RESEARCHERID
THOMSON REUTERS

[ICRTCET-2018]

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

A SURVEY OF AI TECHNIQUES USED IN INTRUSION DETECTION SYSTEM

Huda Mirza Saifuddin[*1] & Priyanka Padki[2]

[*1&2]Assistant Professor, Dept. of CSE,, RajaRajeswari College Of Engineering, Bangalore, India

ABSTRACT

Most companies today are making use of artificial intelligence to strengthen their security arsenal. Artificial intelligence acts as a cyber weapon, which can provide intelligent learning processes that can contribute hugely to the security aspect of an organization. In this paper we focus on intrusion detection, its classification and mainly provide a review on the artificial intelligence (AI) techniques that could be used in intrusion detection.

Keywords: *Cyber Security, Artificial Intelligence, Intrusion detection.*

I. INTRODUCTION

Artificial Intelligence – The term artificial intelligence refers to how we can enact machines humanly i.e. make decisions of their own, ability to think by themselves. AI enabled systems are having their own knowledge and are capable of taking actions and respond to any events. So how can we make machines knowledgeable? One method is to make the machine learn the rules that the experts know (Expert system) .The other way is to train the machines to work on past data and build their own set of knowledge.

Security systems continuously conform to the changing environment and threats that are involved in the digital play to provide persistent protection. Cyber reality, to some extent is quite distinctive and the methodologies used in security are custom fitted on a routine basis to the known assaults. Indeed, even with human interaction, the adaptation processes are likely to go slow and insufficient. Artificial Intelligence has a flexible and adaptable system behavior which can help in defeating the various deficiencies of today's cyber security tools. The perception of AI is seen in different ways. Sometimes AI is seen as a developing existential hazard for mankind, whereas sometimes experts have expressed caution at the expanding role of AI substances and have worries about their moral reasonability [1]. AI concentrates on how the human brain thinks, and how the people learn and choose to work while tackling an issue. These AI techniques can be used for creating a premise of Intelligent Software System. This paper provides a survey of all AI techniques that can be used for intrusion detection which can be used in cyber security

II. APPLICATIONS OF AI BASED TECHNIQUES

In order to prevent cyber assault numerous AI techniques can be utilized. Since we are heading towards a future in which the interaction with machines will be way smarter than the human beings, likewise even the technologies are developing at a high pace. This not only has benefits but is also leading to deadly threats and assaults.

A. Intelligent Agents

Intelligent agents are self-sustaining systems that allows computer systems to communicate with their peers to share information and must be equipped to respond with appropriate action in the event of an hostile situation. The conditions specified very much determines the mobility and flexibility of these systems. Their intelligence and synergy makes them a very qualified candidate for fighting cyber-attacks. A basic level of cyber -police along with portable intelligent agents can be built up in order to settle some lawful and business issues. But on a broader spectrum, a Multi-agent tool is required for the entire operational picture. For instance, a Neural network based intrusion detection and hybrid multi agent techniques which have been already depicted. The advantages of intelligent agents are that they are proactive, reactive and mobile in nature.

B. Neural nets

The neural nets history started when Frank Rosenblatt in 1957 created the perceptron- an artificial neuron. Perceptrons are meant to learn and tackle intriguing issues by joining in limited numbers. Neural nets comprise of countless artificial neurons. These neural nets help in a high pace parallel learning and decision making. Hence these neural nets are applied for learning patterns, pattern recognition, arrangement and so forth. Hence neural nets are quite beneficial in detecting and preventing intrusion. Not only intrusion, these can be applied in DoS identification, malware classification, spam recognition, zombie detection, computer identification and in forensic investigations. When installed in hardware or in a graphics processor component, because of their high speed the neural nets are quite popular in cyber defense. By the utilizing Field Programmable Gate Arrays (FPGA), a great advancement is reported for neural nets and their conformity to changing threats. Neural nets in AI have been quite advantageous due to their high speed of operation and warm detection.

C. Expert systems

The components of expert systems include knowledge base, inference engine and a shell. The knowledge base possesses information about expert knowledge that is captured and saved for a specific application. The inference engine is used for inferring answers in the context of current knowledge and also has further information about a circumstance. the shell has a void inference engine and a knowledge base, the software used must support this shell. This system helps for arranging security. To detect network intrusion we require rules, knowledge base and other configurations for which expert systems can be made use of. Knowledge base will contain intrusion specific features (rule set). All real time data packets must pass the rule set.

III. INTRUSION DETECTION SYSTEMS (IDS)

Intrusion Detection System is designed to be used for securing computer systems [2]. IDS monitors and defends computer systems against intrusions. It analyzes the events that occur in the system and finally decide the degree of their legitimateness.

A. Categorization of intrusion detection systems (IDS)

i. Criteria of categorization: Data collection and storage unit

- **Host-based IDS**: if IDS monitors a host i.e. it gathers data from host to be protected.
- **Network-based IDS** : if IDS monitors a network i.e. collects data directly from network in the form of packets.

ii. Criteria of categorization: Data analysis and processing unit

- **Signature-based IDS** : maintains a database of known signatures i.e. a comparison is carried out between the data collected from data collection unit and content stored on the database for detecting intrusion
- **Anomaly -based IDS**: the system reacts to the anomalous behavior in comparison with its history of previous behavior or profile of the system. Any variation detected is alarmed as an attack.

iii. Criteria of categorization: Method of generating response

- **Passive IDS** : these systems notify the respective authority when an attack is detected. They do not take any measures on their own to bring down the damage.
- **Active IDS** : these systems respond to attacks by taking an action against it.

IV. AI BASED TECHNIQUES FOR INTRUSION DETECTION

There exists many new knowledge standards that can be made use of for the decision making process. They may possess modular knowledge and hierarchical architecture. To incorporate these new standards we require a knowledge management system that is automated in nature. Going forward large knowledge bases may infer that they may be

ISSN 2348 – 8034
Impact Factor- 5.070

involved in more extensive applications. So the knowledge acquisition will itself demand extensive investment also focus should be on building large knowledge bases. . In this section we will discuss in details of all the AI techniques that could be used to thwart intrusion detection.

A. Decision Tree based approach (DT)

Decision Tree based approach has nodes, arcs and leaves. To label each node a feature attribute is utilised. Every arc from a node is labeled with a feature value for node's feature. Each leaf is labeled with a feature value for node's feature, then a decision tree can be used to classify a data point from root to leaf node. Each data item is defined by values of the attribute. It is initially constructed from a set of pre-classified data. Data items are partitioned based on the values of these attributes. Each partitioned set of data items goes through this process recursively. It terminates when all data items belongs to the same class. Intrusion is also a classification problem. Each connection here is scrutinized to check whether it is a normal connection or an attack. The scrutiny is done on the basis if a predefined database. Thus DT can be used in IDS. The advantages of using DT in IDS are that are capable of learning a model by the data and can predict a future event (like an attack). They work well with large datasets [3] and the multi-class classification aspect of DT is very useful for IDS. The limitations are identifying the attribute that best divides the data items into their classes.

B. Genetic Algorithm (GA)

Genetic algorithm is a growing area of artificial intelligence and is based on natural process of human evolution. Genetic algorithm processes more than one string at a time and thus works with numerous solutions instead of one solution at a time. Also previously found solution is efficiently used in the next iteration ie population of solutions are applied to get better solutions.

The first step in this technique is to generate Primary population of candidate solutions [4]. Second step involves generating fitness of all listed solutions. Then these solutions are assessed and modified based on a few stochastic operators. These operators help produce most successful solutions (i.e. selection, crossover and mutation). These fitness solutions lead to new population and the same procedure is applied again and again.

The process stops when either a maximum number of generations have been generated or when a required fitness rate has been obtained for the population. It works very well with complex problems. Its is very fast and can be used to classify attacks and also to make specific rules for different attacks. Hence it makes it a good candidate to be used in IDS. The advantages of using GA in IDS are that it suits well for optimization problems and inherent parallelism. The limitations of GA are that it has a high resource consumption, the quality of final solution reduces subsequently as they work on very large and complex problems and consumes a lot of time.

C. Support vector machine (SVM)

SVM is one of the best-suited learning algorithms for binary classification. Basically it was a type of pattern classifier based on statistical learning technique for classification. SVM is very successfully in the field of pattern recognition and lately its been proving its vigor in intrusion detection also [5]. Due to its great generalization nature capability and its ability to overcome the taboo of dimensionality it can be efficiently used in intrusion detection. It can be used to choose setup parameters and also its great speed makes it a good choice. These systems are able to scale easily and can learn large set of patterns. Its great strength is its ability to update training patterns dynamically. The advantages of SVM are its capability in selecting appropriate setup parameter, SVM are independent of dimensionality of feature space- thus can learn larger set of patterns and can scale better and update training patterns dynamically. Its limitations are that it being a supervised machine learning method requires labeled information for efficient learning, its pre-existing knowledge is required for classification which may not be easily available always and it processes raw features for classification - which increases architecture complexity and decreases accuracy of detecting intrusion.

D. Fuzzy Logic

This technique was in practice since 90's. it is a great and most suitable choice for problems where uncertainty is involved, intrusion detection is one such problem. Fuzzy logic based IDS require human expertise for determining fuzzy rules and fuzzy sets and these are time-consuming procedures. So AI usage supports us in developing these rules and fuzzy sets automatically and hence it reduces our time in developing a good intrusion classifier method. This leads to a shorter development time for making or updating an intrusion classifier [6]. The main idea applied here is to evolve two rules, the first rule is applied to normal class and other rule for the abnormal class using a profile data set with information related to the computer network during the normal behavior and during intrusive (abnormal) behavior. Applying fuzzy methods for IDS been advantageous when compared to classical approach. Fuzzy Logic advantages are that it is well suited for scenarios involving problems based on vague, improper and incomplete data and is very useful for solving problems that are not easy to model mathematically. Its limitations are that it needs more fine tuning and simulation before it's operational.

E. K-Nearest Neighbor (kNN)

System behaviors can either be normal or intrusive. When the need arises for classifying these systems based on their normal or intrusive behavior, one can make use of k-nearest neighbor technique [7]. These behaviors can be treated as some kind of system calls that require attention. These system calls can be combined to form a complete document that needs to be executed. Since the content of the document are system calls, one can perform classification of these system calls using a kNN classifier. The classification will be done by calculating the Euclidean distance of each of these system calls. The lesser the distance, the better the document fits into the group or cluster of similar behavior. kNN Advantages are that its easy and fast to implement and categorization of a sample can have many class labels. Its limitations are that its computational costs shoot up when there is an increase in the number of neighbors, it is sensitive to the local structure of the data and it suffers a large search problem to find the nearest neighbors.

F. Artificial Neural Network (ANN)

An information processing model inspired by the human biological nervous system such as the brain is an Artificial Neural Network. This model helps in processing information on wide basis. The structure of ANN has a main component called as a neuron. Each of these neurons independently process information and are linked together with other neurons to form a network [8]. The neurons with their processing information are summed up to form an element which will be fed to an activation function. This feeding process can either happen in a single layer or in multiple layers. But, how do these neurons basically learn or process the information fed into them? This can be achieved by training the entire network based on some example of past data or behaviors. Once the network has been trained, whatever information will be processed will classify itself based on the corresponding class of the trained network. Hence this quality of classifying based on trained network, helps in detecting intrusion attacks. ANN advantages include that it does not need to be reprogrammed as it learns by itself due to self-learning capability and in cases of failure, ANN can still keep working due to its parallel nature. Limitations are that ANN needs training to be functional and needs to be emulated.

G. Bayesian curve/network (BN)

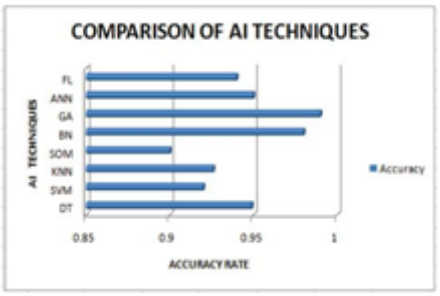
A Bayesian network, when viewed graphically comprises of nodes. These nodes are used for representing some random variables. The structure of Bayesian network has directed arcs in it, which follow a direction from parent nodes to child nodes. This kind of direction favors an indication of conditional dependence. This directed graph has an acyclic structure, due to which cycles are forbidden in this network. These set of nodes then form a complete joint distribution [9]. This distribution will be later defined by some conditional probability distribution so the each node has a relation to its parent node. Hence, this kind of a structure caters the means for decomposition and for rapidly computing conditional probability distributions. Detecting IDS in the network using Bayesian Network is possible and henceforth it can be normal or anomalous. BN favors the advantages to predict result of an intervention before intervening. Limitation of BN is that its knowledge tends to be unstable and it demands high maintenance efforts and costs.

H. Self-Organizing Map (SOM)

A neural network model that has widely been used for analysis and visualualization for data that has high dimensions is a Self-Organizing Map. It hails from a family of competitive learning network [10]. It inherits a quality of mapping inputs that are having high dimensions onto a regular 2D array of neurons. It makes use of a 1D or 2D topology preserving map describing relationships among the neighboring points in the dataset. In regard to intrusion detection SOM classifies the network traffic and prepares a topological map by understanding the underlying data, which is further used for intrusion detection. Advantages of SOM are that it works well with non-linear data set and is simple and easy. Limitations are that as the number of neurons increases the computation increases thus it could become time-consuming.

V. COMPARATIVE ANALYSIS OF AI TECHNIQUES

The below figure 1, gives us the comparison of accuracy rate [11] of different AI techniques discussed above. The figure shows that Genetic Algorithms suits best and provides most accurate results with respect to Intrusion Detection.



The Figure 1, depicts a comparative analysis based on survey of various above referenced research papers. The comparison infers that Genetic Algorithm (GA) methodology is most widely used and also yields convincing results when compared to other methodologies.

VI. CONCLUSION

AI is the future of defense against various cyber-attacks. In the above discussion we have highlighted the top 8 techniques that could be tailored effectively to thwart intrusion attacks and can further be extended to detect and prevent the same. Many of these techniques could also be combined to give a multiple AI approach, serving as an IDS. We also provide a comparison based on various such implementations carried out for IDS.

REFERENCES

1. *Fernando Maymí, Robert Bixler, Randolph Jones, Scott Lathrop," Towards a Definition of Cyberspace Tactics, Techniques and Procedures", 2017 IEEE International Conference on Big Data (BIGDATA), 978-1-5386-2715-0/17/$31.00 ©2017 IEEE.*
2. *Gulshan Kumar, Krishan Kumar, Monika Sachdeva, "Artificial Intelligence Based Intrusion Detection Techniques - A Review", © Institute of Management Studies, Noida.*
3. *Manish Kumar, Dr. M. Hanumanthappa, "Intrusion Detection System Using Decision Tree Algorithm", 978-1-4673-2101-3/12/$31.00 ©2012 IEEE.*
4. *Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis, "Intrusion Detection System Using Genetic Algorithm", Science and Information Conference 2014 August 27-29, 2014 | London, UK.*
5. *Jayshree Jha, Leena Ragha, "Intrusion Detection System using Support Vector Machine", International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868.*

6. *Depak.S1,Manikanta.A1,Thanuja.R2,Umamakeswari.A "Comparative Analysis Of Different Techniques Used In Anomaly-Based Intrusion Detection", International Journal of Pure and Applied Mathematics, Volume 115 No. 7 2017, 175-182, ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)*
7. *M.Govindarajan, Rlvl.Chandrasekaran, "Intrusion Detection Using k-Nearest Neighbor", 978-1-4244-4787-9/09/$25.00 ©2009 IEEE*
8. *L.P. Dias, J. J. F. Cerqueira, K. D. R. Assis, R. C. Almeida Jr, "Using Artificial Neural Network in Intrusion Detection Systems to Computer Networks", 978-1-5386-3007-5/17/$31.00 ©2017 IEEE.*
9. *Nagaraju Devarakonda, Srinivasulu Pamidi, Valli Kumari V, Govardhan A, "Intrusion Detection System using Bayesian Network and Hidden Markov Model", 2212-0173 © 2012 Published by Elsevier Ltd. doi: 10.1016/j.protcy.2012.05.081*
10. *Liberios VOKOROKOS, Anton BALÁŽ, Martin CHOVANEC, "INTRUSION DETECTION SYSTEM USING SELF ORGANIZING MAP ",ISSN 1335-8243 © 2006 Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovak Republic.*
11. *Janu Gupta ,Jasbir Singh, "Detecting Anomaly Based Network Intrusion Using Feature Extraction and Classification Techniques", International Journal of Advanced Research in Computer Science, ISSN No. 0976-5697.*