A Proposal: Distributed Agent-Based IDS to Detect Anomalous SNMP Situations Using Unsupervised Learning

Álvaro Herrero, Emilio Corchado, Leticia Curiel, José Manuel Sáiz

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

Abstract. The present work is included in a research line approaching the anomalous situations detection issue from a pattern recognition point of view. Previous works have shown the viability and effectiveness of an Intrusion Detection System (IDS) based on the neural method called Cooperative Maximum Likelihood Hebbian Learning (CMLHL), which has never been applied to the IDS and network security field before this research. This system has been shown to be able to solve the difficult problem of identifying anomalous traffic patterns related to Simple Network Management Protocol (SNMP). The present work introduces a new approach for this IDS, based on a distributed architecture, where different agents cooperate to detect anomalous SNMP situations in big-size networks.

1 Introduction

IDS are hardware or software systems that monitored the events occurring in a computer system or network, analyzing them to identify computer security problems. They have become a necessary additional tool to the security infrastructure as the number of network attacks has increased very fast during the last years.

Among all the techniques that can be used today to implement IDS (such as state-transition diagrams, expert systems, petri nets, signature verification, etc.), connectionist models have been identified as a very promising method of addressing the Intrusion Detection (ID) problem due to the ability to detect day-0 (previously unknown) attacks and to classify patterns (attack classification, alert validation). Up to now, there have been several attends to apply connectionist models (such as Self-Organising Maps [1, 2] or Elman Network [3]) to the network security field [4]. This paper presents an IDS based on a connectionist model which has never been applied to the ID problem before this research: CMLHL [5, 6, 7]. It has been shown a very effective technique in the identification of anomalous SNMP situations [8, 9, 10, 11].

The actual demands of effectiveness and complexity have caused the development of new computing paradigms. One of these new paradigms is the agents and multiagent systems one. A software agent can be defined as a system with capacity of adaptation and provided with mechanisms allowing it to decide what to do (according to their objectives) [12]. This kind of systems has been previously used in the field of IDS [13, 14].

2 Problem Description

A protocol in a computer network context is a specification that describes low-level details of host-to-host interfaces or high-level exchanges between application programs. Among all the implemented network protocols, there are some of them that can be considered quite dangerous for network security. Among those, we have focused our effort in the study of SNMP because an attack based on this protocol may severely compromise system security [15]. SNMP was one of the top five most vulnerable services in order of importance identified by CISCO [16].

In the short-term, SNMP was oriented to manage nodes in the Internet community [17]. That is, it is used to control routers, bridges, and other network elements, reading and writing a wide variety of information about the devices: operating system, version, routing tables, default TTL (Time To Live), and so on. Some of this data can be extremely sensitive. The IAB (Internet Activities Board) recommended that all IP (Internet Protocol) and TCP (Transmission Control Protocol) implementations were network manageable [18]. The implementation of the Internet Management Information Base (MIB) and at least one of the management protocols like SNMP is the consequence of this suggestion. The MIB can be roughly defined as a database that contains information about some elements or devices that can be network-controlled. It stores the information about the elements that SNMP controls

There are some dangerous anomalous situations related to SNMP [8, 9, 10, 11], as an SNMP port sweep (a scanning of network computers using sniffing methods to verify if SNMP protocol is active in any ports) and an MIB information transfer (a transfer of some information contained in the SNMP MIB). The latter 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.

SNMP can also be used in a segmented network. In this case, the terms SNMP Agent, Proxy Agent and SNMP party are introduced [19]. An SNMP Agent is the operational role assumed by an SNMP party (generally a device controlled by this protocol) when it performs SNMP management operations in response to received SNMP protocol messages. An SNMP Proxy Agent is an SNMP Agent that performs management operations by communicating with another logically remote party. In the case of segmented networks, “logically remote” means that each party can be located in a different network segment.

The transparency principle [19] defines the behavior of an SNMP party and says that the way in which one SNMP party processes SNMP protocol messages received from another SNMP party is entirely transparent to the latter. Implicit in this principle is the requirement that, throughout its interaction with a Proxy Agent, a management station is supplied with no information about the nature or progress of the proxy mechanisms by which its requests are performed. That is, it should seem to the management station as if it were interacting via SNMP directly with the proxied device.

3 Overview of the Distributed Agent-Based IDS

To upgrade the model previously developed [8, 9, 10, 11], the IDS is split out in different agents that work together in order to detect intrusive actions.

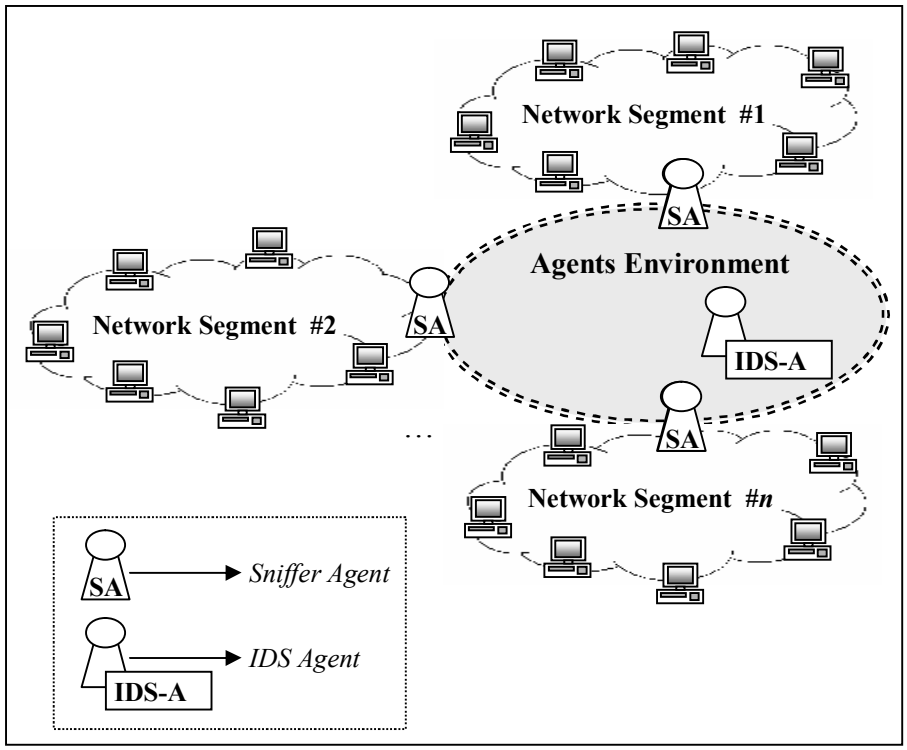

Fig. 1. The Distributed Agent-Based IDS

Corporative networks can be very big-size ones, where hosts are set up into different network segments. It is mainly caused by IP addresses limitations. All the different SNMP anomalous situations can be produced in every different segment (where SNMP Agents are set up). Here is where a distributed IDS can take advantage. By using a distributed IDS (where "listener" entities capture the traffic traveling along each different network segment), the model is able to identify all the anomalies caused in a segment-divided network. Otherwise, only the anomalies caused in the segment where the IDS is located could be identified. In order to detect these situations we propose this distributed agent-based IDS. Its structure is shown in Fig. 1 and it consists on two different kinds of agents:

- **Sniffer Agent (SA)**: one agent of this kind is in charge of one segment in which the network is divided.
- **IDS Agent (IA)**: there is only one IDS agent, which is in charge of processing the information sent by Sniffer Agents and alerting the network administrator.

This structure allows the system using source and destination IP addresses in such a way that each agent could know from where network segment the packet is coming.

3.1 Sniffer Agent

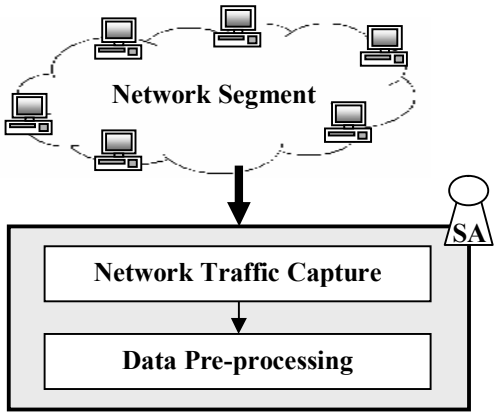

Fig. 2. Sniffer Agent Structure

The structure of Sniffer Agents is shown in Fig. 2 and they carry out two main functions:

- **Network Traffic Capture**: one of the network interfaces (of the host where the agent is located) is set up as "promiscuous" mode. It captures all the packets traveling along this network segment.
- **Data Pre-processing**: the captured data is selected, pre-processed and sent to the IDS Agent.

The necessary information for the traffic analysis is obtained from the headers of the packets that travel along the network. This data can be obtained by using a network analyser. The study of SNMP is the reason why the system selects packets based on UDP (User Datagram Protocol). This means that in terms of TCP/IP (Transmission Control Protocol/Internet Protocol) protocol stack, the model captures only the packets using UDP at transport layer and IP at network layer. In addition to SNMP packets, the data sets contain traffic related to other protocols installed in our network, like NETBIOS and BOOTPS.

In the Data Pre-processing step, the agent performs a data selection of all the information captured. After that, the data sets contain the following variables extracted from the packet headers:

- **Timestamp**: time when the packet was sent.
- **Protocol**: all the protocols contained in the data set should be codified.
- **Source Port**: port number of the source host that sent the packet.

- **Destination Port**: port number of the destination host where the packet is sent.
- **Source IP address**: numeric value that codifies the source host IP address.
- **Destination IP address**: numeric value that codifies the destination host IP address.
- **Size**: total packet size.

Only the pre-processed data is sent to the IDS Agent. This does not imply a huge increase of network traffic because only a reduced portion of the information is sent.

3.2 IDS Agent

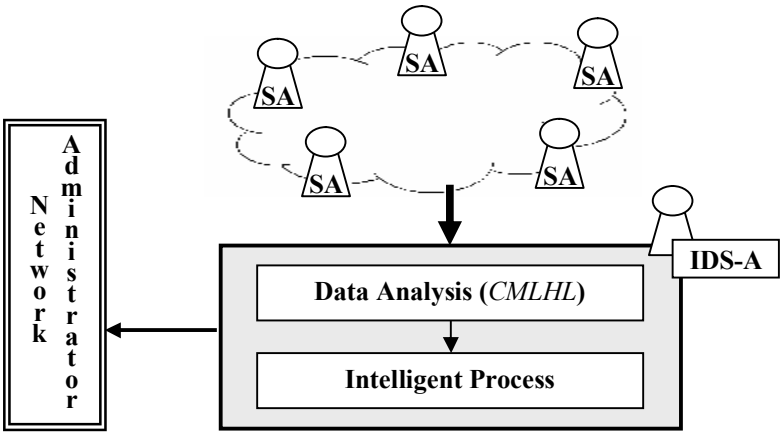

Fig. 3. IDS Agent Structure

Once the pre-processed data is received, the unsupervised connectionist model (see section 4) analyses the data and identifies anomalous patterns. This distributed structure allows the IDS Agent to identify anomalous situations concerning different network segments. That is, the identification of anomalous situations that can affect SNMP Agents located in different segments. Finally, the network administrator is alerted about the anomalous situations that are happening (or had happened) along all the segments in which the network is divided.

To handle and analyse all the pre-processed information concerning the whole network, only the IDS Agent should have a very big calculus power. This Agent can be located in the most powerful machine in the organization. It does not matter what network segment this machine is located in. Additionally, the IDS agent can be equipped with different mechanisms to abort an attack. That is, after identifying an anomalous situation, it can carry out concrete actions to abort the attack like turning SNMP off.

4 The Cooperative EPP IDS Model

Exploratory Projection Pursuit (EPP) [20, 21, 22, 23] 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 where 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 Analysis step performed by the IDS Agent (Fig. 3) is based on the use of a neural EPP model called Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [5, 6, 7]. It was initially applied to the field of artificial vision [5, 6] to identify local filters in space and time. Here, we have applied it to the field of computer security. It is based on Maximum Likelihood Hebbian Learning (MLHL) [22, 23] but adding lateral connections [5, 6], which have been derived from the Rectified Gaussian Distribution [24]. 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. 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. CMLHL can be expressed as:

Feed forward:

$$y_i = \sum_{j=1}^{N} W_{ij} x_j, \forall i . \quad (1)$$

Application of lateral connections:

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

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

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

Weight change:

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

Where: τ is the "strength" of lateral connections, b is the bias parameter, η is the learning rate and p is a parameter related to the energy function [6, 22, 23].

A is a symmetric matrix used to modify the response to the data whose effect is based on the relation between the distances among the output neurons. It is based on the Cooperative Distribution [24], but to speed learning up, it can be simplified to [25]:

$$A(i, j) = \delta_{ij} - \cos\left(2\pi\left(i - j\right)/ M\right). \quad (5)$$

δ_{ij} is the Kronecker delta.

5 Results, Conclusions and Future Work

The presented IDS is capable of identifying anomalous situations by means of temporal visualization of the system response. These situations are related for instance to those which are non parallel (for example, a port sweep [8, 9, 11]) or with higher temporal concentration of packets (for example, a MIB information transfer [9, 10, 11]) than the normal ones.

The application of this structure provides the following advantages:

- SNMP anomalous situations concerning different network segment (where SNMP Agents are set up) can be detected.
- Only a very powerful machine is needed to contain the IDS Agent. To process and classify the information in the agent located in each segment, all the agents should have a high calculus power.
- It does not matter in which segment the IDS Agent is located. It can be reallocated in a different segment or host in the case of the host or segment where it was located falls down, what implies a high level of robustness.
- Automatic mechanisms can be run to abort an attack.

Future work will be focused on the upgrading of this agent-based IDS and the application of parallel computing to speed up the data analysis process.

Acknowledgments

This research has been supported by the McyT project TIN2004-07033 and the project BU008B05 of the JCyL.

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. 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. Zanero S., Savaresi S.M.: Unsupervised Learning Techniques for an Intrusion Detection System. ACM Symposium on Applied Computing (2004) 412 – 419
4. Ghosh, A. Schwartzbard A., Schatz A.: Learning Program Behavior Profiles for Intrusion Detection. Workshop on Intrusion Detection and Network Monitoring (1999)
5. 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 (ICANNGA 2005). Springer Computer Science (2005) 454 – 457
6. Herrero, A., Corchado, E., Sáiz, J.M.: Identification of Anomalous SNMP Situations Using a Cooperative Connectionist Exploratory Projection Pursuit Model. International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2005). Lecture Notes in Computer Science, Vol. 3578. Springer-Verlag, Berlin Heidelberg New York (2005) 187 – 194
7. Herrero, A., Corchado, E., Sáiz, J.M.: A Cooperative Unsupervised Connectionist Model Applied to Identify Anomalous Massive SNMP Data Sending. International Conference on Natural Computation (ICNC 2005). Lecture Notes in Computer Science, Vol. 3610. Springer-Verlag, Berlin Heidelberg New York (2005) 778 – 782
8. Corchado, E., Herrero, A., Baruque, B., Sáiz J.M.: Detecting Compounded Anomalous SNMP Situations Using Unsupervised Pattern Recognition. International Conference on Artificial Neural Networks (ICANN 2005). Lecture Notes in Computer Science, Vol. 3697. Springer-Verlag, Berlin Heidelberg New York (2005) 905 – 910
9. Wooldridge, M. Multiagent Systems: A Modern Approach to Distributed Artificial Inteligence. Gerhard Weiss (1999)
10. Spafford, E.H., Zamboni, D.: Intrusion Detection Using Autonomous Agents. Computer Networks: The International Journal of Computer and Telecommunications Networking, 34(4) (2000) 547 – 570
11. Hegazy, I.M., Al-Arif, T., Fayed, Z.T., Faheem, H.M.: A Framework for Multiagent-based System for Intrusion Detection. Intelligent Systems Design and Applications. Advances in Soft Computing Serie. Springer-Verlag, Berlin Heidelberg New York (2003)
12. Myerson, J.M.: Identifying Enterprise Network Vulnerabilities. International Journal of Network Management 12 (2002)
13. Cisco Secure Consulting: Vulnerability Statistics Report (2000)
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. Davin, J., Galvin, J., McCloghrie, K.: SNMP Administrative Model. RFC-1351 (1992)
17. Friedman J., Tukey. J.: A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers 23 (1974) 881-890
18. Hyvärinen A.: Complexity Pursuit: Separating Interesting Components from Time Series. Neural Computation 13 (2001) 883-898
19. 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
20. Fyfe, C., Corchado, E.: Maximum Likelihood Hebbian Rules. European Symposium on Artificial Neural Networks (2002)
21. 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

22. 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
23. Corchado, E., Corchado, J.M., Sáiz, L., Lara, A.: Constructing a Global and Integral Model of Business Management Using a CBR System. First International Conference on Cooperative Design, Visualization and Engineering (2004)
24. Seung, H.S., Socci, N.D., Lee, D.: The Rectified Gaussian Distribution. Advances in Neural Information Processing Systems 10 (1998) 350
25. Charles, D.: Unsupervised Artificial Neural Networks for the Identification of Multiple Causes in Data. Ph.D. Thesis. University of Paisley (1999)