

# Intelligent Intrusion Detection System using Machine Learning Algorithm (HMM)

Prerana Agale<sup>1#</sup>, Sanjana Gawali<sup>2#</sup>, Rutuja Gawade<sup>3</sup> and Sandhya Ghorpade<sup>4</sup>

*Prof. Prabodh Nimat<sup>5</sup>*

*1 UG Scholar, DIT, Pimpri, Pune, 411018, India*

*2 UG Scholar, DIT, Pimpri, Pune, 411018, India*

*3 UG Scholar, DIT, Pimpri, Pune, 411018, India*

*4 UG Scholar, DIT, Pimpri, Pune, 411018, India*

*5 Professor, DIT, Pimpri, Pune, 411018, India*

*Dr. D. Y. Patil Institute of Technology*

.....  
Corresponding Authors List#: Sanjana Gawali<sup>2#</sup>

Email: <sup>2</sup>sanjana.gawali99@gmail.com # Mobile: +91-8087247941, 2 UG Scholar, DIT, Pimpri, Pune, 411018, India

<sup>1</sup>agaleprerana@gmail.com, <sup>3</sup>rutugawade1998@gmail.com, <sup>4</sup>sandhyaghorpade0@gmail.com, <sup>5</sup>psnimat@gmail.com

.....

## ABSTRACT

The impact of information security breaching is becoming bigger and complicated day-by-day. Intrusion Detection Systems (IDS) are considered one of the basic building blocks for the protection against intrusive activities through detecting it before it hits the network systems. Artificial neural networks have been used successfully for addressing the high accuracy and precision demands of intrusion detection systems. Intrusion Detection system are security system monitoring traffic activities over information system and network for safe detecting of hostile intrusion from either outside or inside of an organization . IDS can be trained, validated and tested using CICIDS dataset. Dataset was used for validation and testing. In this paper, an intelligent intrusion detection system using Hidden Markov Model, a Machine Learning Algorithm will be built.

**Keywords:** Intrusion detection, machine learning, HMM Algorithm, CICIDS dataset

## 1. INTRODUCTION

Generally, an intrusion is an illicit action or access into system and person who tries to enter into the system and if becomes successful is called an intruder. Intrusion refers to any illegal action that affects integrity, confidentiality and availability of system. Detecting such malicious activities is called intrusion detection. An IDS is system software that monitors the malicious activities or policy violation and produce report. Intrusion detection systems are security systems monitoring traffic activities over information

systems and networks, for sake of detecting possible hostile intrusions originating from either outside or inside an organization. Intelligent intrusion detection systems can be built using four basic approaches: Clustering Techniques, Genetic algorithms Fuzzy logic and Artificial neural networks: supervised and unsupervised.

An intrusion detection system is built based on the clustered version of SOM neural network. In existing system we have use to 100,000 connected records, which consists of 80% of NSL\_KDD dataset. Dataset was used for

validation and testing. The proposed system provides a security system, named Intrusion Detection System. The performance of proposed system is measure with the help of CICIDS dataset. The proposed technique provides good detection rate and detects malicious behaviours launched toward a system.

## 2. LITERATURE REVIEW

In paper, [1]Detection of Intrusions in KDDCup Dataset using GA by Enumeration Technique author Vishal R. Chaudhary, R. S. Bichkar presents an efficient GA based methodology to produce the classification rules for Network intrusion detection system. The chromosome structure has been selected by applying enumeration technique in which the computational time required to produce the population is significantly reduced and near optimal rules are generated. These classification rules are used to find networking attacks or intrusions. The proposed system is applied on KDDCup99 Dataset to yield more efficient and effective classification rules.

The paper presented by Sulaiman Alhaidari, Ali Alharbi and Mohamed Zohdy[2] describes Distributed Denial Of Service (DDOS) attacks using Hidden Markov Models. The performance of the approach they used is better and achieves higher detection rate and has lower false positive rate compared to Naïve Bayes and Neural Network Machine Learning Algorithms.

The paper presented by Ranjit Panigrahi and Samarjeet Borah[3] describes the detailed analysis of CICIDS dataset for designing Intrusion Detection System. The contribution of the paper is it identifies and provides effective solutions to the shortcomings of CICIDS dataset, relabels CICIDS dataset to reduce high class imbalance problem.

## 3. INTRUSION DETECTION:

## 4. PROPOSED SYSTEM:

Intrusions can be defined as the set of actions that attempt to compromise the confidential harmony, integrity or availability of a computer resource, that deliberate unauthorized access of the resource, they try to make an attempt to;

1. Accessing of information
2. Manipulating of data
3. Rendering of information in a system to make unreliable or unusable.

An intrusion detection system (IDS) is a union of hardware and software components that detect harmful or malicious attempts in the network. IDS can monitor all the network activities and hence can detect the signs of intrusions. The main aim of IDS is to inform the system administrator that any doubtful activity happened. There are two kinds of Intrusion detection techniques:

### A) Anomaly Detection:

Recognize malicious activities based on deviations from the normal conduct are considered as attacks. Although it can detect unknown intrusions, the rate of missing report is low.

### B) Misuse Detection:

Recognize intrusions based on a standard pattern of the malicious activity. It can be very helpful for known attack patterns. Also, the rate of misplaced report is high. One disadvantage of Misuse Detection over Anomaly Detection is that it can only notice intrusions which contain known patterns of attack. An intrusion detection system (IDS) monitors the activities of a given environment and decides whether these activities are malicious or normal based on system integrity, confidentiality and the availability of information resources. When building IDS one needs to consider many issues, such as data collection, data pre-processing, intrusion recognition, reporting, and response.

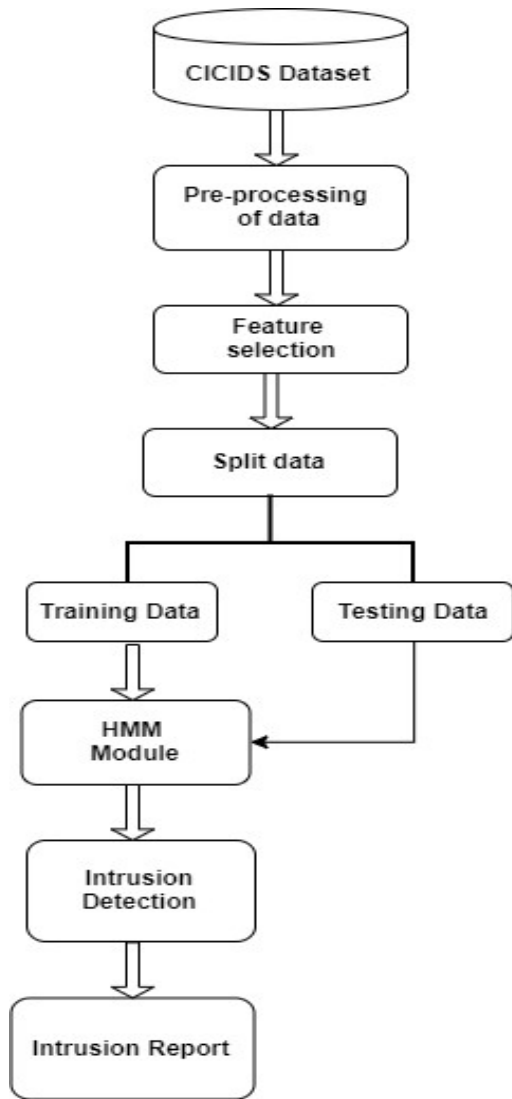


Fig1: Flowchart for intrusion detection

**4.1. Description:**

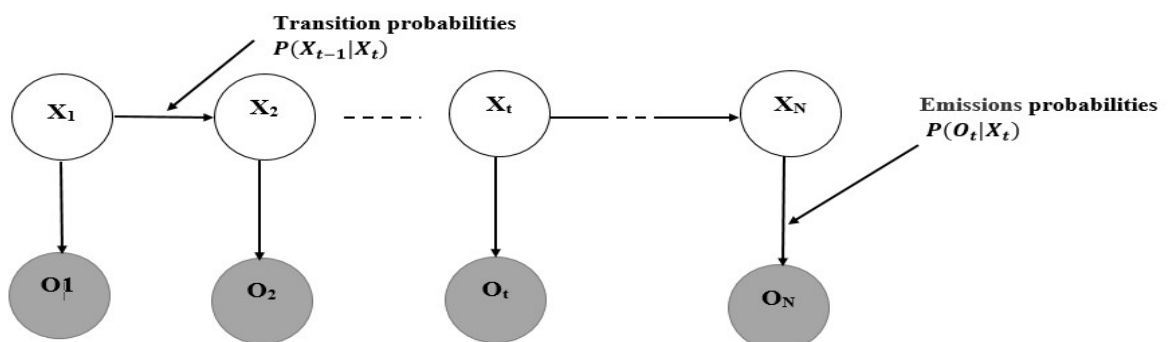
A Dataset is a collection of data. Load Dataset into the program. Data Pre-processing is an important step in machine learning process. Data pre-processing is a technique that is used to convert the raw data into a clean dataset. Data is cleaned through process such as handling

the missing value, noisy data or resolving the inconsistencies in the data.

Data transformation is the process of converting data from one format or structure into another format or structure. It primarily involves the way in which source data element will be changed for the destination. Machine learning uses features (i.e. variables or attributes) to generate predictive models. Using a suitable combination of features is essential for obtaining high precision and accuracy. Because too many (unspecific) features pose the problem of over fitting the model, we generally want to restrict the features in our models to those that are most relevant for the response variable we want to predict. Using as few features as possible will also reduce the complexity of our models, which means it needs less time and computer power to run and is easier to understand. After feature selection dataset is split into training data and test data. Apply training dataset to the HMM algorithm. Applying the HMM algorithm generate the machine learning model. Send test data to the model and first check the performance and based on that identify and predict Intrusion Detection.

**4.2. Hidden Markov Model**

HMMs have been extensively utilized in many applications such as speech recognition, finance, computer vision and bioinformatics. HMM is composed of hidden states, and observable emissions. States are the desirable events in a system, which are not visible to the observer, while emissions are the observable symbols emitted from the states. Using a sequence of emissions, an HMM can predict whether a system is in each state at a certain time. Fig shows the first-order HMM, where the observations are shaded in grey.



**Fig.** First-order HMM, where the observations are shaded in grey [2]

There are three fundamental sub-problems to HMM. The first is the evaluation problem. This addresses calculating the probability that the model can generate the indicated output sequence. The second is the decoding problem. This strives to derive the model history—that is, sequence of states—that was most likely responsible for the generation of a specified output sequence. The third is the so-called learning problem. This problem endeavours to deduce model parameters from a set of output sequences in a manner that offers the greatest fidelity, that is, likelihood of correct sequence generation.

## 5. ADVANTAGES

- Accuracy of detecting suspicious user is efficient than existing system.
- Internal Intrusion Detection, which detects malicious behaviors of users.
- Other systems consume longer time for data analysis than the IDS does.
- This can also detect malicious behaviors for systems employing GUI interfaces.

## 6. FUTURE SCOPE

- We can use such IDS system for different topology and network in future.
- We can also develop the system dynamically using wire shark which detect the intrusion live.
- Use of parallelism in population generation and in detection phase, the computational time of the system will be reduced and also impact on calculation of fitness of individual.

## 7. CONCLUSION

In this paper, IDS with HMM technique is represented to detect the different types of attacks. The performance of system is measured with the help of CICIDS dataset. The proposed technique provides good intrusion detection rate that applied and tested to detect DDoS attacks.

## 8. REFERENCES

- [1] Vishal R. Chaudhary, R. S. Bichkar, "Detection of Intrusions in KDDCup Dataset using GA by Enumeration Technique," in International Journal of Innovative Research in Computer and Communication Engineering Issues Vol. 3, Issue 3, March 2015.
- [2] Sulaiman Alhaidari, Ali Alharbi, Mohamed Zohdy , " Distributed Denial Of Service (DDOS) attacks using Hidden Markov Models," IJCSI International Journal of Computer Science Issues, Volume 15, Issue 5, September 2018.
- [3] Panigrahi, Ranjit & Borah, Samarjeet. (2018). A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems. 7. 479-482.
- [4] Shrivastava, Gulshan & Sharma, Kavita & RAI, SWARNLATA. (2010). The Detection & Defense of DoS & DDoS Attack: A Technical Overview.
- [5] Al-Dabbagh, Ahmad & Li, Yuzhe & Chen, Tongwen. (2017). An Intrusion Detection System for Cyber Attacks in Wireless Networked Control Systems. IEEE Transactions on Circuits and Systems II: Express Briefs. PP. 1-1. 10.1109/TCSII.2017.2690843.
- [6] M. Panda, A. Abraham, and M. R. Patra, "Discriminative multinomial naive bayes for network intrusion detection," in Information Assurance and Security (IAS), 2010 Sixth International Conference on, 2010, pp. 5-10.
- [7] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self-organization map (SOM) artificial neural network," Journal of Engineering Science and Technology, vol. 8, pp. 107-119, 2013.
- [8] N. F. Haq, A. R. Onik, M. A. K. Hridoy, M. Rafni, F. M. Shah, and D. M. Farid, "Application of machine learning


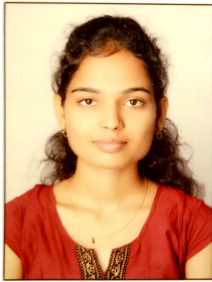


approaches in intrusion detection system: a survey," IJARAI-International Journal of Advanced Research in Artificial Intelligence, vol. 4, pp. 9-18, 2015.

[9] S. Zhong, T. M. Khoshgoftaar, and N. Seliya, "Clustering-based network intrusion detection,"

International Journal of reliability, Quality and safety Engineering, vol. 14, pp. 169-187, 2007.

[10] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," Knowledge-based systems, vol. 78, pp. 13-21, 2015.

## 9. AUTHORS DETAILS

<b>First Author</b>		<p>Name: Prerana Agale            Email: agaleprerana@gmail.com            Current Affiliation: UG Student            Current Institute: Dr. D. Y. Patil Institute of Technology            Institute Email: info.engg@dypvp.edu.in            Institute Address: Pimpri, Pune</p>
<b>Second Author</b>		<p>Name: Sanjana Gawali            Email: sanjana.gawali99@gmail.com            Current Affiliation: UG Student            Current Institute: Dr. D. Y. Patil Institute of Technology            Institute Email: info.engg@dypvp.edu.in            Institute Address: Pimpri, Pune</p>
<b>Third Author</b>		<p>Name: Rutuja Gawade            Email: rutugawade1998@gmail.com            Current Affiliation: UG Student            Current Institute: Dr. D. Y. Patil Institute of Technology            Institute Email: info.engg@dypvp.edu.in            Institute Address: Pimpri, Pune</p>
<b>Fourth Author</b>		<p>Name: Sandhya Ghorpade            Email: sandhyaghorpade0@gmail.com            Current Affiliation: UG Student            Current Institute: Dr. D. Y. Patil Institute of Technology            Institute Email: info.engg@dypvp.edu.in            Institute Address: Pimpri, Pune</p>