ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

Intrusion detection pattern recognition using an Artificial Neural Network

José Ernesto Luna Domínguez[1] , Anabelem Soberanes Martín[1]
[1] Universidad Autónoma del Estado de México,
Centro Universitario UAEM Valle de Chalco

***Abstract*— Computer security has always been essential to control access and detect intruders in information systems, data networks and database. It has sought the privacy and integrity of the information is as secure as possible, using intelligent techniques for the analysis and application of data mining, in order to obtain the behavior patterns of the intruder. Technology advances or evolves exponentially, which causes the development of sophisticated and efficient systems of information, what new methods are derived unauthorized accesses and vulnerabilities.**

In this article the results of intrusion detection with automatic technique that allows the neural network is developed, this technique effectiveness tube by using certain training techniques as well as the weight matrix. Thanks to the network user patterns were obtained and a solution to detect unauthorized access to computer systems is proposed.

***Index Terms*— Pattern recognition, artificial neural network, intrusion, intrusion detection.**

I. INTRODUCTION

There are several computer systems in the world, and these constitute a large amount of information in network known as the Internet. Thousands of people (users) access to say data network without realizing that leave or generate data behavior, i.e. when they use the systems to share information on its geographical position, activities, visited sites and even personal preferences. This is the pattern of user activity and data can be used as a unique identifier for each person, this identifier facilitates the task of Intrusion Detection Systems (IDS) effectively.

The generated information about the activity patterns is stored in multiple databases systematically, making analysis more difficult. Due to the above, we have sought to develop tools (software) to solve the difficulty of the analysis. The tools can also generate patterns of user behavior, which in turn makes it possible to generate a personal profile to all users who use the system.

Taking into consideration that every user on the network has a regular behavior (aka standard), this behavior must be made allowing dynamic fit to the daily needs of each person, which is why we have made robust techniques, allowing efficient and agile to fit the constantly changing personal behavior. For the development of this study data analysis techniques were analyzed using Artificial Neural Networks (ANN), as they are used for pattern recognition, clustering and classification. The ANN constitute a learning method in order to mimic the way that the biological neurons process information; ANN techniques were applied in order to generate a dynamic profile for the user, that behaves as a unique identifier for access to information systems, and improve the intrusion detection systems.

II. ARTIFICIAL NEURAL NETWORK (ANN)

They are a learning paradigm and automatic processing, which were inspired by the way of functioning of the nervous system of some animals. They consist of an interconnection of neurons working together to generate the stimulus output. Within the area of Artificial Intelligence (AI) are known as neural networks, neurons are cells with the main feature of the electrical excitability (pulses) generated by the plasma membrane [1]. As shown in Fig. 1, neurons consist of:

- Axon: Transmits messages that are sent to other neurons
- Dendrites: Are the outer fibers of a neuron, such fibers are those that receive messages from other neurons.
- Myelin: Fundamental, which prevents neurons to enter into short circuit surrounds the axon. On the other hand contributes to the message transmission speed.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

• Terminals: Bumps on the end of the axon that sends messages to other neurons.

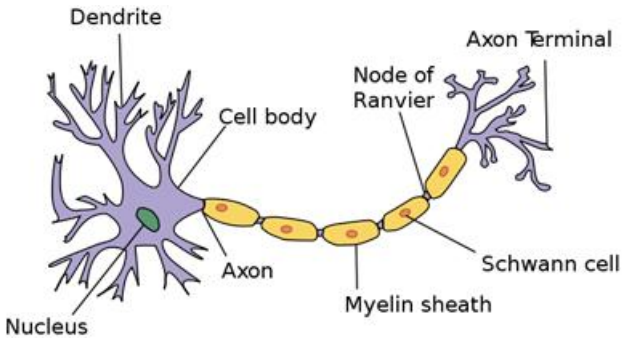

Fig. 1 Structure of a Biological Neuron (www.usdbiology.com)

Informatics model or computational

The artificial neuron is also known as PLC, which has an activation level, which receives signals that change the state of the neuron, these signals are known as activation function transition. The pulses received in the neuron from abroad or from other neurons of the same ma network [2].

The level of activation depends on the entries received and the synaptic values. You need to know the value of the activation state, and this is achieved by calculating the total input of the automaton, i.e. buckles inputs to certain values, in Fig. 2 the typical artificial neuron shown added.

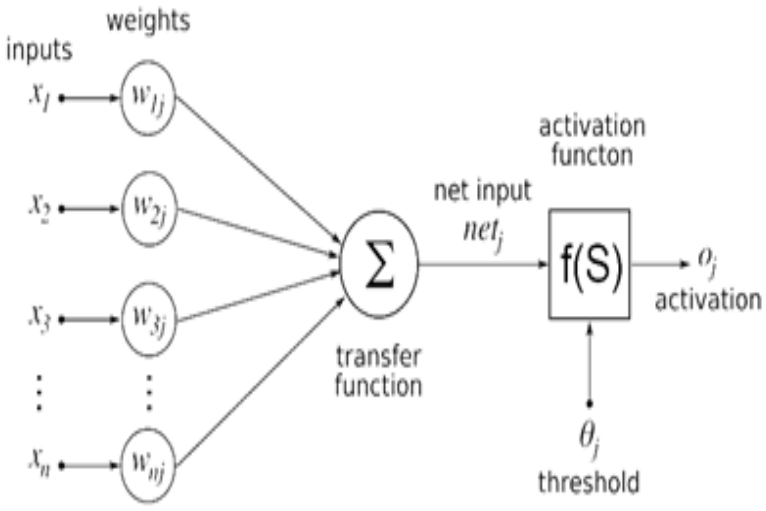

Fig. 2 Structure of an artificial neuron (www.dspguide.com)

In the Fig. 3 shows the structure of an artificial neural network is shown in her input units which form the input layer are observed in this layer shows that the output of each neuron is the input for the other network this form of interconnection between neurons called connectivity pattern. Each entry has a vector, when being introduced by the network is copied each value of the vector in the corresponding neuron, how do each input unit receives the full vector information, processes it and generates a single output that is disseminated by connections between neurons. When the input has spread to the entire network, an output vector whose components are generated each of the input values of the output neurons [3].

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

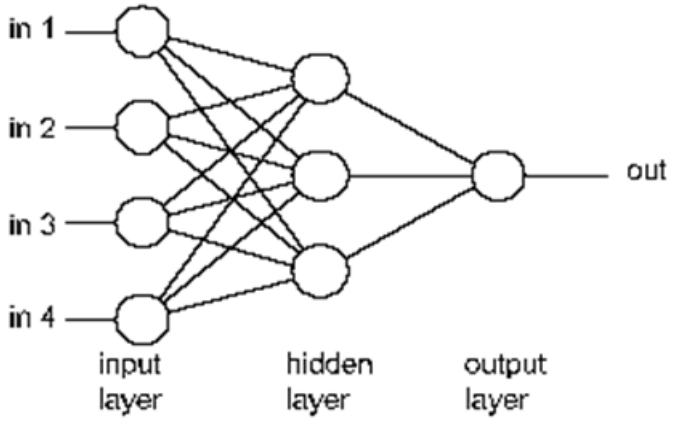

Fig. 3 Structure of a basic artificial neural network (www.cheshireeng.com)

For this proper operation is necessary for the neural network to generate learning, which depends on the scheme with the learner, as this will be the types of problems that solve achieved. There are several schemata for learning, these schemes are:

- Supervised learning
- Unsupervised learning
- Learning by strengthening

Within research supervised learning better known as outputs are used, it is a technique of deduction from the data with which the network is trained, these data are vectors that are composed of a pair of input and other output desired. This learning scheme aims to create a function that is able to predict the value corresponding to the input value after having seen a number of examples, in Fig. 4 the process of supervised learning is outlined [4].

The way it looks for the purpose of this learning algorithm is following the steps to do listed below:

1. Initialize the weights randomly
2. Presentation of the training set (TS)
3. Obtaining outputs for TS
4. Comparison with current outputs desired.
5. If the termination criterion is checked to the next step, but go to the wall so 2.
6. Complete the process.

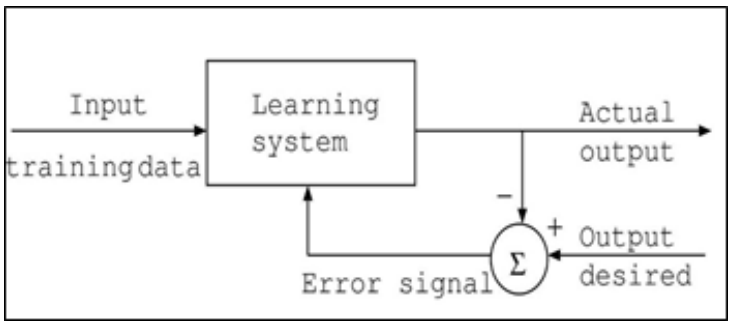

Fig. 4 Supervised learning algorithm (www.intechopen.com)

III. PATTERN RECOGNITION

The knowledge gained on the basis of data processed pa seeks a unique pattern of behavior for each user, this pattern should be dynamic and more close to reality, which is necessary to understand the models for data analysis.

Classification tasks were performed, of which a set of bosses based on the relationship between the data, it seeks not only to determine the class for each new instance to be evaluated, but also want to find will define the degree of

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

certainty of predictions, which will improve the classification of the data, it attempts to perform tasks detection of outliers, which suggest fraud, failure or intrusion.

For classification of ANN was used with the intention of detecting more efficiently intrusions into the systems that users use. Collected the information shown in Table 1, which is composed of four values: date, time, method and exercise, are records of system users, these data allow fully identify whether the user is or is not an intruder, this network was designed and developed in Matlab 2013 with the input values shown in Table 1, came home after several tests until they managed to get the expected result, in Figure 5 the diagram obtained by the software shown above [5] [6].

Table 1. Sample input data

Date	Time	Method	Exercise
2014/02/25	**12**	**3**	**1**
2014/02/26	**11**	**2**	**1**
2014/02/27	**13**	**2**	**1**
2014/02/27	**12**	**2**	**1**
2014/03/01	**14**	**3**	**1**

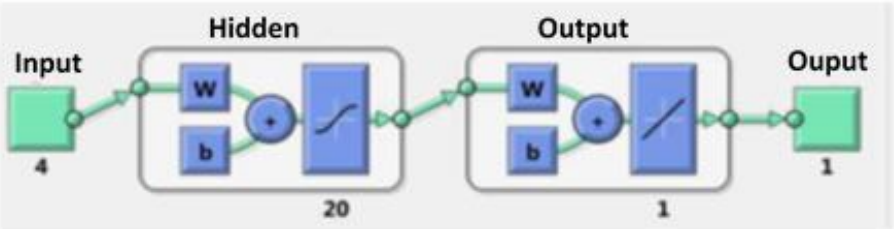

Fig. 5 Network generated in Matlab

These data are organized into two matrices, the input X and the target T, each column i of the input array has 4 elements representing the date, time, method and exercise, while each corresponding to the target column matrix only count on an element, this element represents the user causing the intrusion.

Since this network is initiated with random weights, the results of the implementation vary slightly each running, implementation was in two layers, and it is just a layer hides which contains 24 neuron, depending on the complexity of the problems determine the number of neurons, in this case once the neural network design, we proceeded to train her, which required a set of training data, validation and testing other.

Once the training was completed in Matlab, it is necessary to observe the performance window inside the tab run, to see if performance improved during training. The trained network is tested with test input vectors, these tests identify wait to get the results when performed with data in real systems [6].

Correct and incorrect classifications are shown in Table 2 with the test input vectors to see if the network has a right or proper training, the percentage of incorrect classifications must be very small, if it was contrary, that means keep training differently, or even need to have more neurons in the hidden layer.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

Table 2. Test rankings

Test data	% incorrect	% correct
315	1%	99%
260	3%	97%
390	1%	99%

Matlab has a tool to measure the false positives and false negatives, which allow us to measure whether the training of the network is optimal, in Image 6 the results generated by Matlab is.

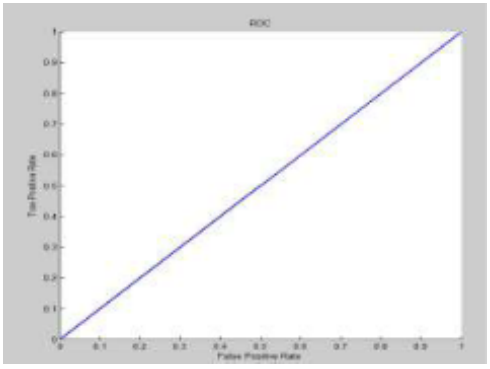

Fig. 6 False positives and false negatives

In the Fig. 6 shows that the line goes from the bottom left to the upper right corner, indicating a low rate of false positives, as compared to other research, the line matches the other mechanisms considered classifiers incorrect data, which shows that the training of the network is adequate and can proceed to perform intrusion detection with this network [7].

IV. RESULTS

In Matlab a small test interface allows authenticate if users are not intruders or is fed with data obtained above and designed. You need a user focus and detail their behavior, having as variables, date and time, with these data the ANN classifies the user pattern and shows the graph of variables, which shows the difference in behavior compared with another user. The answers can be "correctly identified user" implying that it is a correct access, the other answer may be "Intruder" when the user has a different behavior of others.

In Fig. 7 the results of normal behavior are shown in 8 and atypical respectively, in which detection the neuronal network and the performance achieved is checked through the interface used

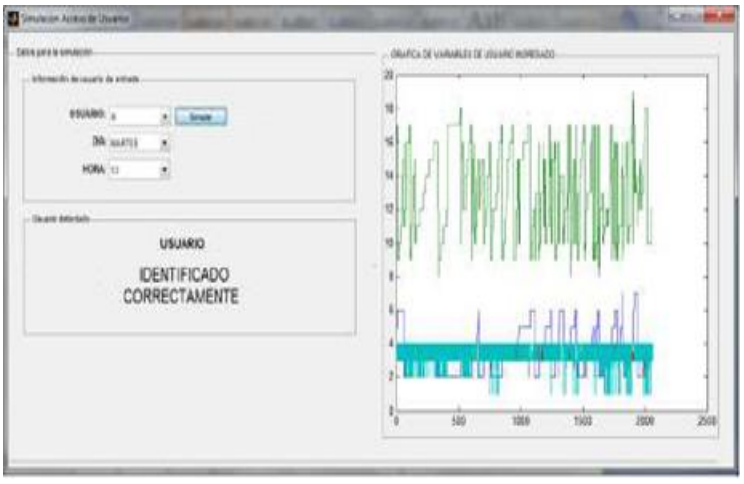

Fig. 7 Normal user behavior

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

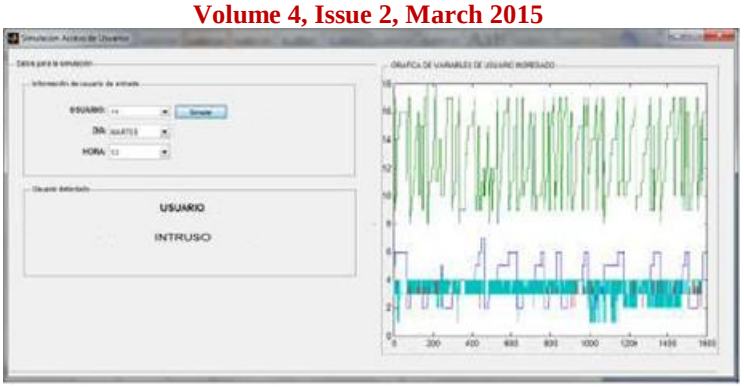

Fig. 8 A typical user behavior

The behavior may be normal if the input information corresponds to usual or common behavior in the system, while the graphs of the behavioral variables are identical. Furthermore atypical behavior in the input information must be of different actions to the normal pattern and the graphs of the variables do not match each other. This way we know which cases arise intrusions and in which cases users authenticate traditional way, through the recognition of patterns of user behavior.

V. CONCLUSION

The use of neural networks allow streamline pattern recognition and classification of them, thanks to that the results can be more real. For this investigation has been made to obtain different patterns of behavior for each user and you can get your identity correct manner. It was necessary debug information generated by a system of an institution that cannot be mentioned above due to confidentiality. This institution was identified intruders from thousands of users.

Because debugging information, patterns of behavior for each user and thus generate an automation control intrusion in the system were identified, and thus the profiles of behavior for each user were developed for this work the result by comparing the different behaviors and results of previous work is improved even when put to the test was conducted in a different institution of ancestors.

As future work, the concern of tracking in experiments in which different variables involved in computer systems, whether the interaction with the environment are studied arises, the signals emitted by each node of the network, as well as some movements user (physical), the study of these variables could generate a more accurate profile for each user and evaluate human behavior.

REFERENCES

[1] Anderson, James A.. Introducción a las redes neuronales. Cambridge, Mass. MIT Press, 1993.

[2] Casal, Jordi "Universidad Autónoma de Barcelona" Tipos de muestreo. Web. 28 de Marzo <2014. Minnie.uab.es/veteri/21216/TiposMuestreo1.pdf>

[3] Guevara Maldonado César Byron., "Reconocimiento de patrones para identification de usuarios en accesos informáticos". Universidad Complutense de Madrid, España 2012.

[4] Gurney, Kevin N.. An introduction to neural networks. London: UCL Press, 2002

[5] Pajares G., Inteligencia Artificial e ingeniería del conocimiento, ed. RA-MA, 2006

[6] Santos Peña Matilde. "Clasificación no supervisada con imágenes a color de cobertura terrestre". ISSN 1405-1436. Colegio de Postgraduados, 2012 http://redalyc.uaemex.mx/pdf/302/30215554009.pdf.

[7] Viñuela, Pedro. Redes de neuronas artificiales: un enfoque práctico. Madrid: Prentice Hall, 2004

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 4, Issue 2, March 2015

AUTHOR BIOGRAPHY

José Ernesto Luna Domínguez he is a computer engineer by the Autonomous University of the State of Mexico, teacher in computer science from the Autonomous University of the State of Mexico. He currently teaches at the Academy of networks in the University Center UAEM Valle de Chalco

Anabelem Soberanes Martín has a degree in Computer Systems Management from Universidad del Valle de Mexico. Master in Education from the University of the Americas, course Masters in Computer Science at the Instituto Tecnologico de Estudios Superiores de Monterrey, PhD in Educational Sciences at the Colegio de Postgraduados de la Ciudad de Mexico, has PROMEP profile member of the SNI-C, Certified Activity Coordinator Distance from San Diego Global Knowledge University, a Certified Instructor Taught courses by the CONOCER, working at the Universidad del Estado de Mexico with as a full professor of the Centro Universitario UAEM Valle de Chalco., served as coordinator of the Bachelor of Administrative Computing currently Leader and Member of the Academic Body Applied Computing.