An Unsupervised Cooperative Pattern Recognition Model to Identify Anomalous Massive SNMP Data Sending

Álvaro Herrero, Emilio Corchado, José Manuel Sáiz

Department of Civil Engineering, University of Burgos, Spain
escorchado@ubu.es

Abstract. In this paper, we review a visual approach and propose it for analysing computer-network activity, which is based on the use of unsupervised connectionist neural network models and does not rely on any previous knowledge of the data being analysed. The presented Intrusion Detection System (IDS) is used as a method to investigate the traffic which travels along the analysed network, detecting SNMP (Simple Network Management Protocol) anomalous traffic patterns. In this paper we have focused our attention on the study of anomalous situations generated by a MIB (Management Information Base) information transfer.

1 Introduction

IDS are hardware or software systems that monitor the events occurring in a computer system or network, analysing them to automatically identify security problems.

Connectionist models have been identified as a very promising method of addressing the ID problem due to two main features [1]: their generalisation capability and their ability to classify patterns. Up to now, there have been several attempts to apply Artificial Neural Networks (ANN) (such as Self-Organising Maps [2], Elman Network [3]) to the network security field [4, 5].

Our IDS is based on a neural Exploratory Projection Pursuit (EPP) architecture. The aim of EPP [6, 7, 8, 9] is to reveal possible interesting structures hidden in the high-dimensional data so that a human can investigate the projections by eye.

2 A Novel Unsupervised Neural IDS Model

We can classify our IDS as a network-based [1] one because the data for the traffic analysis is obtained from the packets travelling along the whole network. This data can be extracted from the captured packets headers by using a network analyser.

We have developed a system for detecting anomalous traffic patterns; these include proper attacks and dangerous situations without being an attack.

The novel IDS model is structured as follows:

– **1st step**.- Network Traffic Capture: setting up one of the network interfaces as "promiscuous" mode, it can capture all the packets traveling along the network.

- **2nd step**.- Data Pre-processing: the captured data is pre-processed (see Section 3).
- **3rd step**.- Data Classification: once the data has been captured and pre-processed, the connectionist model presented below is used to analyse the data and to identify the anomalous patterns.
- **4th step**.- Result Display: this visualization tool displays data projections highlighting anomalous situations clearly enough to alert the network administrator, taking into account aspects as the traffic density or "anomalous" directions.

The Data Classification and Result Display steps performed by this IDS model are based on the use of a neural EPP architecture called Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [10, 11]. It was initially applied to the field of Artificial Vision to identify local filters in space and time [10, 11]. It is based on the neural architecture called Maximum Likelihood Hebbian Learning (MLHL) [8, 9]. The final neural model (CMLHL) can be described as follows: consider an N-dimensional input vector, x, and an M-dimensional output vector, y, with W_{ij} being the weight linking input j to output i and let η be the learning rate.

$$\text{Feed forward: } y_i = \sum_{j=1}^{\mathbf{N}} W_{ij} x_j, \forall i \ . \tag{1}$$

$$\text{Lateral activation passing: } y_i(t+1) = \left[y_i(t) + \tau(b - Ay)\right]^+ . \tag{2}$$

$$\text{Feedback: } e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i \ . \tag{3}$$

$$\text{Weight change: } \Delta W_{ij} = \eta . y_i . sign(e_j) \mid e_j \mid^{p-1} . \tag{4}$$

We use the standard MLHL with lateral connections. These lateral connections [10, 11] have been derived from the Rectified Gaussian Distribution [12] and applied to the negative feedback network [13]. The resultant net [10, 11] can find the independent factors of a data set but do so in a way which captures some type of global ordering in the data set.

3 Real Intrusion Detection Scenario Specific Data Set

Among all the implemented network protocols, there are some of them that can be considered quite dangerous for the network security. Among those, we have focused our effort in the study of SNMP because an attack based on this protocol may severely compromise the network security.

In the short-term, SNMP was oriented to manage nodes in the Internet community [14] and the MIB can be defined as a database which contains information about some elements or devices that can be network-controlled. The data set used in this work contains a transfer of some information contained in a SNMP MIB. This kind of transfer is considered a quite dangerous situation because a person having some free tools, some basic SNMP knowledge and the community password (in SNMP v. 1 and v. 2) can come up with all sorts of interesting and sometimes useful information.

In the Data Pre-processing step the system selects packets based on UDP (User Datagram Protocol). In this step, the system also performs a data selection and only the following 5 variables (extracted from the packet headers) are used: **timestamp** (the time when the packet was sent), **protocol** (we have codified all the protocols contained in the data set), **source port** (the port number of the source host which sent the packet), **destination port** (the port number of the destination host where the packet is sent) and **size** (total packet size in Bytes).

4 Experimental Results, Conclusions and Future Work

Through a simple visual analysis of the figure Fig. 1.a, it is easy to identify several packet groups. Two of them (Groups 1 and 2 in Fig. 1.a) are different from other groups related to normal traffic as it is explained below. The packets belonging to each protocol contained in the data set are identified and visualized in the same group in Fig. 1.a, except in the case of SNMP packets (this case is explained later).

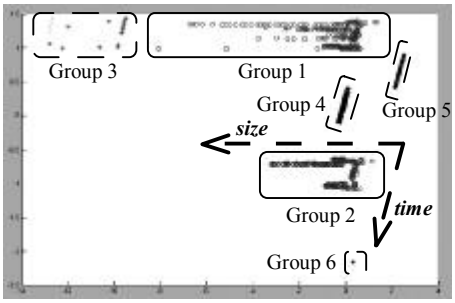

Fig. 1.a. Data projection displayed by the our neural IDS model

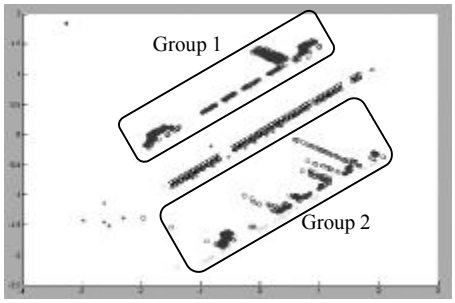

Fig. 1.b. Best Maximum Likelihood Hebbian Learning projection

After an analysis (labeling and studying most of the represented points) of the packets belonging to Groups 1 and 2 (Fig. 1.a) we have identified several features:

- These Groups are related to the SNMP MIB transfer mentioned above. They contain packets sent and received during the transfer embedded in the data set. All the packets belonging to SNMP are contained in one of these two groups.
- Group 1 contains all the traffic going from destination to source, while Group 2 contains all the traffic in the other way (from source to destination).
- Each group extends over two main axes: one related to the packet size and the other related to the timestamp.

We have labeled Groups 1 and 2 (Fig. 1.a) as anomalous ones due to two combined issues: the high temporal concentration of packets, and because they are made up of different size packets, situation related to the MIB information transfer.

This IDS model has been previously used to identify a SNMP port sweep [5] and it worked properly, identifying the anomalous situation in a very clear way.

We have applied different ANN such as Principal Component Analysis (PCA) [15, 16] or MLHL (Fig. 1.b) to the same data set. PCA is not able to detect the anomalous situation contained in the data set, because it shows the “anomalous” packets in the same way in which the rest of the traffic is shown. Fig. 1.b shows how MLHL is

capable of detecting the anomalous situation (Groups 1 and 2) but it is not detected as clearly as by using CMLHL (Fig. 1.a).

As conclusions, there are some issues to highlight: The visualization tool can show the packets grouped by their protocol and only the network administrator has the authority to decide whether a situation classified as anomalous is dangerous or not.

Future work will have the following work lines: the application of grid computation [17] in both Data Classification and Result Display steps and the use of distributed systems based on agents and multiagents and the exchange of information.

Acknowledgments

This research has been supported by the McyT projects: TIN2004-07033.

References

1. Planquart, J-P.: Application of Neural Networks to Intrusion Detection. Information Security Reading Room - SANS (SysAdmin, Audit, Network, Security) Institute (2002)
2. Hätönen, K., Höglund, A.,Sorvari, A.: A Computer Host-Based User Anomaly Detection System Using the Self-Organizing Map. International Joint Conference of Neural Networks (2000)
3. Ghosh, A. Schwartzbard, A., Schatz, A.: Learning Program Behavior Profiles for Intrusion Detection. Workshop on Intrusion Detection and Network Monitoring (1999)
4. Debar, H., Becker, M., Siboni, D.: A Neural Network Component for an Intrusion Detection System. IEEE Symposium on Research in Computer Security and Privacy (1992)
5. Corchado, E., Herrero, A., Baruque, B., Saiz, J.M.: Intrusion Detection System Based on a Cooperative Topology Preserving Method. International Conference on Adaptive and Natural Computing Algorithms. SpringerComputerScience. SpringerWienNewYork, (2005)
6. Friedman, J., Tukey, J.: A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers (23) (1974) 881-890
7. Hyvärinen, A.: Complexity Pursuit: Separating Interesting Components from Time Series. Neural Computation 13 (2001) 883-898
8. Corchado, E., MacDonald, D., Fyfe, C.: Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. Data Mining and Knowledge Discovery. Kluwer Academic Publishing 8(3) (2004) 203-225
9. Fyfe, C., Corchado, E.: Maximum Likelihood Hebbian Rules. ESANN(2002)
10. Corchado, E., Han, Y., Fyfe, C.: Structuring Global Responses of Local Filters Using Lateral Connections. JETAI 15 (4) (2003) 473-487
11. Corchado, E., Fyfe, C.: Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. International Journal of Pattern Recognition and Artificial Intelligence 17(8) (2003) 1447-1466
12. Seung, H.S., Socci, N.D., Lee, D.: The Rectified Gaussian Distribution. Advances in Neural Information Processing Systems 10 (1998) 350
13. Fyfe, C.: A Neural Network for PCA and Beyond. Neural Processing Letters 6 (1996)
14. Case, J., Fedor, M.S., Schoffstall, M.L., Davin, C.: Simple Network Management (SNMP). RFC-1157 (1990)
15. Postel, J.: IAB Official Protocol Standards. RFC-1100 (1989)
16. Oja, E.: Neural Networks, Principal Components and Subspaces. International Journal of Neural Systems 1 (1989) 61-68
17. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. 1st edition. Morgan Kaufmann Publishers (1998)