Hybrid Neural Networks for Intrusion Detection System

Chaivat Jirapummin[1], Naruemon Wattanapongsakorn[1] and Prasert Kanthamanon[2]
[1]Department of Computer Engineering, Faculty of Engineering,
King Mongkut's University of Technology Thonburi,
91 Prach-Uthit Road, Bangkok 10140, Thailand
Tel. +66-2-470-9089, Fax.: +66-2-872-5050
[2]School of Information Technology
King Mongkut's University of Technology Thonburi,
91 Prach-Uthit Road, Bangkok 10140, Thailand
e-mails : naruemon@cpe.kmutt.ac.th, prasert@it.kmutt.ac.th

Abstract: Network based intrusion detection system is a computer network security tool. In this paper, we present an intrusion detection system based on Self-Organizing Maps (SOM) and Resilient Propagation Neural Network (RPROP) for visualizing and classifying intrusion and normal patterns. We introduce a cluster matching equation for finding principal associated components in component planes. We apply data from The Third International Knowledge Discovery and Data Mining Tools Competition (KDD cup'99) for training and testing our prototype. From our experimental results with different network data, our scheme archives more than 90 percent detection rate, and less than 5 percent false alarm rate in one SYN flooding and two port scanning attack types.

Keywords: Intrusion Detection System (IDS), Network Security, Self-Organizing Maps (SOM), Visualization, Cluster Matching, Resilient Propagation Neural Network (RPROP)

1. Introduction

Intrusion detection system (IDS) is used as a second line of defense in computer security measures. It can alert a network administrator when the network is attacked. Typically, data in network audit log is displayed in text format. In order to check for network intrusion activities, a network security officer may have to look for all data during the suspicious attack time, which is cumbersome and error prone.

In this paper, we present an alternative methodology for both visualizing intrusions by using the Kohonen's Self-Organizing Map (SOM), and classifying intrusions using Resilient Propagation Neural Network (RPROP). We gather major beneficial characteristics of both neural network models into our hybrid IDS, consisting of both unsupervised and supervised learning algorithms.

The rest of this paper is organized as follows. Section 2 discusses related works in IDS. Section 3 explains the concepts of SOM, RPROP and our cluster matching algorithms. Section 4 presents our experimental results and analysis. Lastly, section 5 provides a conclusion.

2. Related Works

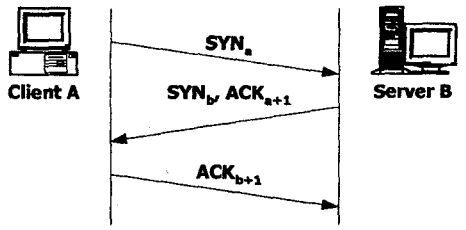
SOM approach is a relatively new choice for anomaly detection. Concerning intrusion detection, SOM is used as a postmortem or off-line analysis. Girardin [1] used SOM to visualize the network data and let the operator judge for anomaly packets. Hoglund et al. [2] used SOM as an anomaly detector to UNIX audit data. In addition, our previous work [3] introduced self- organizing map application for an IDS with visualization.

Lee and Heinbuch [4] used hierarchical backpropagation neural network to detect TCP SYN flooding and port scanning intrusions. There is also a combinational approach using backpropagation and expert system for an IDS [5]. Nevetheless, visualizing together with classifying intrusion data has not been introduced in any network IDS.

3. TCP SYN Flooding and Port Scanning Attacks

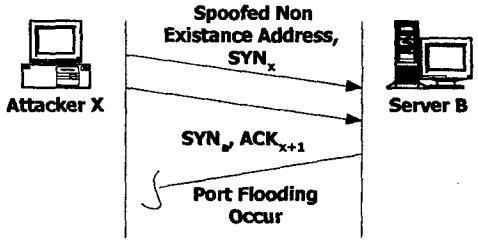
In our IDS, we focus on detection of network protocol attacks; TCP SYN flooding and port scanning which are probably the most common attacks. TCP SYN flooding is one of denial of service attacks. First, an attacker sends a large set of SYN packets to a server using unused IP address. Then, the server acknowledges these packets and waits for response which never arrives. Finally, the memory of the server becomes exhausted. We can compare a normal TCP 3-way handshake with a TCP SYN flooding handshake presented in Figures 1a and 1b. Readers can find more details of the TCP handshakes in Scuba et al [6].

Port scanning attack is a kind of probing or surveillance attacks. It does not intend to damage a system. However, it tries to gather information from a target network. There are many variants of scanning attacks in many protocols. We study TCP scanning attacks, which are depicted in Figures 2a and 2b. If the target ports are closed, the server sends reset packets. TCP connect differs from TCP half-connect in the third packet. In TCP connect, the attacker acknowledge the server response in third packet. For more information, Kanlayasiri et., al. [7] provides a good review in port scanning attacks.



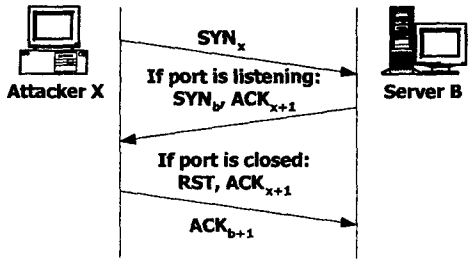
(a)

Figure 1. a) The Normal TCP 3-Way Handshake.



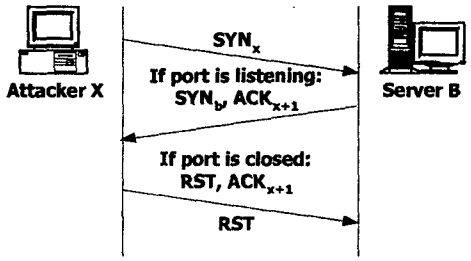
(b)

b) TCP SYN Flooding .



(a)

Figure 2. a) TCP Connect Scanning.



(b)

b) TCP Half-Connect Scanning.

4. Self-Organizing Map, RPROP Neural Network and Cluster Matching

Self-organizing map (SOM) [8] is an unsupervised neural network algorithm. In our experiments, we employ a batch version of SOM to cluster and visualize data. A SOM weight is adapted according to an average of input data in a Voronoi region, i.e. the data x(t) which has the same best matching unit in SOM feature map, as presented in Eq(1) and Eq(2).

$$\forall i, \quad \overline{x_i} = \frac{\sum_{x(t)\in V_i} x(t)}{n_i} \tag{1}$$

$$m_i^* = \frac{\sum_j n_j h_{ji} \overline{x}_j}{\sum_j n_j h_{ji}} \tag{2}$$

From Eq(1) $\overline{x_i}$ and n_i are mean value of data and number of data in each map unit according to the Voronoi set V_i, respectively. m_i^* denotes an equilibrium state of a map vector. Neighborhood function h_{ji} is gaussian. Distances of each map unit to each of its immediate neighbors are calculated and visualized by using gray scales of Unified Distance Matrix (U-Matrix) [9].

We use Resilient Propagation algorithm (RPROP) [10] as an intrusion classifier. It is an accelerated version of supervised Back-propagation neural network algorithm with the following weight updating rule, shown in Eq(3).

$$\Delta w_{ij}(t) = \begin{cases} -\Delta_{ij}(t) & , \ if \ \frac{\partial E}{\partial w_{ij}}(t) > 0 \\ +\Delta_{ij}(t) & , \ if \ \frac{\partial E}{\partial w_{ij}}(t) < 0 \\ 0 & , \ otherwise \end{cases} \tag{3}$$

In our hybrid scheme, output weight information from SOM is fed into the RPROP network, as shown in Figure 3. In this paper, we propose a cluster matching equation, Eq(4), to facilitate the interpretation of SOM. Our scheme is similar to the SOM adaptation rule in Eq(2).

$$\%matching_i = \frac{\sum_{j\in Nc} h_{ji} L_i}{\sum_{j\in Nc} h_{ji} n_j} \times 100 \tag{4}$$

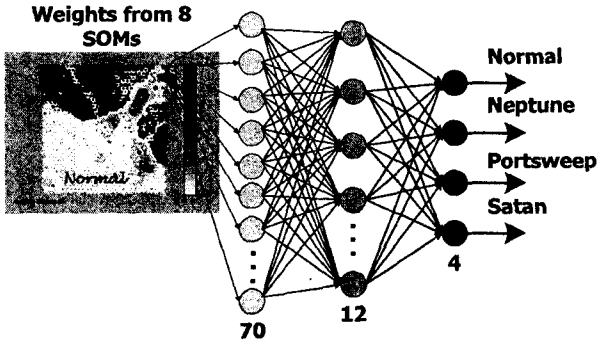


Figure 3. Hybrid Neural Network IDS.

where L_i is the i^{th} labeled unit located in a component plane according to a U-Matrix. h_{ji} is the neighborhood function of j^{th} unit in the component plane around i^{th} labeled unit. n_j is the number of neighborhood and kernel units, which are bounded by the threshold value of the component plane. Matching surface from Eq(4) is depicted in Figure 4. The peak of the gaussian function represents the exact matching unit, where the labeled unit in a component plane is located exactly at the same place as the one in the U-Matrix.

Matching percentage is decreased if the founded units located far away from the exact matching unit.

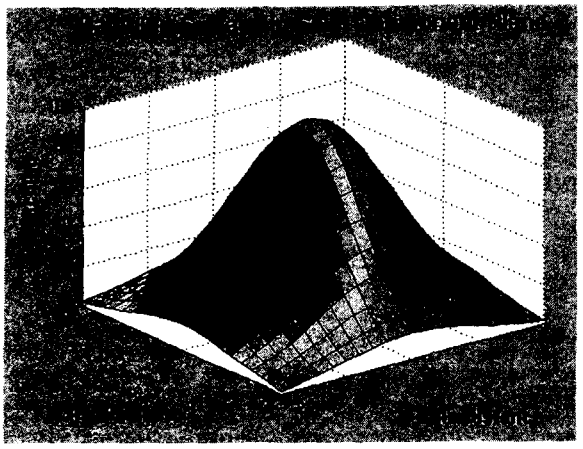


Figure 4. Matching Surface in SOM Component Plane.

5. Experimental Result

We select normal dataset, Neptune attack (SYN flooding), Portsweep attack (port scanning) and Satan attack (port scanning) datasets from [11] to train and test our IDS prototype. The datasets were already preprocessed by Lee et., al. [12] where readers can find the complete description of features. We divide 121,820 training data patterns equally into 8 sets. Each set is then clustered by a 1,234-unit SOM network. In RPROP setting, we use a 3-layer network, consisting of 70 neurons in the first hidden layer, 12 neurons in the second hidden layer and 4 neurons in the output layer, resulting to a 70-12-4 feed-forward neural network, as shown in Figure 3. The transfer functions for the first hidden layer, the second hidden layer and the output layer of RPROP are tan-sigmoidal, log-sigmoidal and log-sigmoidal, respectively. To achieve a reliable result, we perform 20 different trainings & testings.

There are two main testing datasets used in our experiment. Testset 1 contains 98,648 data, which was captured from the same network as the training data. Testset 2 includes 126,373 unseen normal and attack data from a different network. The average detection accuracy resulted from testset 2 is illustrated in Table 1. Nearly all Neptune attacks can be detected by our IDS system with very low false alarm rate. Portsweep and Satan attacks can be detected and correctly identified less than those from the Neptune attacks, more likely due to the insufficient datasets available for training and testing our IDS system. Testset 1 has a bit better detection accuracy when compared to testset 2, since the testing and training datasets were captured from the same network; more likely with some similar or routine network activities.

After tested, we use SOM to visualize the testing results in both Testset 1 and Testset 2. Figure 5 shows a U-Matrix of four labels. Shaded color in the vertical bar denotes a cluster border where features of a map unit are differentiated from their neighbors, while white color indicates a cluster center.

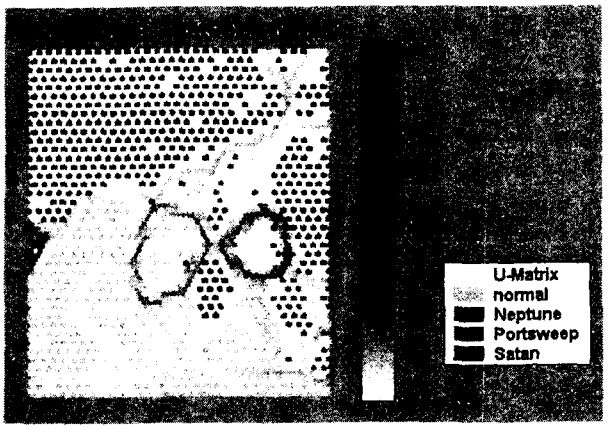


Figure 5. U-Matrix of Testset 1 Map 2.

Positions in every component plane and U-Matrix are associated with each other. Vertical bar in the component plane indicates approximated value of features in a SOM unit. Figures 6-9 display principle associated component planes in each class. The four most matching components are founded by sorting % matching in Eq(4), in descending order. Interpretation of each class characteristics is performed by the definition of principal associated components. Normal activities can be described as connections that use relevant services (same_srv_rate, dhst_same_srv_rate and dhst_count). They are fully opened and closed connections (SF), as shown in Figure 6. Neptune attacks are visualized as flooding activities (dhst_count), half-opened (S0) and SYN error connections (dhst_serror_rate and dhst_srv_serror_rate), shown in Figure 7. Portsweep attacks are illustrated as rejected connections (rerror_rate and REJ). They come from the same source port (dhst_same_src_port_rate), with the destination port scan slowly (dhst_diff_srv_rate), as shown in Figure 8. Satan attacks are portrayed as connections trying to scan a computer (dhst_diff_srv_rate and diff_srv_rate) rapidly (count), as shown in Figure 9.

Attacks	Detection Rate	False Alarm Rate
Neptune	99.7181	0.0591
Port Sweep	97.9123	4.1917
Satan	90.2811	4.4988

Table 1. IDS Simulation Results.

In our experiments, we perform both quantitative and qualitative analysis. The quantitative analysis is done by evaluating detection accuracies. From our IDS simulation results, as shown in Table 1, we achieves more than 90 % detection rate and less than 5 % false alarm rate in three selected attack programs.

Furthermore, we perform qualitative analysis by interpreting principal associated components of each attack using cluster matching. The knowledge that we gain from analyzing principal associated components can facilitate verification process of intrusion detection system.

6. Conclusion

In this paper, we proposed a hybrid neural network approach for IDS. We offered insightful visualization for network intrusion, using the clustering SOM approach. Then, we applied RPROP to classify suspicious network activities visualized initially by our SOM. Our IDS scheme is based on divide and conquer approach. We cover both qualitative analysis (by SOM) and quantitative analysis (by RPROP).

References

[1] L. Girardin, "An eye for Network Intruder-Administator Shootouts," Proc. of the 1st USENIX Workshop on Intrusion Detection and Network Monitoring, 1999.

[2] A.J. Hoglund et al."A Computer Host-based User Anomaly Detection System using Self-Organizing Map," Proc. of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Vol. 5, pp. 411-416, 2000.

[3] C. Jirapummin and N. Wattanapongsakorn, "Visual Intrusion Detection using Self-Organizing Maps", Proc. of Electrical and Electronic Conference (EECON-24), Thailand, Vol. 2, pp. 1343-1349, 2001.

[4] S.C. Lee and D.V. Heinbuch, "Training a Neural-Network Based Intrusion Detector to Recognize Novel Attacks," Information Assurance and Security, pp.40-46, 2000.

[5] J.M. Bonifacio et al., "Neural Networks Applied in Intrusion Detection Systems," Proc of IEEE Joint Conference on Neural Network, Vol. 1, pp. 205-210, 1998.

[6] C.L. Schuba, et al. "Analysis of Denial of Service Attack on TCP," Proc. of IEEE Symposium on Security and Privacy, pp 208-223, 1997.

[7] U. Kanlayasiri, "A Rule-based Approach for Port Scanning Detection," Electrical and Electronic Engineering Conference (EECON-23), Thailand, pp.148-153, 2000.

[8] T. Kohonen, *Self-Organizing Map*, 3rd edition , Springer Springer-Verlag, 501 2001.

[9] A. Ultsch and H.P. Siemon, "Kohonen's Self-Organizing Feature Maps for Exploratory Data Analysis," Proc. of International Neural Network Conference, pp. 305-308, 1990.

[10] M. Reidmiller et al. "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP algorithm," IEEE Inter Conf. on Neural Network, pp.586-591, 1993.

[11] S.J. Stolfo, et al., "KDD cup 1999 dataset," UCI KDD repository, http://kdd.ics.uci.edu

[12] W. Lee and S.J. Stolfo, "A Framework for Buliding Intrusion Detection Systems", Proceedings of the IEEE Symposium on Security and Privacy, p.120 –132, 1999.

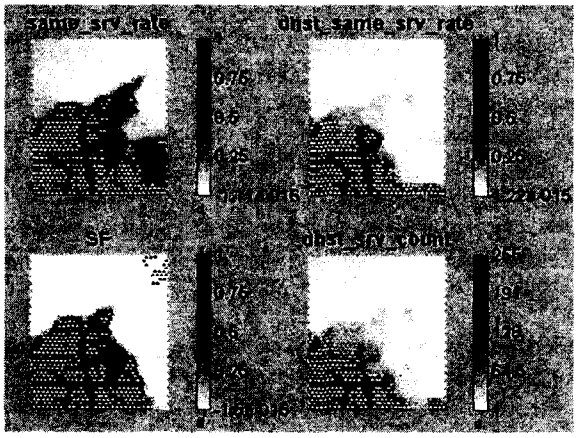


Figure 6 Principle Associated Component Planes of Normal Activities

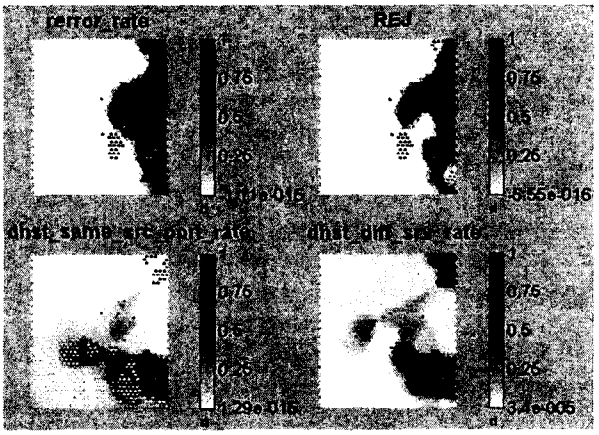


Figure 7 Principle Associated Component Planes of Portsweep Attacks

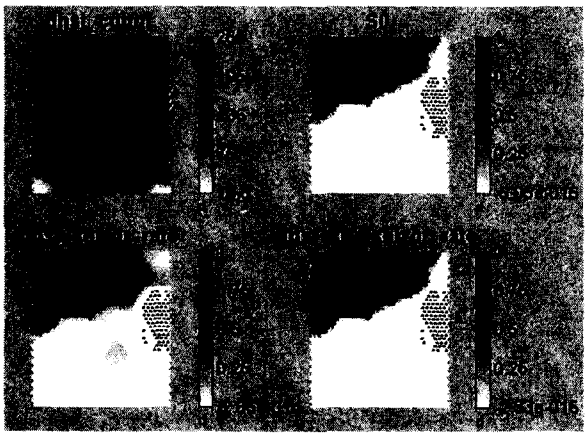


Figure 8 Principle Associated Component Planes of Neptune Attacks

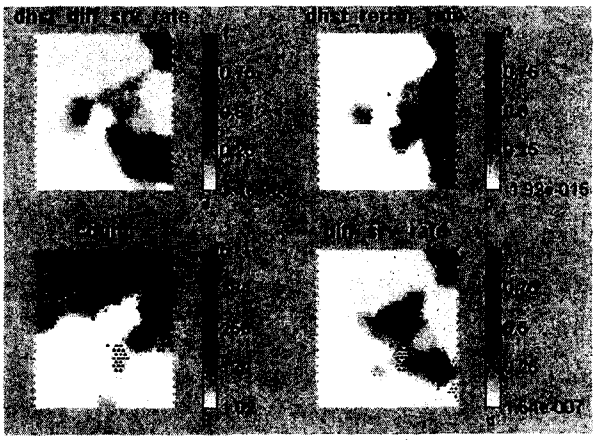


Figure 9 Principle Associated Component Planes of Satan Attacks