INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

APPLICATION OF ARTIFICIAL INTELLIGENCE TO INTRUSION DETECTION SYSTEM

NIKHIL K. SONTAKKE

M.E. 2nd Sem, Dept. Computer Science &Engg, Sipna COET, Amravati, India.

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

Abstract: This paper presents a succinct review of the application of various Artificial Intelligence techniques and their advances in the design, development and application of Intrusion Detection Systems (IDS) for protecting computer and communication networks from intruders.IDS plays an important role in network security. These Artificial Intelligence algorithms have been shown to demonstrate their respective capabilities to produce high performance accuracies in various applications. This study is intended to serve as an all-in-one resource to practitioners and researchers in this inter-disciplinary endeavor while assisting them to take a critical look at the various efforts made so far in order to design and develop new and better algorithms that will have the capability of solving most of the yet unsolved problems in the field of Network Intrusion Detection.

Keywords: Intrusion Detection Systems, Artificial Intelligence, Network Security.

PAPER-QR CODE

Corresponding Author: MR. NIKHIL K. SONTAKKE

Access Online On:

www.ijpret.com

How to Cite This Article:

Nikhil Sontakke, IJPRET, 2014; Volume 2 (9): 888-894

INTRODUCTION

Intrusion Detection System (IDS) is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusion [1]. It is useful not only in detecting successful intrusions, but also in monitoring attempts to break security, which provides important information for timely counter-measures. Basically, IDS can be classified into two types: Misuse Intrusion Detection and Anomaly Intrusion Detection. Traditional protection techniques such as user authentication, data encryption, avoiding programming errors, and firewalls are used as first lines of defense for computer security. These have failed to fully protect networks and systems from increasingly sophisticated attacks and malwares. As a result, intrusion detection systems (IDS) have become an indispensable component of security infrastructure used to detect these threats before they inflict widespread damage. Recently, the use of Artificial Intelligence (AI) techniques has been employed in different data mining and machine learning classification and prediction modeling schemes. In addition to these, hybrid data mining schemes, hierarchical hybrid intelligent system models, and ensemble learning approaches that combine the base models with other hybrid machine learning paradigms, to maximize the accuracy and minimize both root mean squared errors and computational complexity, have also gained popularity in the literature [2].

In this paper, a succinct review has been carried out on the individual capabilities of various AI techniques in their application to network IDS. Such techniques include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Genetic Algorithms (GA) and Fuzzy Neural Networks (FNN). Attempts were also made to propose possible hybrid approaches based on these techniques.

2. Background

2.1.*Overview of Intrusion Detection Systems*

An intrusion detection system dynamically monitors the events taking place in a monitored system, and decides whether these events are symptomatic of an attack or constitute a legitimate use of the system. Figure 1depicts the organization of an IDS where solid arrows indicate data/control flow while dotted arrows indicate a response to intrusive activities.

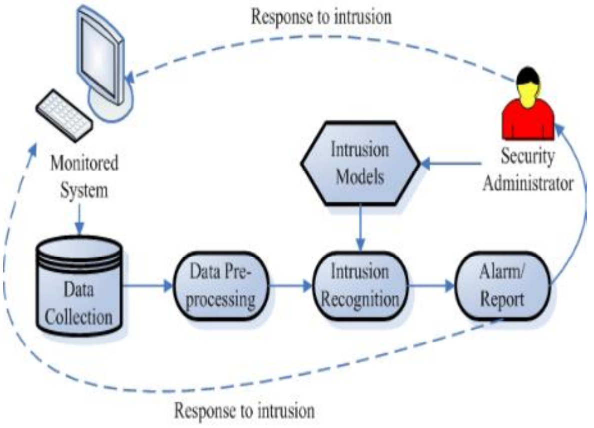

Fig. 1. Organization of a generalized intrusion detection system

In general, IDSs fall into two categories according to the detection methods the employ, namely (i) misuse detection and (ii) anomaly detection. Misuse detection identifies

intrusions by matching observed data with pre-defined descriptions of intrusive behavior. So well-known intrusions can be detected efficiently with a very low false positive

rate. For this reason, the approach is widely adopted in the majority of commercial systems. However, intrusions are usually polymorph, and evolve continuously. Misuse detection will fail easily when facing unknown intrusions.

One way to address this problem is to regularly update the knowledge base, either manually which is time consuming and laborious, or automatically with the help of supervised learning algorithms. Unfortunately, datasets for this purpose are usually expensive to prepare, as they require labeling of each instance in the dataset as normal or a type of intrusion. Another way to address this problem is to follow the anomaly detection model.

Anomaly detection is orthogonal to misuse detection. It hypothesizes that abnormal behavior is rare and different from normal behavior. Hence, it builds models for normal behavior and detects anomaly in observed data by noticing deviations from these models. There are two types of anomaly detection. The first is static anomaly detection, which assumes that the behavior of monitored targets never changes, such as system call sequences of an Apache service; the second type is dynamic anomaly detection. It extracts patterns from behavior habits of end users or networks / hosts usage history. Sometimes these patterns are called profiles.

Clearly, anomaly detection has the capacity of detecting new types of intrusions, and only requires normal data when building the profiles. However, its major difficulty lies in discovering boundaries between normal and abnormal behavior, due to the deficiency of abnormal samples in the training phase. Another difficulty is to adapt to constantly changing normal behavior, especially for dynamic anomaly detection.

In addition to the detection method, there are other characteristics one can use to classify IDSs, as shown in Figure 2.

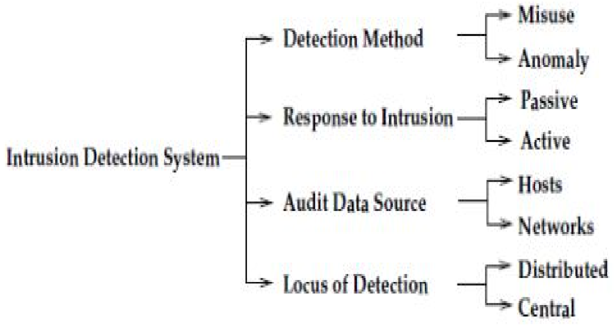

Fig. 2. Classification of Intrusion Detection Systems

2.2. Overview of Artificial Intelligence

The application of the capabilities of Artificial Intelligence techniques has been widely appreciated in Computer and Communication Networks in particular, as well as in other fields. This inter-disciplinary endeavor has created a collaborative link between Computer Scientists and Network Engineers in the design, simulation and development of network intrusion models and their characteristics. Computational Intelligence (CI), an offshoot of AI, covers all branches of science and engineering that are concerned with the understanding and solving of problems for which effective computational algorithms do not yet exist. Thus, it overlaps with some areas of Artificial Intelligence and a good part of Pattern Recognition, Image Analysis and Operations Research. It is based on the assumption that thinking is nothing but symbol manipulation. Thus, it holds out the hope that computers will not merely simulate intelligence, but actually achieve it. CI relies on heuristic algorithms such as in Fuzzy Systems, Neural Networks, Support Vector Machines and Evolutionary Computation.

AI is itself an advancement of the concept of its predecessor, Data Mining (DM). DM is the process of finding previously unknown, profitable and useful patterns embedded in data, with no prior hypothesis. It is the process of analyzing data from different perspectives, summarizing it into useful information and finding correlations or patterns among datasets in large data repositories. The objective of DM is to use the discovered patterns to help explain current behavior or to predict future outcomes. DM borrows some concepts and techniques from several long-established disciplines viz. Artificial Intelligence, Database Technology, Machine Learning and Statistics. The field of DM has, over the past couple of decades, produced a rich variety of algorithms that enable computers to learn new relationships/knowledge from large datasets [3].

3. OVERVIEW OF SOME ARTIFICIAL INTELLIGENCE TECHNIQUES AND THEIR APPLICATION IN IDS

A good number of studies have been carried out on the use of various CI/AI techniques to model various IDS strategies. Some of these techniques will be discussed in the following sections.

3.1. Artificial Neural Networks (ANN)

Attempts to artificially simulate the biological processes that lead to intelligent behavior culminated in the development of ANN. ANN is a mathematical or computational model that is based on biological neural networks. It consists of an interconnected group of artificial neurons which processes information using a connectionist approach to computation. In most cases, ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

Some of the versions of ANN, depending on which algorithm is used at the summation stage, include:

Probabilistic Neural Networks, Generalized Regression Neural Networks and Multi-Layer Perceptron Neural Networks. The most commonly used learning algorithm of ANN is the Feed-Forward Back-propagation algorithm.

B. Fuzzy Inference Systems (FIS)

Fuzzy Inference System include Type-1 Fuzzy System (fuzzy) and Type-2 Fuzzy System (fuzzy fuzzy). Type-2Fuzzy System (FS) was recently introduced as an extension of the concept of Type-1 Fuzzy. Type-2 FS have grades of membership that are themselves fuzzy. For each value of primary variable (e.g. pressure and temperature), the membership is a function (not just a point value). This is the secondary Membership Function(MF), whose domain, the primary membership, is in the interval (0,1), and whose range, secondary grades, may also be in (0,1). Hence, the MF of a Type-2 FS is three-dimensional, and the new third dimension provides new design degrees of freedom for handling uncertainties.

C. Support Vector Machines

Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of Generalized Linear Classifiers. They can also be considered as a special case of Tikhonov Regularization. SVMs map input vectors to a higher dimensional space where a maximal separating hyper plane is constructed.

D. Genetic Algorithms

Genetic Algorithm (GA) is a computing technique used as an exhaustive search paradigm to find exact or approximate solutions to optimization problems. GAs are categorized as global search heuristics. Its paradigm is based on a particular class of evolutionary algorithms that uses techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover. GAs are implemented in a computer simulation framework in which a population of abstract representations (representing chromosomes) of candidate solutions (representing biological creatures, or phenotypes) to an optimization problem produces better solutions. Traditionally, solutions are represented in bits (a set of 0s and 1s), but other encodings are also possible.

The evolution process begins with a population of randomly generated individuals and continues in generations.

In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Usually, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached.

Genetic Algorithms have been widely applied in almost all fields of research. The main property that makes genetic representations in computer simulations convenient is that their parts are easily aligned due to their fixed size, which facilitates simple crossover operations. The fitness function is defined over the genetic representation and measures the quality of the represented solution. Once the genetic representation of a problem has been obtained, and the fitness function defined, GA proceeds to initialize a population of solutions randomly, and then improves it through repetitive application of mutation, crossover, inversion and selection operators.

VI. CONCLUSION

Given the succinct review of the application of Artificial Intelligence techniques and its advances along with their excellent performance in literature, we conclude that further research in this area is necessary as there are very promising results that are obtainable from such techniques. The ensemblage and hybridization of various Artificial Intelligence techniques also indicate a bright future in the analysis of IDS and the prediction of its variousproperties for effective real-time network security.

REFERENCES

1. C.H. Lee, and Y. C. Lin, "Hybrid Learning Algorithm for Neuro-Fuzzy Systems", in Proceedings: Proceedings. 2004 IEEE International Conference on Fuzzy Systems, 2004, pp. 691-696.

2. H. Artail, H. Safa, M. Sraj, I. Kuwatly, and Z. Al-Masri, "A Hybrid Honeypot Framework for Improving Intrusion Detection Systems in Protecting Organizational Networks", Journal of Computers & Security, Vol. 25, 2006, pp 274 – 288.

3. Symeonidis, A. L., and Mitkas, P. A., 2005. Agent Intelligence through Data Mining. Multi-agent Systems, Artificial Societies, and Simulated Organizations Series 14: 200. USA: International Book Series, Springer Business Media.