# Detecting Compounded Anomalous SNMP Situations Using Cooperative Unsupervised Pattern Recognition

Emilio Corchado, Álvaro Herrero, José Manuel Sáiz

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

**Abstract**. This research employs unsupervised pattern recognition to approach the thorny issue of detecting anomalous network behavior. It applies a connectionist model to identify user behavior patterns and successfully demonstrates that such models respond well to the demands and dynamic features of the problem. It illustrates the effectiveness of neural networks in the field of Intrusion Detection (ID) by exploiting their strong points: recognition, classification and generalization. Its main novelty lies in its connectionist architecture, which up until the present has never been applied to Intrusion Detection Systems (IDS) and network security. The IDS presented in this research is used to analyse network traffic in order to detect anomalous SNMP (Simple Network Management Protocol) traffic patterns. The results also show that the system is capable of detecting independent and compounded anomalous SNMP situations. It is therefore of great assistance to network administrators in deciding whether such anomalous situations represent real intrusions.

## 1 Introduction

Intrusion Detection Systems (IDS) are tools designed to monitor the events occurring in a computer system or network, analysing them to detect suspicious patterns that may be related to a network or system attack. They have become a necessary additional tool to the security infrastructure as the number of network attacks has risen very sharply over recent years.

There are currently several techniques used to implement IDS. Some are based on the use of expert systems (containing a set of rules that describe attacks), signature verification (where attack scenarios are converted into sequences of audit events), petri nets (where known attacks are presented with graphical petri nets) or state-transition diagrams (representing attacks with a set of goals and transitions). One of the main disadvantages of these techniques is the fact that new attack signatures are not automatically discovered without updating the IDS.

Connectionist models have been identified as very promising methods of addressing the ID problem due to two key features: they are suitable to detect day-0 attacks and they are able to classify patterns (attack classification, alert validation). There have recently been several attempts to apply artificial neural architectures [1, 2] (such as Self-Organising Maps [3, 4] or Elman Network [5]) to the field of network security. This paper presents an IDS based on a neural architecture that has never before been applied to the problem of ID.

## 2 The Cooperative Unsupervised IDS Model

Exploratory Projection Pursuit (EPP) [6, 7, 8, 9] is a statistical method for solving the complex problem of identifying structure in high dimensional data. It is based on the projection of the data onto a lower dimensional subspace in which its structure is searched by eye. It is necessary to define an "index" to measure the varying degrees of interest generated by each projection. Subsequently, the data is transformed by maximizing the index and the associated interest. From a statistical point of view the most interesting directions are those that are as non-Gaussian as possible.

The Data Classification and Result Display steps performed by this IDS model are based on the use of a neural EPP model called Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [10, 11, 12]. It was initially applied to the field of Artificial Vision [10, 11] to identify local filters in space and time. Here, we have applied it to the field of Computer Security [2, 13, 14]. It is based on Maximum Likelihood Hebbian Learning (MLHL) [8, 9]. Consider an N-dimensional input vector, $\mathbf{x}$, and an M-dimensional output vector, $\mathbf{y}$, with $W_{ij}$ being the weight linking input $j$ to output $i$ and let $\eta$ be the learning rate. MLHL can be expressed as:

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

The activation ($e_j$) is fed back through the same weights and subtracted from the input:

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \forall j . \tag{2}$$

Weight change:

$$\Delta W_{ij} = \eta . y_i . sign(e_j) | e_j |^{p-1} . \tag{3}$$

Lateral connections [10, 11] have been derived from the Rectified Gaussian Distribution [15] and applied to the MLHL. The resultant net can find the independent factors of a data set but do so in a way that captures some type of global ordering in the data set. So, the final CMLHL model is as follows:

Feed forward step: Equation (1)

Lateral activation passing: $y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ .$ **(4)**

Feed back step: Equation (2)

Weight change: Equation (3)

Where: $\eta$ is the learning rate, $\tau$ is the "strength" of the lateral connections, *b* is the bias parameter and $p$ is a parameter related to the energy function [8, 9, 11].
Finally *A* is a symmetric matrix used to modify the response to the data. Its effect is based on the relation between the distances among the output neurons.

## 3 Model Structure

The aim of this research is to design a system capable of detecting anomalous situations within a computer network. The information analysed by our system is obtained from the packets that travel along the network, meaning that it is a Network-Based IDS. The data needed to analyse the traffic is contained on the captured packets headers, obtained using a network analyser.

The structure of the IDS model is described as follows:

First step.- Network Traffic Capture: one of the network interfaces is set up in "promiscuous" mode. It captures all the packets travelling along the network.

Second step.- Data Pre-processing: the captured data is pre-processed and used as an input data in the following stage.

Third step.- Data Classification: once the data has been pre-processed, the connectionist model (section 2) analyses the data and identifies anomalous patterns.

Fourth step.- Result Display: the last step is related to the visualization stage. Finally the output is presented to the network administrator.

## 4 Real Data Sets Containing Compounded and Independent Anomalous SNMP Situations

We have decided to study anomalous SNMP situations because an attack based on this protocol may severely compromise system security [17]. CISCO [18] ranked the top five most vulnerable services in order of importance, and SNMP was one of them. In the short-term, SNMP was oriented to manage nodes in the Internet community [19].

Our efforts have focussed on the study of two of the most dangerous anomalous situations related to SNMP [2, 13, 14]:

SNMP port sweep: it is a scanning of network computers for the SNMP port using sniffing methods. The aim is to make a systematic sweep within a group of hosts to verify if SNMP is active in any port. Both default port numbers (161 and 162) and random port number (3750) are used.

MIB information transfer: the MIB (Management Information Base) can be defined in broad terms as the database used by SNMP to store information about the elements that it controls. This situation is a transfer of some information contained in the SNMP MIB. This kind of transfer is potentially quite a dangerous situation because anybody who possesses some free tools, some basic SNMP knowledge and the community password (in SNMP v. 1 and SNMP v. 2) will be able to access all sorts of interesting and sometimes useful information.

In this work, the IDS analysed three different data sets:

1st Data set (Fig 1): this includes an example of each one of the anomalous situations defined above: an SNMP port sweep and an MIB information transfer. We have called this a compounded anomalous SNMP situation because it involves simple but different anomalous events that occur at the same time.

2nd Data set (Fig 2.a): this contains an example of an SNMP port sweep situation (an independent anomalous SNMP situation).

3rd Data set (Fig 2.b): an example of an MIB information transfer situation (another independent anomalous SNMP situation).

In addition to the SNMP packets, these data sets contain traffic related to other protocols installed in our network, such as NETBIOS and BOOTPS.

In the Data Pre-processing step, the system performs a data selection from all of the captured information. As a result, all of the above-mentioned data sets contain the following five variables extracted from the packet headers: timestamp (the time when the packet was sent in relation to the first one), protocol (all the protocols contained in the data set have been codified, taking values between 1 and 35), source port (the port number of the source host that sent the packet), destination port (the destination host port number to which the packet is sent) and size (total packet size in Bytes).

## 5 Results, Conclusions and Future Work

Scatterplot Matrix is used to analyse pairwise relationships between variables in high dimensional data sets. Each factor pair highlights different structures or clusters in the projections of the same data set. It was used to analyse the results obtained from the connectionist IDS model. The system identified (Fig 1.a) the two anomalous situations contained in the real compounded data set. The analysis took account of such aspects as traffic density or "anomalous" traffic directions.

Factor pair 2-1 (Fig 1.a) contains the best representation of this anomalous situation, where the horizontal axe is related with the time feature and the vertical axe represents a combination of the protocol and size features. There are several issues to highlight about this figure: *Group 1* (Fig 1.a) identifies the sweep by means of normal and abnormal directions. It is clear that packets contained in this group do not progress in the same direction as the rest of packets groups (related to normal situations). On the other hand, *Groups 2* and *3* (Fig. 1.a) bring together packets related to the MIB information transfer. These groups are identified as anomalous due to their high temporal packets concentrations.

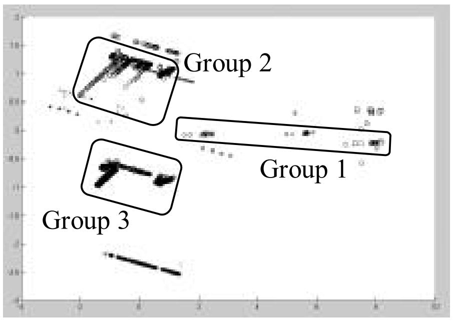

**Fig. 1.a**. Scatterplot Matrix factor pair 2-1 generated by the model for the 1st data set

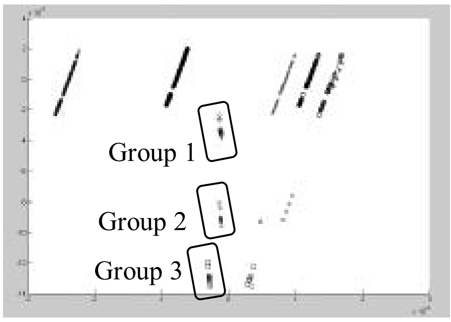

**Fig. 1.b**. PCA projection for the 1st data set

We have applied different connectionist methods such as Principal Component Analysis (PCA) [20] (Fig. 1.b) or MLHL to the same data set. CMLHL provides more sparse projections than the others [11]. CMLHL is able of identifying both anomalous situations while PCA (Fig. 1.b) is only able to identify the sweep (*Groups 1, 2 and 3*).

On the other hand, as can be seen in Fig. 2.a and Fig. 2.b, the neural IDS is capable of identifying both anomalous situations independently. The following figures (Fig. 2.a and Fig 2.b) show how the system performs successfully in those cases where there is only one anomalous situation within normal ones (2nd and 3rd Data Sets). In Fig 2.a we have identified the sweep (*Groups 1, 2 and 3*) by means of normal/abnormal direction and in Fig 2.b we have identified the MIB transfer (*Groups 1 and 2*) by means of high temporal concentration of packets.

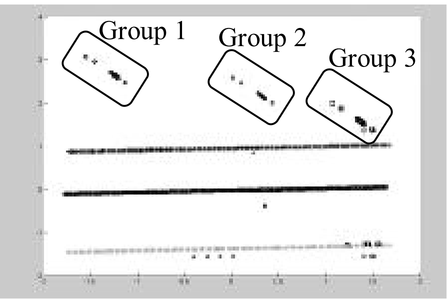

**Fig. 2.a**. Independent SNMP anomalous situation by a port sweep (2nd data set)

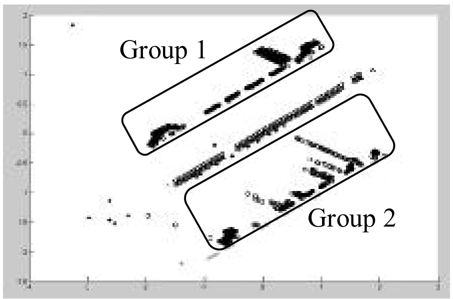

**Fig. 2.b**. Independent SNMP anomalous situation by a MIB transfer (3rd data set)

This research demonstrates the effectiveness and robustness of this novel IDS due to its capability to identify anomalous situations in two different ways: whether or not they are contained in the same data set. In summary, the connectionist IDS described in this paper is able to identify both independent and compounded anomalous SNMP situations showing its capability for generalization.

The visualization tool used in the Result Display step, shows data projections that highlight anomalous situations sufficiently clearly to alert the network administrator, taking into account such aspects as traffic density or “abnormal” directions.

One of the most common IDS techniques is the one called signature verification [20b]. Most of signature verification systems use pattern matching algorithms based on previously established rules included in a database. To reduce the number of posterior false alarms, this database should be adapted to the work environment by studying the traffic patterns that circulate along the network segment where the IDS is set up. One disadvantage of this method is the high processing time consume. This can be reduced by speeding up the packets analysis [21]. In comparison with this method, the advantages of our novel neural IDS are the following: it does not require any previous knowledge in the form of rules and it is able to detect unknown attacks day-0 ones.

Further work will be focused on the application of GRID [22] computation with more complex data sets and the use of multi-agent distributed systems.

## References

1. 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)
2. Corchado, E., Herrero, A., Baruque, B., Sáiz, J.M.: Intrusion Detection System Based on a Cooperative Topology Preserving Method. International Conference on Adaptive and Natural Computing Algorithms. Springer Computer Science. SpringerWienNewYork (2005)
3. 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)
4. Zanero, S., Savaresi, S.M.: Unsupervised Learning Techniques for an Intrusion Detection System. ACM Symposium on Applied Computing (2004) 412-419
5. Ghosh, A., Schwartzbard, A., Schatz, A.: Learning Program Behavior Profiles for Intrusion Detection. Workshop on Intrusion Detection and Network Monitoring (1999)
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. European Symposium on Artificial Neural Networks (2002)
10. Corchado, E., Han, Y., Fyfe, C.: Structuring Global Responses of Local Filters using Lateral Connections. Journal of Experimental and Theoretical Artificial Intelligence 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. Corchado, E., Corchado, J.M., Sáiz, L., Lara, A.: Constructing a Global and Integral Model of Business Management Using a CBR System. 1st International Conference on Cooperative Design, Visualization and Engineering (2004)
13. Herrero, A., Corchado, E., Sáiz, J.M.: A Cooperative Unsupervised Connectionist Model Applied to Identify Anomalous Massive SNMP Data Sending. 1st International Conference on Natural Computation (2005) ("*In press*")
14. Herrero, A., Corchado, E., Sáiz, J.M.: Identification of Anomalous SNMP Situations Using a Cooperative Connectionist Exploratory Projection Pursuit Model. 6th International Conference on Intelligent Data Engineering and Automated Learning (2005) ("*In press*")
15. Seung, H.S., Socci, N.D., Lee, D.: The Rectified Gaussian Distribution. Advances in Neural Information Processing Systems 10 (1998) 350-356
17. Myerson, J.M.: Identifying Enterprise Network Vulnerabilities. International Journal of Network Management 12 (2002)
18. Cisco Secure Consulting: Vulnerability Statistics Report (2000)
19. Case, J., Fedor, M.S., Schoffstall, M.L., Davin, C.: Simple Network Management (SNMP). RFC-1157 (1990)
20. Oja, E.: Neural Networks, Principal Components and Subspaces. International Journal of Neural Systems 1 (1989) 61-68
21. Aldwairi, M., Conte, T., Franzon, P.: Configurable string matching hardware for speeding up intrusion detection. ACM SIGARCH Computer Architecture News 33(1) (2005)
22. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. 1$^{t}$ edn. Morgan Kaufmann Publishers (1998)