, (COLING' 02), ACM, Stroudsburg, PA, USA., pp: 1-7. DOI: 10.3115/1072228.1072246

Zhou, C.V., C. Leckkie and S. Karunasekera, 2010. A survey of coordinated attacks and collaborative intrusion detection. Comput. Secu., 29: 124-140. DOI: 10.1016/j. Chose. 2009.06.008