# A Comparison of Neural Projection Techniques Applied to Intrusion Detection Systems

Álvaro Herrero[1], Emilio Corchado[1], Paolo Gastaldo[2], and Rodolfo Zunino[2]

[1] Civil Engineering Department
University of Burgos
C/ Francisco de Vitoria s/n, 09006, Burgos, Spain
{ahcosio,escorchado}@ubu.es
[2] Department of Biophysical and Electronic Engineering (DIBE)
Genoa University
Via Opera Pia 11a, 16145 Genoa, Italy
{paolo.gastaldo,rodolfo.zunino}@unige.it

**Abstract.** This paper reviews one nonlinear and two linear projection architectures, in the context of a comparative study, which are used as either alternative or complementary tools in the identification and analysis of anomalous situations by Intrusion Detection Systems (IDSs). Three neural projection models are empirically compared, using real traffic data sets in an IDS framework. The specific multivariate data analysis techniques that drive these models are able to identify different factors or components by studying higher order statistics - variance and kurtosis - in order to display the most interesting projections or dimensions. Our research describes how a network manager is able to diagnose anomalous behaviour in data traffic through visual projection of network traffic. We also emphasize the importance of the time-dependent variable in the application of these projection methods.

**Keywords:** Unsupervised Learning, Neural Networks, Exploratory Projection Pursuit, Auto-Associative Back-Propagation, Principal Component Analysis, Computer Network Security, Visualization, Intrusion Detection.

## 1 Introduction

An Intrusion Detection System (IDS) is designed to monitor computer systems or network events and to detect undesired and unauthorised entries, mainly via the internet. An IDS has become a necessary additional tool to the security infrastructure of a computer system as network attacks have risen dramatically over recent years.

Our research addresses the use of projection methods as either an alternative or a complementary tool that allows the network administrator to visualize traffic data patterns. In complex clustering domains, some data sets may hide their own structures. Projection models [1] are used as tools to identify and remove correlations between problem variables, which enables us to carry out dimensionality reduction, visualization or exploratory data analysis. These tools search for interesting

F. Sandoval et al. (Eds.): IWANN 2007, LNCS 4507, pp. 1138–1146, 2007.

projections or dimensions based on the analysis of different statistical features, such as Principal Component Analysis (PCA) [2], [3] and Exploratory Projection Pursuit (EPP) [1], [4], among others.

The projection system that we propose exploits projection models to arrive at a compact visualization of traffic evolution. The resulting IDS is designed to assist the network manager by providing an effective visual tool to detect anomalous situations based on the identification of high temporal distributions of the packets, moving in non-parallel or unorthodox directions to the normal ones. One of the main benefits of these neural network-based models is their ability to identify new attacks, known as "day-0 attacks", without updating the IDS.

This paper reports a review of one nonlinear and two linear neural projection models that can all be effectively applied to an IDS, which is followed by a comparative study of their respective performances. Following this introduction, section 2 goes on to introduce the projection models under examination; section 3 describes the data sets used in the experiments; and, finally, section 4 discusses the results, puts forward a number of conclusions and pointers for future work.

## 2 Unsupervised Neural Projection Models

Several attempts have been made to apply Neural Architectures (such as Self-Organising Maps [5], [6] Multilayer Perceptron [7], Radial Basis Function Networks [8]) to the field of network security [9], [10]. Most of them have focused on a classificatory approach to the intrusion detection task. A different approach is followed in this research, in which the main goal is to provide the network administrator with a snapshot of the network traffic, not only to detect anomalous situations but also to visualize protocol interactions and traffic volume. Three different models dealing with this issue are applied and their performance has been compared.

Unsupervised learning was chosen for this research, because in a real-life situation there is no target reference with which to compare the response of the network. The use of this kind of learning is very appropriate, for instance in the case of identifying "day-0 attacks".

### 2.1 Principal Component Analysis

PCA originated in work by Pearson [2], and independently by Hotelling [3] to describe the multivariate data set variations in terms of uncorrelated variables each of which is a linear combination of the original variables. Its main goal is to derive new variables, in decreasing order of importance, that are linear combinations of the original variables and are uncorrelated with each other. It is a well-known technique, and it can be implemented by a number of connectionist models [11], [12].

### 2.2 Neural Implementation of Exploratory Projection Pursuit

The standard statistical method of EPP [1] also provides a linear projection of a data set, but it projects the data onto a set of basis vectors which best reveal the interesting

structure in data; interestingness is usually defined in terms of how far the distribution is from the Gaussian distribution.

One neural implementation of EPP is Maximum Likelihood Hebbian Learning (MLHL) [4], [13]. It identifies interestingness by maximising the probability of the residuals under specific probability density functions which are non-Gaussian.

One extended version of this model is the Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [14] model. CMLHL is based on MLHL [4], [13] adding lateral connections [14], [15] which have been derived from the Rectified Gaussian Distribution [16]. The resultant net can find the independent factors of a data set but does so in a way that captures some type of global ordering in the data set.

Considering an N-dimensional input vector ( $x$ ), and an M-dimensional output vector ( $y$ ), with $W_{ij}$ being the weight (linking input $j$ to output $i$ ), then CMLHL can be expressed [14], [15] as:

1. Feed-forward step:

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

2. Lateral activation passing:

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

3. Feedback step:

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \forall j \ . \tag{3}$$

4. Weight change:

$$\Delta W_{ij} = \eta . y_i . sign(e_j) | e_j |^{p-1} \ . \tag{4}$$

Where: $\eta$ is the learning rate, $\tau$ is the "strength" of the lateral connections, $b$ the bias parameter, $p$ a parameter related to the energy function [4], [13], [14] and $A$ a symmetric matrix used to modify the response to the data [14]. The effect of this matrix is based on the relation between the distances separating the output neurons.

### 2.3 Nonlinear Principal Component Analysis

Nonlinear Principal Component Analysis (NLPCA) [17] was designed to circumvent the limitations of linearity inherent in the PCA model. NLPCA is based on an auto-associative neural network and employs a Multi-Layer Perceptron (MLP) structure, which belongs to the feedforward class of neural networks [18].

The conventional MLP model implements a stimulus-response behaviour by combining several layers of elementary units ('neurons'). Each unit involves a simple, nonlinear transformation of weighted inputs; theoretical proof is available that

feedforward networks embedding a sigmoidal nonlinearity support arbitrary mappings [19], [20]. A conventional MLP includes three layers (input, 'hidden' and output), and associates an input vector, $\boldsymbol{x} \in \Re^D$, with an output vector, $\boldsymbol{y} \in \Re^Q$, computed as:

$$y_q(\mathbf{x}) = w'_{q,0} + \sum_{u=1}^{N_h} \left[ w'_{u,q} \cdot \sigma\left( w_{u,0} + \sum_{k=1}^{D} w_{u,k} x_k \right) \right]; q = 1,...,Q \ . \tag{5}$$

where, $N_h$ is the depth of the sigmoid series expansion, and W represents the coefficients of the weights for the interconnections between the two upper layers. An empirical fitting process tunes the weights, W, so that the network best reproduces the desired ($\boldsymbol{x}$, $\boldsymbol{y}$) mapping over a given training set. The classical cost function measuring the mapping distortion is the mean square error, $E_{\mathrm{W}}$, between the desired response (or 'target'), for a given input vector and the actual network output. Thus, the network-training process is formulated as an optimization problem expressed in the following terms:

$$\min_{\mathrm{W}} E_{\mathrm{W}} = \min_{\mathrm{W}} \frac{1}{n} \sum_{s=1}^{n} \left\| \mathbf{t}^{(s)} - \mathbf{y}\left(\mathbf{x}^{(s)}\right) \right\|^2 \ . \tag{6}$$

where $\boldsymbol{t}^{(s)}$ is the desired output for the s-th training vector, $\boldsymbol{x}^{(s)}$, and $n$ is the number of training pairs ($\boldsymbol{x}^{(s)}$, $\boldsymbol{t}^{(s)}$). In practice, the learning problem (6) is tackled efficiently and effectively by the Back-Propagation (BP) algorithm [18], which uses a stochastic gradient-descent strategy over the weight space.

Auto-Associative BP (AABP) networks constitute an unsupervised variant of the general MLP model, in which the desired outputs coincide with the network inputs: t ≡ x. The aim is a reduction in dimensionality by forcing the network to replicate the training sample distribution in this way, as the hidden layer is typically smaller than the input/output ones. At run-time, an AABP network associates each input vector with the 'coding' values computed by the hidden neurons; these mapping outputs support the (lossy) transformation from the input space into a lower-dimensional representation. A three-layer AABP network implements a mapping that is, in fact, affine to PCA. As such, the resultant mapping can suffer from the same drawbacks that affect PCA-like representations, such as a remarkable sensitivity to outliers in the training set. Hence, the NLPCA architecture (Fig. 1) that involved a sophisticated AABP model was proposed to tackle this issue.

As with conventional three-layer AABP, the output layer imposes the input values as targets and a hidden layer continues to support dimensionality reduction. The crucial difference with regard to the conventional AABP lies in the compression and reconstruction sections, each of which include an additional layer of neurons, whence the five-layer network. The NLPCA architecture retains the universal approximation ability of BP networks [17], and still adheres to the principle of unsupervised training. The run-time use of the resulting network, after completion of training, is equivalent to the use of a three-layer AABP structure, as the mapping outputs of the middle 'coding' layer provide the low-dimensional representation of each input vector.

Moreover, this increased power of representation conveyed by the NLPCA augmentation is remarkable. The problem is, of course, that nobody knows the $N_h$ target values in advance that should be imposed on the lower section of the network

for learning the compression task. In the NLPCA approach, those target values are implicitly imposed by forcing the network to reconstruct the original sample in the upper section. Thus, the 'reconstruction' section and the compression section will always be symmetrical and will therefore always yield equivalent, universal (inverse) mapping capabilities.

The main advantage is that the compressed representation does not relate to any linear model (as in PCA), but stems instead from a mainly general, internal representation that is empirically learned. NLPCA techniques seem to fit those domains in which 1) a nonlinear representation best encompasses the observed empirical phenomenon, and at the same time, 2) a considerable number of empirical samples are available.

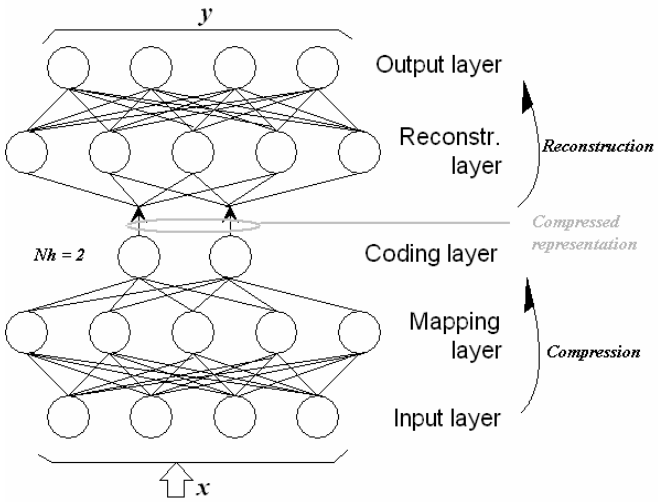


**Fig. 1.** A Nonlinear AABP network includes five layers to reduce data dimensionality

### 2.4 A General Architecture for an Intrusion Detection System

The projection models described above could work embedded in an IDS, performing a data analysis step, in an architecture akin to the MObile VIsualization Connectionist Agent-Based IDS (MOVICAB-IDS) [21], [22]. It has also been shown that this general architecture can be improved by the inclusion of the multiagent paradigm.

## 3 Real Data Set

In this work, the above mentioned neural models have been applied to a real traffic data set [22] containing normal traffic and anomalous situations. These anomalous situations are related to Simple Network Management Protocol (SNMP), known for its vulnerabilities [23]. The data set includes: SNMP ports sweeps (scanning of network computers for different ports - a random port number: 3750, and SNMP

default port numbers: 161 and 162 - using sniffing methods), and a Management Information Base (MIB) - the SNMP database - information transfer.

The used data set contains only five variables extracted from the packet headers: timestamp (the time when the packet was sent), protocol, source port (the port 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). This 'made-to-measure' data set was generated by the research team in a medium-sized network such that the "normal" and anomalous traffic flows were known in advance.

Emphasizing the importance of the time-dependence variable is one of the aims of this work. In line with this idea, 2 different data set variations (inclusion or exclusion of time information) were used.

## 4 Results and Conclusions

In this work, we have compared the performance of the three projection models (PCA, CMLHL and NLPCA) under review in order to analyse their response to the data set described above.

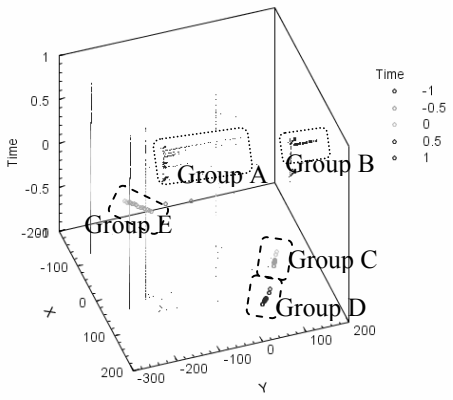


**Fig. 2.** NLPCA projection against time

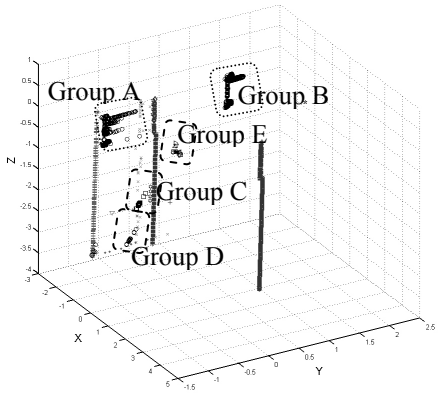


**Fig. 3.** CMLHL projection

In Fig. 2 and Fig. 3, we can see how the NLPCA and the CMLHL model are both able to identify the two anomalous situations contained in the data set. The MIB information transfer (Groups A and B in Fig. 2 and Fig. 3) is identified due to its orthogonal direction with respect to the normal traffic (vertical and parallel straight lines) and to the high density of packets. The sweeps (Groups C, D and E in Fig. 2 and Fig. 3) are identified due to their non-parallel direction to the normal one.

The results shown in Fig. 2 were obtained by using only four variables (excluding time information) for training the NLPCA network and plotting the 2-D projection (X, Y) against time. In contrast, the results on the CMLHL network (Fig. 3) were obtained by using the five variables to obtain a 3-D projection (X, Y and Z). This

outcome shows the intrinsic robustness of CMLHL, which is able to respond properly to a complex data set that includes time as a variable.

Finally, PCA was applied to the problem, firstly by including time information as a variable (Fig. 4), and then by excluding time information and plotting the two principal components against time (Fig. 5).

PCA was only able to identify the port sweeps (Group A in Fig. 4 and Fig. 5). As may be seen, it failed to detect the MIB information transfer because the packets in this anomalous situation evolve in a direction parallel to the "normal" one.

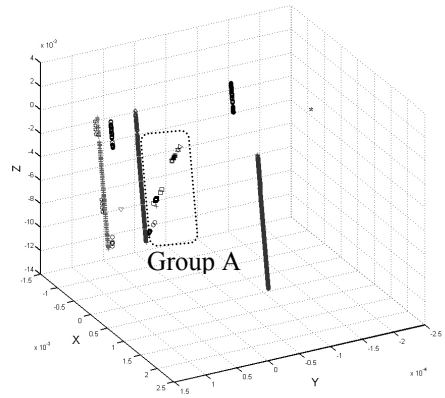


**Fig. 4.** The three first principal components

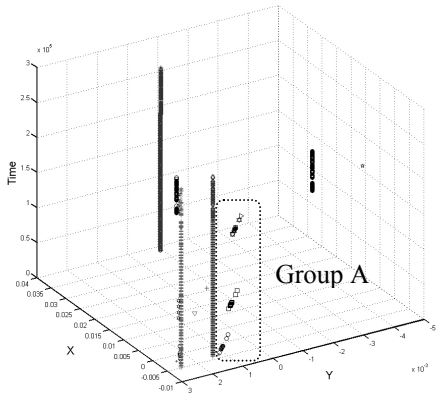


**Fig. 5.** The two first principal components against time

In conclusion, our work upholds the view that projection methods are an interesting and powerful tool in the identification of anomalous situations through visualization. A network administrator can easily identify a network scan represented by its evolution along a non-parallel direction to the normal one while an MIB transfer is characterized by its high packet density and its orthogonal direction with respect to the normal traffic. Another interesting issue is the capability of CMLHL to process time information as one of the data variables. In contrast, the NL-PCA network can not deal with the time variable, even though it obtains similar results to CMLHL by plotting its results against time, as shown in Fig. 2.

These two models outperform PCA, as the latter is unable to identify one of the anomalous situations: the MIB information transfer (including and excluding time infomation), while it always identifies the port sweeps.

Further work will focus on the application of high-performance computing clusters. Increased system power will be used to enable the IDS to process and display the traffic data in real time.

**Acknowledgments.** This research has been partially supported by the MCyT project TIN2004-07033 and the project BU008B05 of the JCyL.

## References


1. Friedman, J.H., Tukey, J.W.: A Projection Pursuit Algorithm for Exploratory Data-Analysis. IEEE Transactions on Computers 23(9), 881–890 (1974)
2. Pearson, K.: On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine 2(6), 559–572 (1901)
3. Hotelling, H.: Analysis of a Complex of Statistical Variables Into Principal Components. Journal of Education Psychology 24, 417–444 (1933)
4. Corchado, E., MacDonald, D., Fyfe, C.: Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. Data. Mining and Knowledge Discovery 8(3), 203–225 (2004)
5. Zanero, S.: Analyzing TCP Traffic Patterns Using Self Organizing Maps. In: Roli, F., Vitulano, S. (eds.) ICIAP 2005. LNCS, vol. 3617, pp. 83–90. Springer, Heidelberg (2005)
6. Sarasamma, S.T., Zhu, Q.M.A., Huff, J.: Hierarchical Kohonenen Net for Anomaly Detection in Network Security. IEEE Transactions on Systems Man. and Cybernetics 35(2), 302–312 (2005)
7. Carpinteiro, O.A.S., Netto, R.S., Lima, I., de Souza, A.C.Z., Moreira, E.M., Pinheiro, C.A.M.: A Neural Model in Intrusion Detection Systems. In: Kollias, S., Stafylopatis, A., Duch, W., Oja, E. (eds.) ICANN 2006. LNCS, vol. 4132, pp. 856–862. Springer, Heidelberg (2006)
8. Zhang, C.L., Jiang, J., Kamel, M.: Intrusion Detection Using Hierarchical Neural Networks. Pattern Recognition Letters 26(6), 779–791 (2005)
9. Debar, H., Becker, M., Siboni, D.: A Neural Network Component for an Intrusion Detection System. In: Proc. of the 1992 IEEE Computer Society Symposium on Research in Security and Privacy, pp. 240–250 (1992)
10. Ryan, J., Lin, M.J., Miikkulainen, R.: Intrusion Detection with Neural Networks. Advances in Neural Information Processing Systems (NIPS'97), vol. 10, pp. 943–949. The MIT Press, Cambridge (1998)
11. Fyfe, C.: PCA Properties of Interneurons: from Neurobiology to Real World Computing. Proc. of the Int. Conf. on Artificial Neural Networks, ICANN 1993, vol. 93, pp. 183–188. Springer Verlag, Berlin Heidelberg (1993)
12. Oja, E.: A Simplified Neuron Model as a Principal Component Analyzer. Journal of Mathematical Biology 15(3), 267–273 (1982)
13. Fyfe, C., Corchado, E.: Maximum Likelihood Hebbian Rules. Proc. of the 10th European Symposium on Artificial Neural Networks (ESANN 2002), pp. 143-148 (2002)
14. Corchado, E., Fyfe, C.: Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. Int. Journal of Pattern Recognition and Artificial Intelligence 17(8), 1447–1466 (2003)
15. Corchado, E., Han, Y., Fyfe, C.: Structuring Global Responses of Local Filters Using Lateral Connections. Journal of Experimental & Theoretical Artificial Intelligence 15(4), 473–487 (2003)
16. Seung, H.S., Socci, N.D., Lee, D.: The Rectified Gaussian Distribution. Advances in Neural Information Processing Systems 10, 350–356 (1998)
17. Kramer, M.A.: Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. Aiche Journal 37(2), 233–243 (1991)
18. Rumelhart, D.E., McClelland, J.L.: Parallel Distributed Processing. MIT Press, Cambridge, MA (1986)
19. Hornik, K., Stinchcombe, M., White, H.: Multilayer Feedforward Networks Are Universal Approximators. Neural Networks 2(5), 359–366 (1989)

20. Cybenko, G.: Approximations by Superpositions of Sigmoidal Functions. Mathematics of Control, Signal and Systems 2(4), 303–314 (1989)
21. Herrero, A., Corchado, E., Sáiz, J.M.: MOVICAB-IDS: Visual Analysis of Network Traffic Data Streams for Intrusion Detection. In: Corchado, E., Yin, H., Botti, V., Fyfe, C. (eds.) IDEAL 2006. LNCS, vol. 4224, pp. 1424–1433. Springer, Heidelberg (2006)
22. Corchado, E., Herrero, A., Sáiz, J.M.: Detecting Compounded Anomalous SNMP Situations Using Cooperative Unsupervised Pattern Recognition. In: Duch, W., Kacprzyk, J., Oja, E., Zadrożny, S. (eds.) ICANN 2005. LNCS, vol. 3697(2), pp. 905–910. Springer, Heidelberg (2005)
23. Cisco Secure Consulting. Vulnerability Statistics Report (2000)