
International JOINT Conference 2006
iberamia
SBIA
SBRN
Organized by
USP
Sponsered by SBC
Co-sponsered by
AEPIA, APPIA, SMIA
ISBN 85-87837-11-7
Proceedings of the International Joint Conference,
10th Ibero-American Artificial Intelligence Conference,
18th Brazilian Artificial Intelligence Symposium,
9th Brazilian Neural Networks Symposium,
IBERAMIA-SBIA-SBRN, Ribeirão Preto, Brazil,
October 23-28, 2006
Edited by Solange Oliveira Rezende - General Chair
Antonio Carlos Roque da Silva Filho - General Co-Chair

Proceedings of the International Joint Conference IBERAMIA/SBIA/SBRN 2006 - 1st Hybrid Artificial Intelligence Systems Workshop (HAIS'2006), Ribeirão Preto, Brazil, October 23–28, 2006. CD-ROM. ISBN 85-87837-11-7

Mobile Hybrid Intrusion Detection System

Álvaro Herrero and Emilio Corchado

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

Abstract. This paper is part of a research line that approaches the anomalous situations detection issue combining projectionist methods and software agents. This Intrusion Detection System (IDS) responds to the challenges presented by traffic volume and diversity. It is a connectionist agent-based model extended by means of a functional and mobile visualization interface. Previous works have shown the viability and effectiveness of an IDS based on a neural method that has never been applied to the IDS and network security field before this multidisciplinary 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).

1 Introduction

Intrusion Detection Systems (IDSs) are tools designed to monitor and analyse computer system or network events in order to detect suspect patterns that may relate to a network or system attack. An IDS that analyses packets travelling over an entire network is referred to as a network-based IDS (NIDS).

Many different forms of Artificial Intelligence (such as Genetic Programming [1], Data Mining [2], [3] or Neural Networks [4], [5], [6] among others), and statistical [7] and signature verification [8] techniques have been applied in the field of IDSs. There are several IDSs that can generate different alarms when an anomalous situation occurs, but they can not provide a general overview of what is happening inside a network. Various visualization techniques have been applied in the field of IDSs [4], [5], [9], [10], [11], [12] to tackle this issue. Some of them (The Multi Router Traffic Grapher [12] for example) offer visual measurements of network traffic. The proposed model goes further and offers a complete and more intuitive visualization of network traffic by depicting each simple packet and providing the network administrator with a snapshot of network traffic, protocol interactions, and traffic volume, generally in order to identify anomalous situations.

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) [13]. This kind of systems has been previously used in the field of IDS [14, 15].

2 Computer Security Framework

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 Simple Network Protocol (SNMP) because an attack based on this protocol may severely compromise system security [16]. SNMP was one of the top five most vulnerable services in order of importance identified by CISCO [17].

In the short-term, SNMP was oriented to manage nodes in the Internet community [18]. 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 [19]. 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 [4], [5], 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.

3 The Hybrid Intrusion Detection System

The presented model may be defined as an IDS formed of different software agents [13] that work in unison in order to detect anomalous situations by taking full advantage of an unsupervised connectionist model [4], [5], [20], [21], [22]. It is designed to split massive traffic data sets into segments and analyse them, thereby providing administrators with a visual tool to analyse the kinds of events taking place on the computer network. This tool also provides an analysis of several subsequent segments as unique ones (simple segments) and also as an accumulated data set.

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. 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.

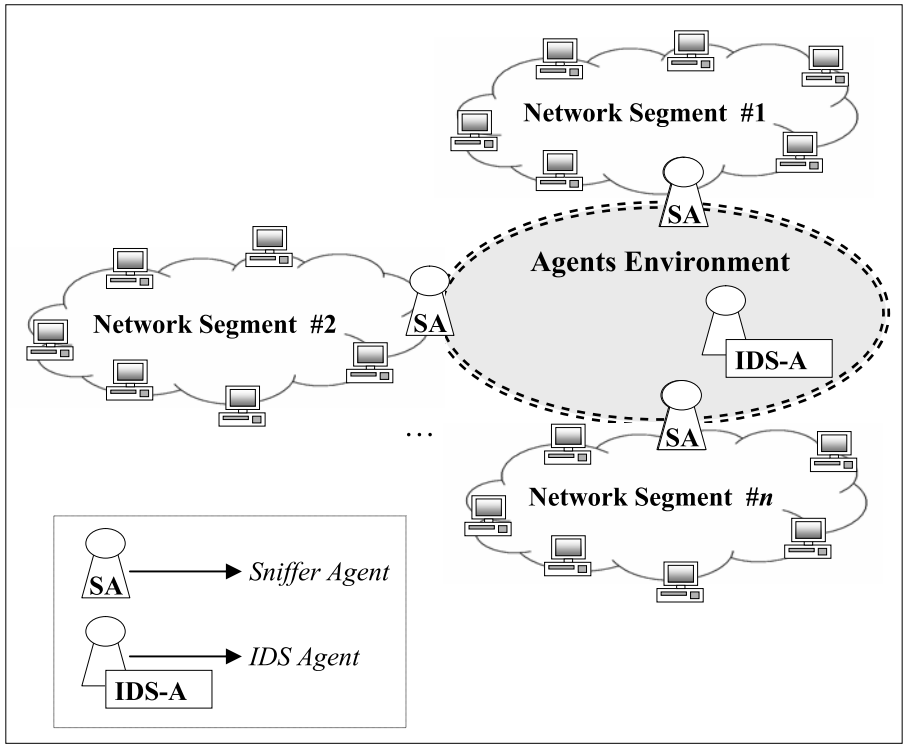

Fig. 1. The Distributed Agent-Based IDS

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

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 travelling along the network. This data can be obtained by using a network analyser. In addition to SNMP packets, the data sets contain traffic related to other protocols installed in our network, like NETBIOS and BOOTPS.

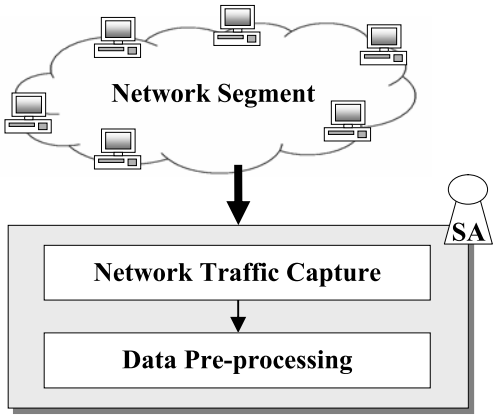

Fig. 2. Sniffer Agent Structure

In the Data Pre-processing step, the agent performs a data selection of all the captured information. After that, the data sets contain 5 variables extracted from the packet headers. 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

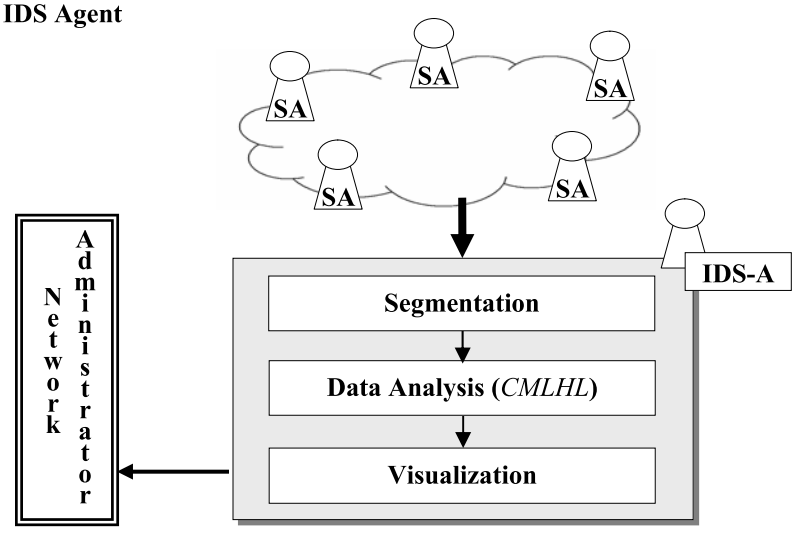

Fig. 3. IDS Agent Structure

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 Pojectionist Analysis Model

The Data Analysis step performed by the IDS Agent (Fig. 3) is based on the use of a neural Exploratory Projection Pursuit (EPP) [23], [24] model called Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [20], [21], [22]. It was initially applied in the field of Artificial Vision [20], [21] to identify local filters in space and time. In this work, it is applied in the field of Computer Network Security. CMLHL is based on Maximum Likelihood Hebbian Learning (MLHL) [25], [26] adding lateral connections [20], [21] which have been derived from the Rectified Gaussian Distribution [27]. 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. The architecture of the network is as follows:

Feed forward:

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

Application of lateral connections:

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

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

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

Weight change:

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

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 [4], [5], [21], [22].

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 [27], but to speed learning up, it can be simplified to:

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

δ_{ij} is the Kronecker delta.

5 Traffic Data Stream

NIDSs have to deal with the practical problem of high volumes of quite diverse data [22]. To deal with the problem of high diversity, this model splits the traffic into different groups, taking into account the protocol (either UDP, TCP, ICMP…) over IP. For the sake of simplicity, only UDP traffic is considered in this work due to its potential dangers.

Once the data set is classified by the protocol over IP, our model is based on the analysis of five main numerical variables (timestamp, source and destination port, packet size and protocol) existing on the packets headers. The capability of these variables to identify different anomalous situations has already been demonstrated.

Then, the model divides the data sets follows:

- Equal simple segments. Each simple segment contains all the packets whose timestamps are between its initial and final limits. As can be seen in Fig. 4, there is a time overlap between each consecutive simple segments. This is done because anomalous situations could conceivably take place between simple segment S_x and S_{x+1} (the next segment following S_x). In this case, it would be necessary to consider some packets twice in order to visualize the end of the anomalous situation and the evolution between simple segments. It may be interesting to analyse the beginning of anomalous situations taking place at the end of a simple segment in comparison with the following packets. Both the length (time duration) of the simple segments and the overlap time can be set up by the administrator.
- Accumulated segments. Each one of these segments contains several consecutive simple ones (Fig. 4). The main considerations are, firstly, to present a long-term picture of the evolution of network traffic to the network administrator and, secondly, to allow the visualization of attacks lasting longer than the length of a simple segment. The number of simple segments making up the accumulated ones is configurable.

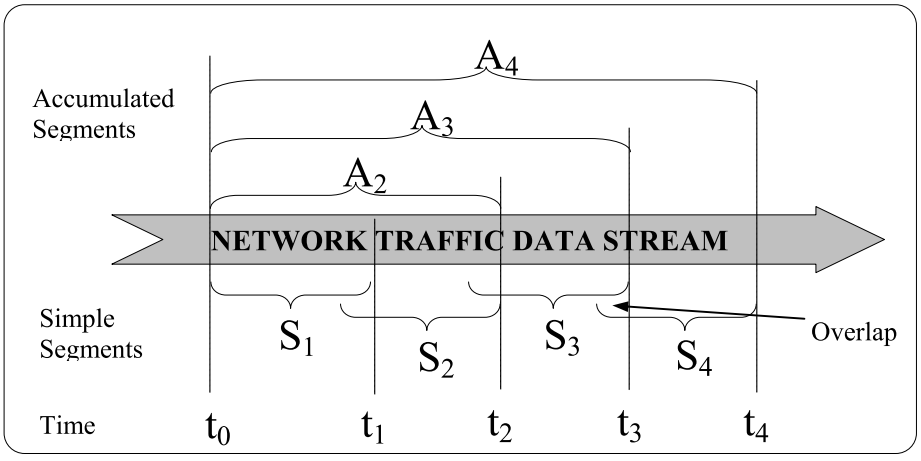

Fig. 4. Data stream fragmentation. Each data set is divided into several simple segments (e.g. S_1, S_2 and so on) and accumulated ones (e.g. A_2, A_3, …)

6 Results and Conclusions

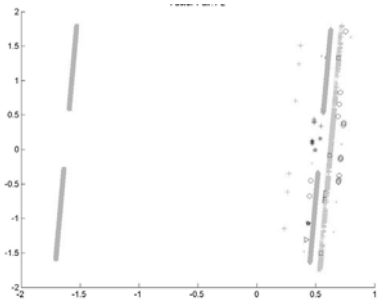

Fig. 5. Visualization of simple segment

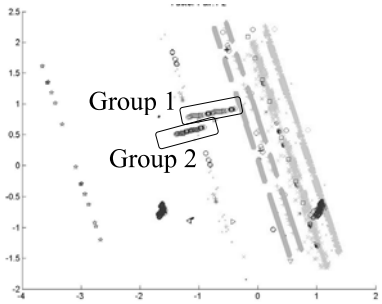

Fig. 6. Visualization of accumulated segment

Fig. 5 shows how the proposed model visualizes a simple segment containing only "normal" traffic. On the other hand, Fig. 6 shows the identification of an anomalous situation related to an SNMP port sweep (Groups 1 and 2).

The presented IDS is capable of identifying anomalous situations by means of temporal visualization of the system response. The 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.

The combination of a multiagent structure and a neural model 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.

This system can be used in combination with other IDS to overcome their limitations (e.g: identification of 0-day attacks).

Further work will focus on the study of different anomalous situations to extend the model to cover several protocols, and the application of different learning rules in the Analysis Step.

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

References

1. Abraham, A., Grosan, C., Martin-Vide, C.: Evolutionary Design of Intrusion Detection Programs. International Journal of Network Security (2006)
2. Julisch, K.: Data Mining for Intrusion Detection: A Critical Review. Research Report RZ 3398, IBM Zurich Research Laboratory. Switzerland (2002)
3. Lee, W., Stolfo, S.J.: A Framework for Constructing Features and Models for Intrusion Detection Systems. ACM Trans. on Inf. and System Sec. (TISSEC), Vol. 3(4). ACM Press, New York (2000)
4. Herrero, A., Corchado, E., Sáiz, J.M.: A Cooperative Unsupervised Connectionist Model Applied to Identify Anomalous Massive SNMP Data Sending. Proc. of the Int. Conf. on Natural Computation (ICNC). LNCS, Vol. 3610. Springer-Verlag, Berlin Heidelberg New York (2005) 778-782
5. Corchado, E., Herrero, A., Sáiz J.M.: Detecting Compounded Anomalous SNMP Situations Using Unsupervised Pattern Recognition. Proc. of the Int. Conf. on Artificial Neural Networks (ICANN 2005). LNCS, Vol. 3697. Springer-Verlag, Berlin Heidelberg New York (2005) 905-910
6. Zanero, S., Savaresi, S.M.: Unsupervised Learning Techniques for an Intrusion Detection System. Proceedings of the ACM Symposium on Applied Computing (2004) 412-419
7. Marchette, D.J.: Computer Intrusion Detection and Network Monitoring: A Statistical Viewpoint. Information Science and Statistics. Springer-Verlag, Berlin Heidelberg New York (2001)
8. Roesch, M.: Snort - Lightweight Intrusion Detection for Networks. Proceedings of the 13th Systems Administration Conference (LISA '99) (1999)
9. Goldring, T.: Scatter (and Other) Plots for Visualizing User Profiling Data and Network Traffic. Proceedings of the ACM Workshop on Visualization and Data Mining for Computer Security (2004)
10. Muelder, Ch., Ma, K-L., Bartoletti: Interactive Visualization for Network and Port Scan Detection. Proceedings of the 8th International Symposium on Recent Advances in Intrusion Detection (RAID). Lecture Notes in Computer Science, Vol. 3858. Springer-Verlag, Berlin Heidelberg New York (2005)
11. Abdullah, K., Lee, Ch., Conti, G., Copeland, J.A.: Visualizing Network Data for Intrusion Detection. Proceedings of the IEEE Workshop on Information Assurance and Security (2002) 100-108
12. MRTG: The Multi Router Traffic Grapher, http://people.ee.ethz.ch/~oetiker/webtools/mrtg/
13. Wooldridge, M. Multiagent Systems: A Modern Approach to Distributed Artificial Inteligence. Gerhard Weiss (1999)
14. 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
15. 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)
16. Myerson, J.M.: Identifying Enterprise Network Vulnerabilities. International Journal of Network Management 12 (2002)
17. Cisco Secure Consulting: Vulnerability Statistics Report (2000)
18. Case, J., Fedor, M.S., Schoffstall, M.L., Davin, C.: Simple Network Management (SNMP). RFC-1157 (1990)
19. Postel, J.: IAB Official Protocol Standards. RFC-1100 (1989)
20. Corchado, E., Han, Y., Fyfe, C.: Structuring Global Responses of Local Filters Using Lateral Connections. Journal of Experimental and Theoretical Artificial Intelligence, Vol. 15(4) (2003) 473-487
21. Corchado, E., Fyfe, C.: Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. Int. Journal of Pattern Recognition and Artificial Intelligence, Vol. 17(8) (2003)
22. Corchado, E., Corchado, J.M., Sáiz, L., Lara, A.: Constructing a Global and Integral Model of Business Management Using a CBR System. Proc. of the 1st Int. Conf. on Cooperative Design, Visualization and Engineering (CDVE). LNCS, Vol. 3190. Springer-Verlag, Berlin Heidelberg New York (2004) 141-147
23. Friedman J., Tukey. J.: A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers, Vol. 23 (1974) 881-890
24. Hyvärinen A.: Complexity Pursuit: Separating Interesting Components from Time Series. Neural Computation, Vol. 13(4) (2001) 883-898
25. Corchado, E., MacDonald, D., Fyfe, C.: Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. Data Mining and Knowledge Discovery, Vol. 8(3), Kluwer Academic Publishing (2004) 203-225
26. Fyfe, C., Corchado, E.: Maximum Likelihood Hebbian Rules. Proceedings of the European Symposium on Artificial Neural Networks (2002) 143-148
27. Seung, H.S., Socci, N.D., Lee, D.: The Rectified Gaussian Distribution. Advances in Neural Information Processing Systems, Vol. 10 (1998) 350-356