# Role of Artificial Intelligence in Intrusion Detection System, Overcoming Securities and Vulnerabilities


DECEMBER 1, 2017

RUTGERS BUSINESS SCHOOL

BY: TEJAS THAKUR
JINIL AMIN

# Contents

# Abstract

As we are moving toward digital world, the generation of data per session is following the exponential path. Which is a tedious task for a security administrator to overview the whole data and identify the intruder? However, with the help of the Machine Learning and AI system, we can able to do the same task considering the error (delta) to be at a minimum. Over the period various Intrusion Detection system has been developed to study and understand the distinct thought process of the cyber intruder that break any system. The most common example of this is a honeypot. That system can direct malicious traffic and gain the knowledge from it. We will discuss here Generic algorithm, fuzzy logic, and Artificial Neural Network and it's future development trend to build more secure system.

# SECTION 1

## Overview of intrusion detection system

Intrusion detection is very much important security process to secure systems and networks. It attempts to identify any nonregular activity in an organizational network, or on a particular host, by analyzing the known or evolving knowledge monitoring. Such events may be initiated by an external intruder or internal misuse. An intrusion can be defined as "**Act of the person or automated system of proxy attempting to break into or misuse a system in violation of an established policies.**" Or "**Any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource.**" [7]

To secure a system, we required Intrusion Detection Mechanism. This method could be hardware based or software base, which can be used for monitoring and detection of user behavior and traffic to identify any illegitimate attempts to break the system. This unauthorized access can be manipulated through the network by malware or any other mean. ID has been used to protect various organizational's IT systems, with the approach of prevention such as authentication and(or) access control mechanisms. ID systems classified as host-based and Network-based according to deployment criteria.[8] Network-based IDS monitors the traffic over the network and looks for attacks, while in host-based IDS a client is installed on the host which controls the host and generate an audit trail. In each domain of network and host base the ID system classified as misuse detection[8] and anomaly detection[8] as per the organization detection approach.

### • Misuse detection

Also known as signature detection which has the power to exploit the already known vulnerabilities of the system to differentiate between anomaly attack patterns and known ID signatures. The main drawback of misuse detection method is that this system will only detect intrusion or pattern for which they are train for or the known signature, i.e., they will not be capable of detecting unknown attacks .

The IDS can operate as centralized, standalone or hybrid integrated with applications that create a distributed system. We can distinguish IDSs as per behavior, i.e., Passive and Active. The passive system merely monitors and generate alerts concerning desire system . whereas Active system is designed to detect and respond to online attacks, attempt's to patch system holes before getting identified by external entities or act proactively by discarding potential intruders or blocking services. [10]

## • Anomaly detection

The anomaly detection detects the issue in intrusion detection associated with a change to the regular system or user behavior. The design of this is derived from the thought that the signature of attacks is significantly different. Anomaly detection has the power to detect unknown attack or variety of known attack if those attacks change considerably during monitoring of the system. The deviations are regardless of the regular user concerning program usage and the policy configuration. The major drawback of the anomaly detections method is that well-known attacks may not be identified, technically if the attacker fits the existing user profile. Another disadvantage of anomaly detection approach is that a malicious user who knows that he or she is begin profiled can change the profile slowly over time to virtually train the anomaly detection system to learn the attacker's malicious behavior as Normal.

The effectiveness of anomaly detection approach is the prior knowledge of the vulnerabilities s in the target systems is not required. Thus, it can able to identify known intrusion and unknown intrusion as well. Also, this approach can detect intrusion that is achieved by the abuse of legitimate users or masqueraders without breaking security policy. However, no system is foolproof the main drawback is high false positive detection error, high computation power and handling gradual misbehavior.

The IDS are classified into [10]

## • Network-based IDS:

The network base Intrusion detection system is responsible for detecting inappropriate anomalous or any kind of data which may consider not appropriate or unauthorized for for the environment. Network IDS design in such way that it shall receive all the packets of the particular network segment ( in fig. 1 switched network). Generally taps or port mirriong [3] technique use to receive packets .

These system( majorly) are pattern base that means they required a signature to detect the intrusion. Also these systems accuricy is depended on the tuning of the system toenvironment.

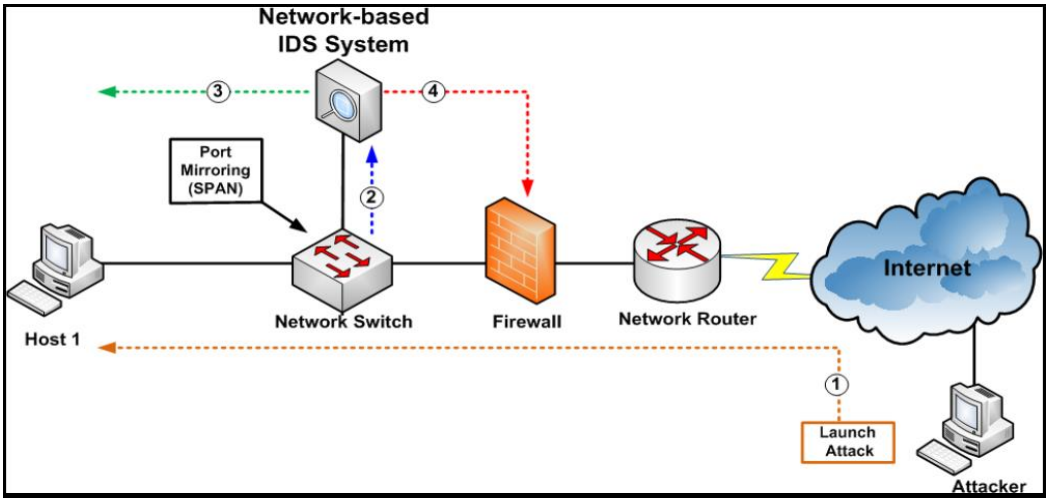



*Fig.1*

• Host-based IDS:

The host base IDS system reside on the specific host .This system scans the host system activities , typically it looks for the host event log, application logs and database logs as per configuration. This system is entirely dependent on the logs files of the respective environment . if somehow if log files data curropted or attacker able to manipulate log file information the HIDS fail to detect occurance of attack . the information gather by system is stored in secure database and compaired with the knowledge base to detect an anomaly.

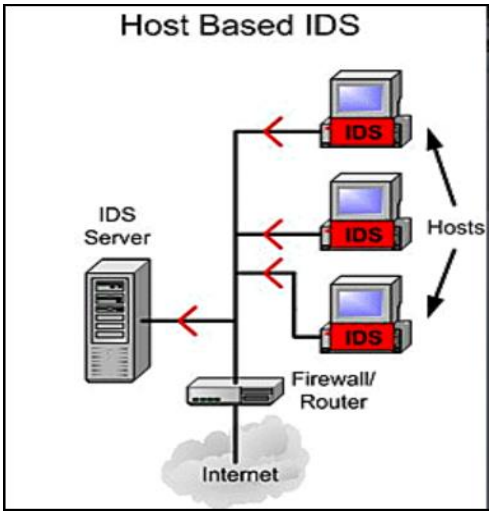



*Fig. 2*

• Hybrid IDS

The network information and host agent data are combined to form a comprehensive view of overall system underlying the IDS. The hybrid IDS work online with encryption and data destination in single host this is the primary need to design this solution.With this approach, the source and target host/ entity can decrypt the network traffic. In general, the intrusion detection system is not core part of any solution; This component will monitor, and trigger an alarm on detecting the intruder and area where the breach occurred, the response of the system depends on the design and configuration of system. Addition to this the triggered alarm shall not provide end to end system security itself; it will help to indicate that some sort of potentially malicious activity is being attempted.

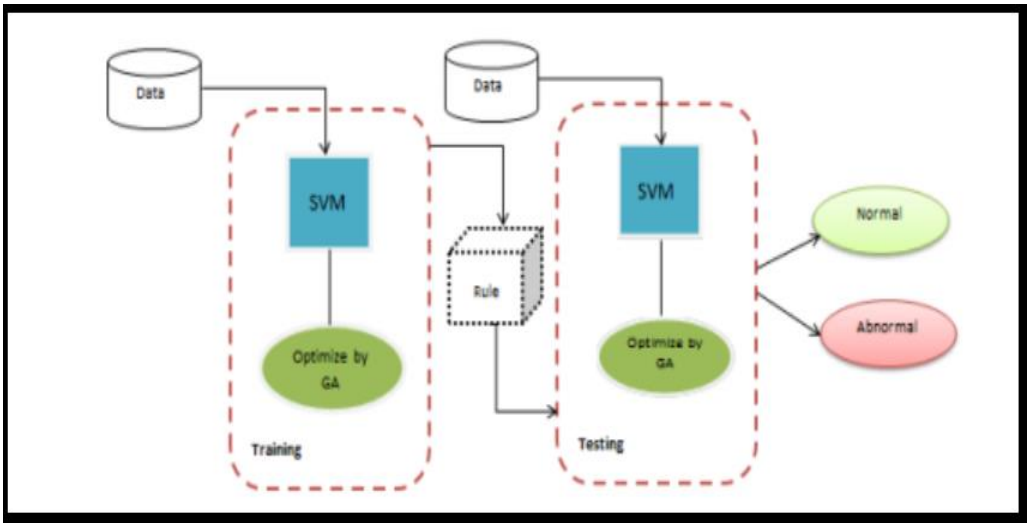



*Fig.3*

IDSs can operate on either a periodic or continuous feed of information (Generally known as Real-time IDS and Interval base IDS) which reflect the different methodologies[10] to be in place. Audit trail analysis is the general method used by interval base system. On another side, the IDS deployable in real-time environments with the primary focus on online monitoring and analyzing system events and user actions which may affect the environment. With online processing, an ID performs verification of various system events and logs . in General, Packets stream is rigorously monitored. With this type of processing, The ID system uses the existing knowledge base of ongoing activities over the network to detect possible intrusion attempts. The figure below shows the Generic view of IDS:

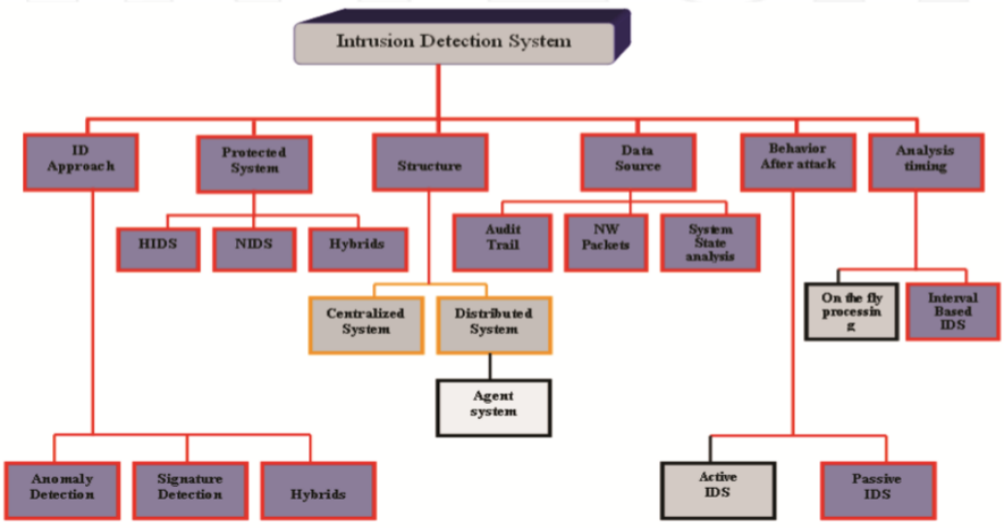



*Fig 4*

## *Problem Definition*

Recently, the primary issue faced by the computing systems is the number of intrusion attacks has grown, and intruders become more and more intelligent and flexible as they evolve over the period.
Its the war between good and bad, build something unbreakable and attacker tries to find a way to in. The reason is that, advance automated hacking tools developed with the latest technology, and these tools, are loaded with various system holes and backdoor entries information which are available on the internet. Although, This problem can be solved by using right application which is designed and configured to detect and prevent such attack. **Intrusion detection (ID)** is an appealing concept as the existing techniques used in computer security are not sufficient to handle dynamic and evolving complex nature of Information systems and security.

The intrusion detector learning goal is to design a predictive model (i.e., a classifier) which is capable of differentiating between "malicious" connection called intrusion and "clean" standard connections. It depends on various attributes that are collected from the header of packet floating on network and system audit trail files (behavior during the contact). Such system could be built using different methodologies as statistical approaches, genetic algorithms, fuzzy systems and neural networks.

To evaluate any recently developed intrusion detection system, we assess the Deep learning method applied to the data collected from hybrid IDs and generate a ROC curve[2] for finding out how good is the given IDS for detecting new and known attacks. There are 39 different attack types as founded by MIT Lincoln Labs[1] that fall into four main Categories:-

1. Probing Attack - Intruder scans the network to gather information and find a known weakness.

2. DOS (denial-of-service) - Intruder makes a computing or memory resource too busy or too full to handles legitimate requests thus denying authorized users access to a machine.

3. U2R (User to root): unauthorized access to root privileges are exploited. Here an Intruder starts out with standard access to a user account on the system and can utilize vulnerability to gain root access to the system.

4. R2L (remote login): unauthorized access from a remote machine. Intruder sends packets to a computer node over a network, then exploits the machine's vulnerability to gain access as a user illegally.

Another major issue, current intrusion detection systems (IDS) examine is that all the features are required for detecting any intrusion pattern.So it is one of the primary requirement to have an AI technique that captures all the necessary elements eliminating nose implanted by attacker thereby increasing efficiency and boost time speed. We will future drill down into details how an appropriate AI technique can help reveal and protection from intrusion.

# SECTION 2

## Basic Techniques & Shortcomings.

### Decision tree

It's a predictive modeling technique to builds a simple tree-like structure model to find a pattern. Decision trees algorithms have been tested and applied to many practical intrusion detection systems. These techniques have also been used to classify and characterize scanning activity as well as to detect novel attacks from cyber adversaries. Once the decision tree is built, it has considerable potential to reduce the amount of data required for analysis which help to identify malicious activity and provide analytics of inside of regular traffic and malicious data.[2]

The general flow of how decision tree algorithm works.

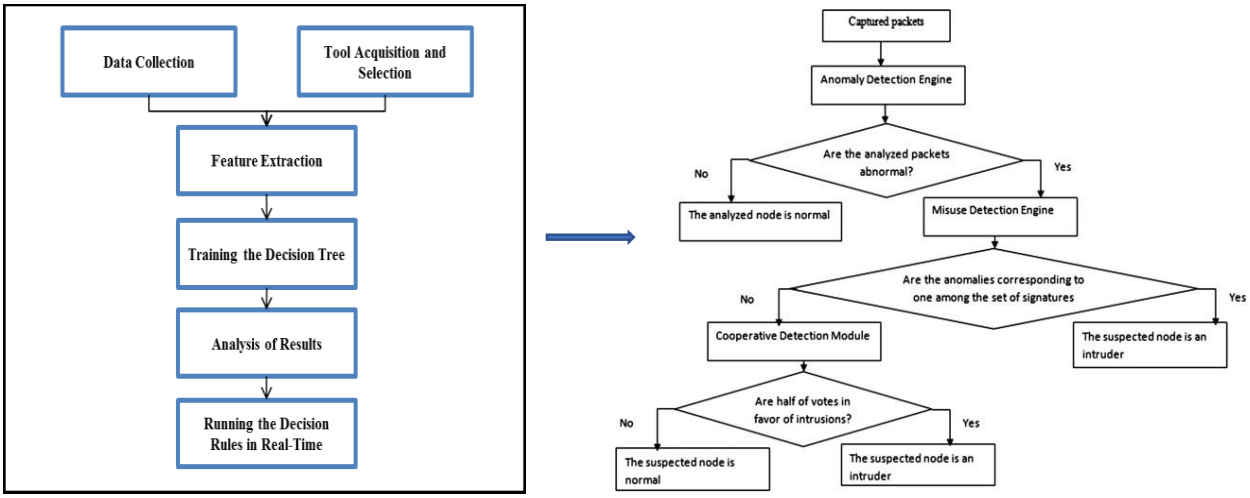



*Fig 5*

Decision trees provide a lot of benefits from a cybersecurity perspective.

- Supplement honeypot analysis by learning from adversary trends and creating rules to detect malicious activity.
- Supplement penetration testing efforts by learning from pen tester's action and creating rules to detect their tactics technique and procedures
- it uses Prioritization techniques; tag identified alerts on various priority level as per system security goals.
- Identify and characterize known scanning activity
- supplement an incident response team's recovery and monitoring process by flapping repeated intrusion attempt early.

Drawbacks

- Time Complexity

- Expert Human intervention
- Ineffective in Feature Extraction.
- Efficiency – misclassification error rate high.

**However, this approach is being static, i.e., the system will not intelligently identify and take a decision and human intervention required to verify. As the dynamic nature of the various network components and hosts with a different factor related to them.**

## Decision Rule

This technique typically involves the application of a set of association rules, and regular subset patterns to classify the audit trail data for the signature detection. This Approach uses rules to represent known intrusions. They are matched with multiple signatures or patterns in the activity data. These rules are not only specific to the identified intrusions, but they can also be specific to the target of the protected network. As the same intrusion may leave different evidence on a different system depending on the network setups. [3]

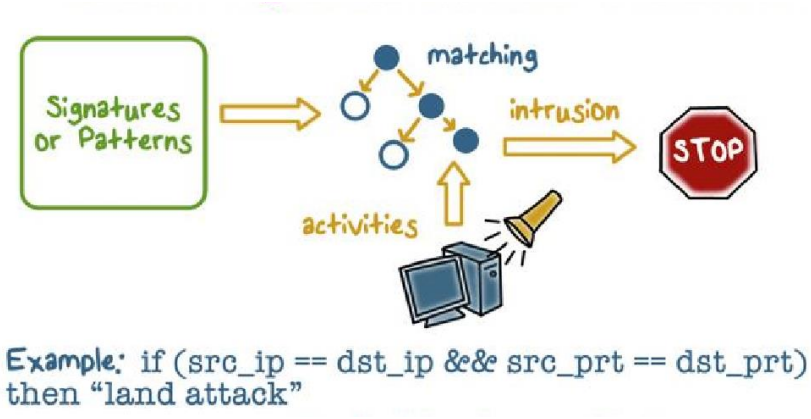



*Fig. 6*

As per the Figure, we first list out all the signature pattern. Then generate a complex decision tree, make rules from that tree To classify the intrusion for damage detection probability. If the rule is satisfied then alert is generated.

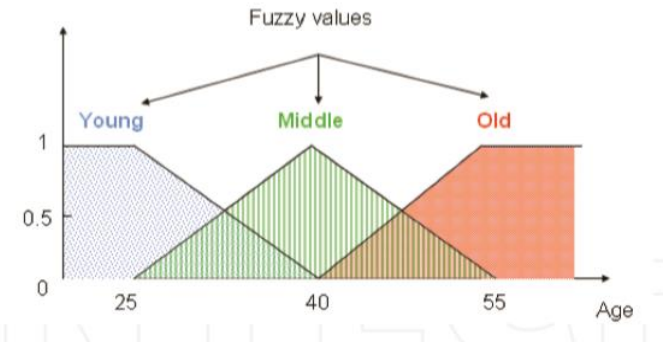
**The advantage of using rules is that they tend to be intuitive and straightforward, unstructured and less rigid. But, The drawback is that rules are difficult to maintain, and in some cases, are inadequate to represent a different kind of information.**

# SECTION 3

## Advance Techniques

### Fuzzy Logic

This technique has been in use in the area of computer and network security especially in intrusion detection for two main reasons. Firstly, various numerical features like CPU usage time, connection interval, etc. can potentially be seen as fuzzy variables. Secondly, the concept of security itself is blurred. In other words, the idea of fuzziness helps to smooth out the abrupt separation of normal behavior from abnormal behavior. That is, a given data point falling outside/inside a defined "normal interval," will be considered anomalous/normal to the same degree regardless of its distance from/within the interval.[3]




Fuzzy is used to summarize qualitative variable into quantitative measure scale. This increases the effectiveness of more "dynamically" labeling of unidentified class/attacks. It is used primarily to optimize static & dynamic Machine learning method to reduce false positive rate.

*Fig 7*

Fuzzy logic applied to a **static** method.

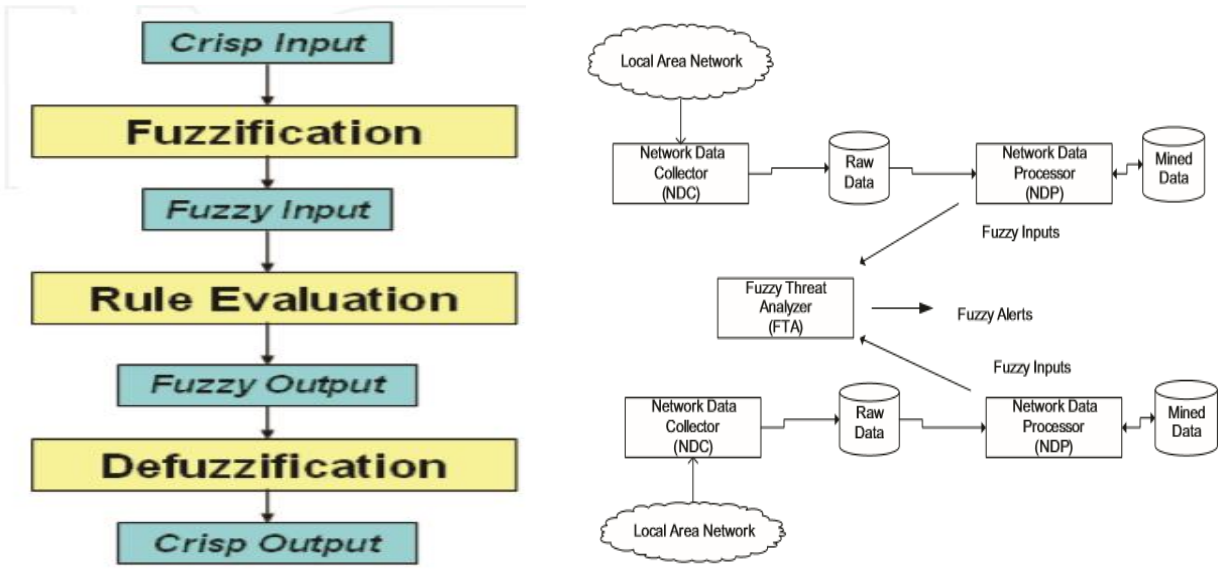



*Fig 8*

Here, we apply fuzzy logic to the given raw data obtained from Hybrid IDS; the Fuzzification Process generates a weighted fuzzy set. Which can be directly applied to the decision tree to create a dynamic rule using Advance RIPPER[4] algorithm.

It is noted that Accuracy of the static method increases on applying Fuzzification process and a static method can detect a new threat in IDS efficiently.

**The only Drawback is that Rule are not Automated, hence Labor intensive human expertise is required.**

## Artificial Neural Network

Its inspired by the human nervous system, which is connected through neurons. Neural networks can understand and learn by training and can be used to identify complex trends. There are two types of ANN architectures, i.e., feed-forward ANN and feedback ANN. In feedforward ANN, the signals move in only one direction from input to output. In feedback ANN, the signals run in both directions. ANN-based intrusion detection can be helpful to eliminate the shortcomings of rule-based IDSs. They are more efficient if adequately trained with both normal and abnormal datasets.[Ref]

For IDS, we use two main approaches for detecting an intruder. If the IDS system is just deployed in the environment, then we go with Single-layer Perceptron or else Multi-layer Perceptron.

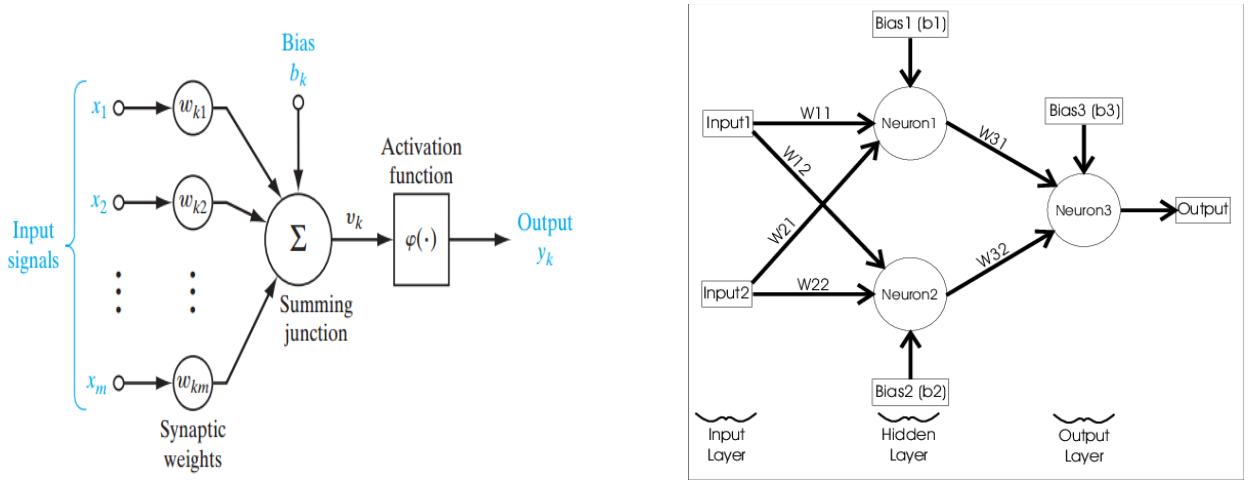



*Fig 9*

Multi-layer Perceptron ANN provides better scaling of raw data feature. ANN is a dynamic method which self-evolves with new feed and efficient when detecting a new threat in the system. A supervised training of known Intrusion attacks is run on ANN so that ANN start evolving & detecting a pattern . after that it is deployed in run-time environment to predict emerging threats.

*Solution:-*

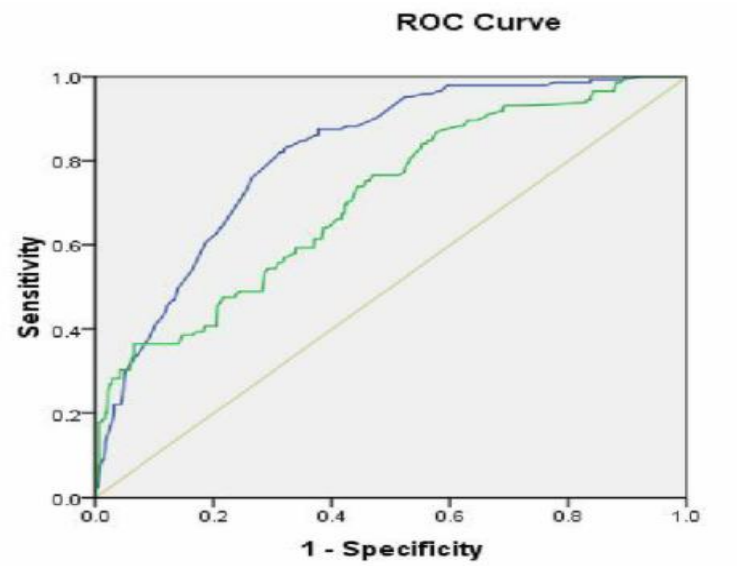
Multi-layer ANN without Fuzzification on IDS - shortcoming

On applying ANN, it is found that IDS becomes very effective on detection but significant drawback happens to be that the input data from IDS being unstructured causes unidentification of Anomaly & outlier in that data which result into the high variance of classified threat.

Multi-layer ANN with Fuzzification on IDS

Provides the best result in term of Dynamic labeling of normal & abnormal behavior.

- feature extraction is performed using PCA.
- normalization of raw data for generalizing the effect of all the features.
- Fuzzification to quantize features with weights.
- ANN for classification & clustering.




The figure shows a comparison between the accuracy obtained from above two approaches. It can be seen that ANN with fuzzy logic (Blue line) appears to have 8% more accuracy rate in comparison to ANN without fuzzy logic. Both of them provide low false rate but it's better to have Fuzzy logic as the Complexity of IDS and Intruder is increasing exponentially.[5]

Fig.10

*Proposed HID System: -*

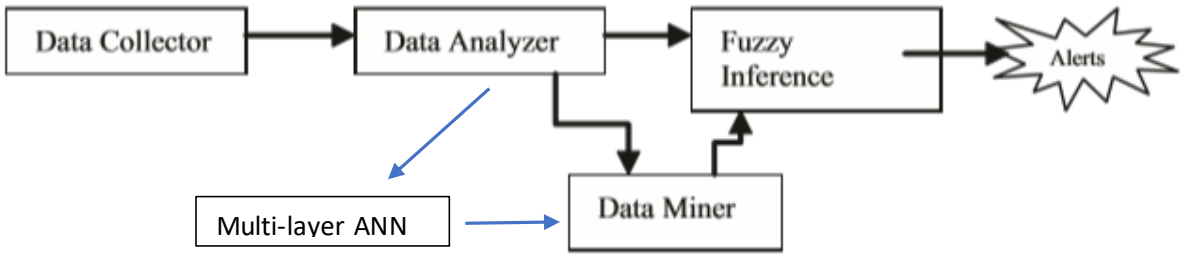



*Fig. 11*

**Drawbacks such as human labor, automation & low accuracy are overcome by implanting the above HIDS.**

## Genetic Algorithm

They are widely used in many areas of computing to solve a complex problem. **It provides robust, adaptive, and optimal solutions**. It is inspired by biological processes such as natural selection, evolution, the theory of mutation, and genetic inheritance.

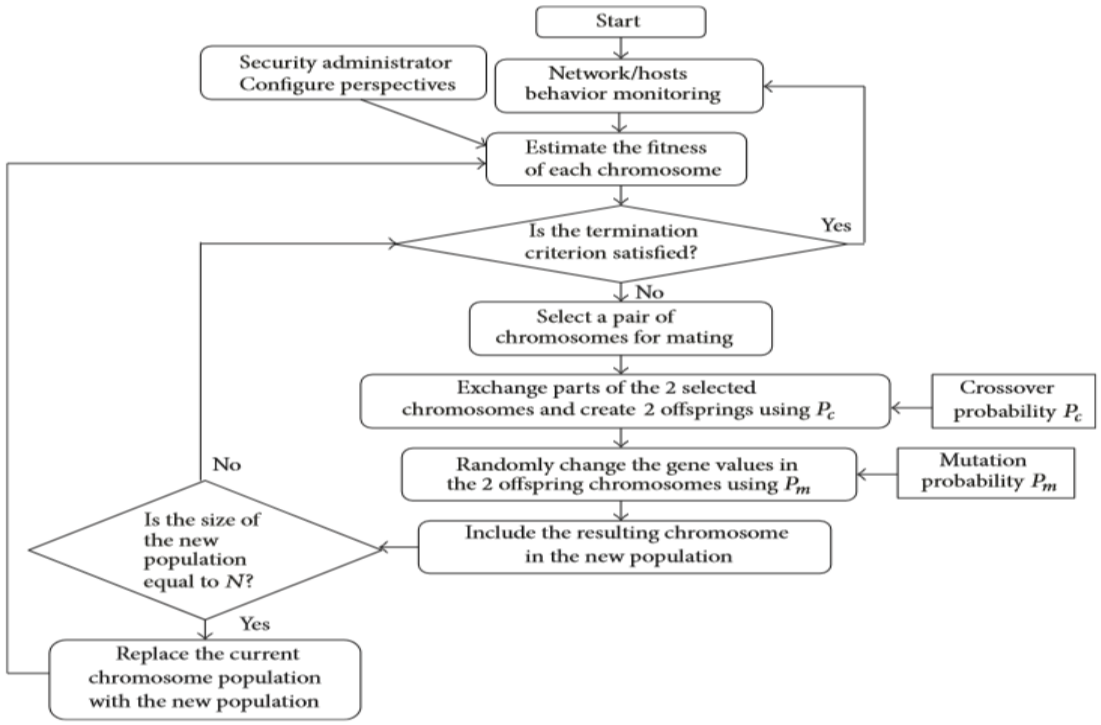



*Fig. 12*

Figure Provides the work-flow of the genetic algorithm.In this algorithm, the selection module derives the most suitable answer for some specific problem. Crossover module selects various parameters that are obtained during different solutions to get new solutions. Mutation module changes one or two features to get optimality in the algorithm. [2]

Genetic algorithm is used in IDS for classification of security attacks and for generating specific rules for different security attacks. It takes into account many Features such as connection status, protocol type, and network services to create rules. The detection mechanism is trained on the specific dataset so that it can accurately identify and classify security attacks. Figure below shows how Genetic algorithm gets integrated with different IDS.

| Genetic Algorithm | | | | | | |
|---|---|---|---|---|---|---|
| Primary Assumption | | | Intruder Type | Intruder Behavior | Type of Approaches | System |
| System activities Cinsider to be | Normal | Applies To | External Intruder | Attempted to breakins Masquerade attacks, penetration of the security control system leakage , Denial of service . Malicious use | Anomaly Detection | Network base Host base Hybrid |
| Intrusive activities | Distinct evidence | | Internal Intruder | | Misuse Detection | |

*Table 1*

*Solution: -*

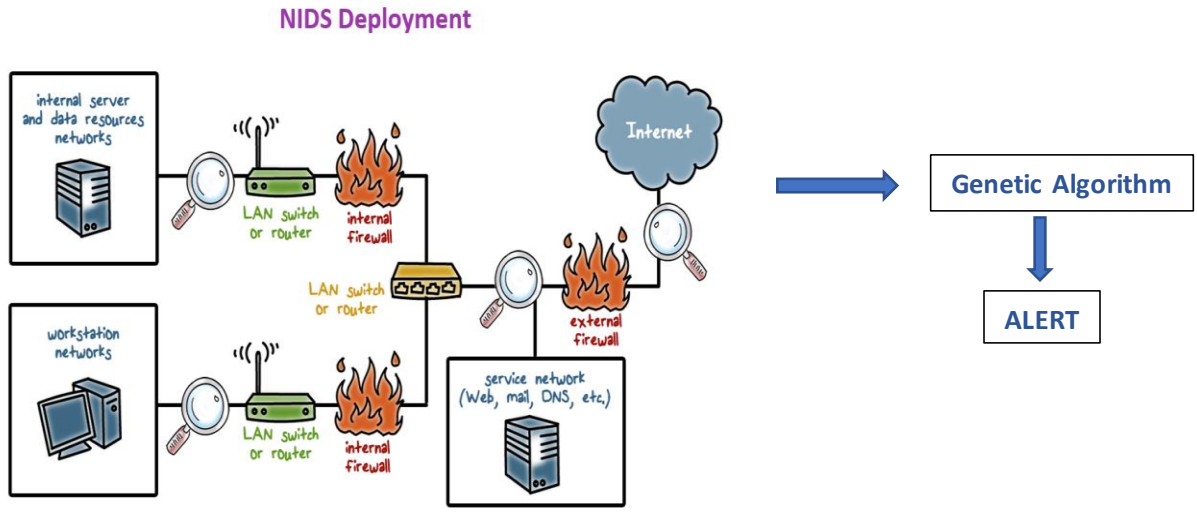



*Fig. 13*

- First, we deploy a NIDS with an inline sensor for detecting any abnormal activity among the nodes. Here we enable both internal & external firewall for noting the Behavior of packets coming in and going out.
- The firewall will filter out any malicious packet with unstructured IP address. After this filtering process, the raw data of the remaining is further sent to the Genetic algorithm as collected by the sensor.
- Genetic algorithm filter raw data and select packet as per feature extraction followed by the crossover of the raw data to normalize the attribute, finally mutation step is performed to optimize the result.
- It is noted using Genetic Algorithm in conjunction with firewall provides a good optimize the probability of detection by 95 % on various security attacks.

*Advantages*

- Genetic Algorithm detects new emerging threats in comparison to rule-based detection.
- Optimization of Non-linear structure of complex attacks is achieved using the Genetic algorithm.

## Conclusion

Security risk for an organization met daily basis. Because of this, it is very much needed to consider more complex and self-learning security implementation just ordinary firewall system.in This paper discussed the various types Intrusion detection system, the advantages and disadvantages of the solution for each type of System (IDS) and future development. This paper finds out the evolution of intrusion detection system and its technique. Recent days intrusion detection systems have become advanced to great extent.

Here we first have discussed the basic technique like decision tree and decision rule. Which are not much efficient and failing to prove the capabilities over the time. Hence, we come up with new approaches which are inclined toward self-learning and evolving .in this category contain the fuzzy logic, Artificial neural network and genetic algorithm.

Each approach is evolved and offering better feasibility the previous on however the complexity and cost factor increasing as well to implement and maintain those more advance system.

These, IDS system is configure as per the domain requirements is not necessary to implement a super heavy system for some manufacturing domain. however, its much needed for the domain like BFSI. so its designer call to opt for which approach and design.

Hence the genetic system is best we are not saying it's 100% proof but yes, the proposed design is best till time and provide maximum benefits of others as we have picket the best of every system which provide the attributes contributing toward intrusion detection in the area of Artificial intelligence and Machine Learning.

## Reference

Below are the potential references which we find to be potential. We may find other references than mention below


1. Comparative Study of Machine Learning Algorithm for Intrusion Detection System, K. Sravani and P. Srinivasu [ 2017]
2. Intrusion Detection Systems Based on Artificial Intelligence Techniques in Wireless Sensor Networks. By Nabil Ali Alrajeh and J. Lloret [Sept 2013] [2-4]
3. Artificial Intelligence and Security: Current Applications and Tomorrow's Potentials by Daniel Faggella [Sept 2017]
4. Intrusion Detection System Using Deep Neural Network for In-Vehicle Network Security, [Min-Joo Kang, Je-Won Kang [June 2017]
5. Using Artificial Intelligence in Intrusion Detection Systems by Matti Manninen, Helsinki University of Technology [2016]
6. A Universal High-Performance Correlation Analysis Detection Model and Algorithm for Network Intrusion Detection System, [Hongliang Zhu, Wenhan Liu, Maohua Sun, Yang Xin] [Feb 2017]
7. The Use of Artificial Intelligence based Techniques for Intrusion Detection [2010] [Dr. Gulshan Kumar Ahuja , Krishan Kumar Saluja, Monika Sachdeva] [2010]
8. Artificial Intelligence and its Application in Different Areas by Avneet Pann , [April 2015]
9. Machine Learning Techniques for Intrusion Detection [Mahdi Zamani, Mahnush Movahedi] [2013]
10. Intrusion Detection Systems, *ISBN 978-953-307-167-1, 334*
11. http://www.omnisecu.com/security/infrastructure-and-email-security/types-of-intrusion-detection-systems.php