*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 10, Issue 3 (Mar. - Apr. 2013), PP 66-70*
*www.iosrjournals.org*

# Intrusion Detection System using Hidden Markov Model (HMM)

Megha Bandgar, Komal dhurve, Sneha Jadhav,Vicky Kayastha,Prof. T.J Parvat
*(Computer, SIT/university of pune,India)*

***Abstract :*** *Intrusion detection systems play a key role in detecting such malicious activities and enable administrators in securing network systems. Intrusion detection can detect malicious attacks that have penetrated preventative mechanisms such as firewalls, which can help provide damage assessment, response and prosecution support.*

*This paper describes a novel approach using Hidden Markov Models (HMM) to detect Internet attacks. In this paper we describe an intrusion detection system for detection of signature based attack. These attack signatures encompass specific traffic or activity that is based on known intrusive activity. We performed single and multiple HMM model for source separation both on IP and port information of source and destination. This approach reduced the false positive rate and we made this type of source separation as our basic step for building HMM.*

***Keywords -*** *Intrusion Detection,* TCP/IP Packet Analysis, *Detection, Hidden Markov Model, algorithm.*

## I. INTRODUCTION

An Intrusion detection system (IDS) is software and/or hardware designed to detect unwanted attempts at accessing, manipulating, and/or disabling computer systems, mainly through a network, such as the Interne Intrusion Detection Systems (IDS) refers to a software or a system built to detect intrusions. In general, detection mechanism used by IDS can be classified into two major categories.

1) *Signature based detection*: Signature-based detection is normally used for detecting known attacks. There are different definitions of attack signatures
2) *Anomaly based detection*: Modeled using normal traffic and deviation from this profile is considered anomalous.

Intrusion detection can be perform by implementing some important tasks on the host computer and network itself like real time traffic analysis and packet login on the IP networks

We use Hidden Markov Model (HMM), a generative model, for modeling input data. The model is proposed to profile TCP based communication channel for intrusions.HMM is used to profile source separated clean traffic and the model thus built is used to classify test traffic. Intrusion-detection systems aim at detecting attacks against computer systems and networks or in general, against information systems. IDS are yet another tool in the network administrator's computer security. Intrusion detection is the process of monitoring for and identifying attempted unauthorized system access or manipulation. A high positive rate is when the IDS says there is a security thread , but the traffic is not malicious or was never intended to be malicious .

Signature detection is the popular type of the IDS, and they work by using database of known bad behaviors. The defined patterns are known as signature. Signature-inspection engines can query any portion of the network packet or look for a specific series of databytes. Signature detection IDSs are proficient at recognizing known threads .Once a good signature is created, signature detectors are great at finding patterns.. Signature detection IDS is that it will specifically identify the threat. Intrusion detection system are yet another tool in a network administrator's computer security arsenal. Often through of as a tertiary extra after antivirus software and firewalls, an IDS is often the best way to detect a security breach.

### 1.1 Why intrusion Detection

Intrusion Detection is the process of monitoring for and identifying attempted unauthorized system access or manipulation. Most network administrators do ID all the time without realizing it.When the IDS notices a possible malicious threat, called an event.it logs the transaction and takes appropriate action. The action may simply be to continue to log, send an alert. Redirect the attack or prevent the maliciousness.IDS support the defense-in-depth security principle and can be used to detect a wide range of rogue events, including the following:

- Password cracking
- Protocol attacks
- Installation of rootkits

- Malicious code, like viruses,worms,Trojans
- Illegal data manipulation
- Unauthorized file access
- Denial of service attack

1.2TCP/IP Packet Analysis

The TCP and IP Protocols work together, hence the name TCP/IP handles routing the packets from source to destination, and TCP works to ensure reliable delivery.The protocol number for ICMP is 1.for TCP is 6, and for UDP is 17.TCP works at the transport layer of the OSI model and contains mechanisms to make sure packets arrived at the destination successfully. A TCP header contains source and destination port number, sequencing and acknowledgement number and six bit flags, among the other data field.TCP is also reliable because it will direct a host to retransmit a packet if it is not acknowledgment by the destination. Different protocol bits are used to tell each communication host whether a TCP packet is part of starting, an established, or a disconnecting session .Each flag can be on(1) or off(0).

The four most important state flags are:

- SYN Synchronization. This starts a TCP session.
- ACK Acknowledge. This Acknowledge successful receipt of a prior packet
- FIN Finish. This will gracefully end a TCP session
- RST Reset. This will forcefully and immediately and TCP session.

## II. Detection

**2.1 Anomaly Detection**

Anomaly Detection it works by establishing accepted baselines and noting exceptional differences.

Baselines can be established for a particular computer host or for a particular network segment.Some IDS vendors refer to AD systems as behavior-based since they look for deviating behaviors.If an IDS looks only at network packet headers for differences, it is called anomaly detection.The goal of AD is to be able to detect a wide range of malicious intrusions, including those for which no previous detection signature exists. By defining known good behaviors, an AD system can alert to everything else.

**2.2 Signature Detection**

Signature Detection are the most popular type of IDS, and they work by using database of known bad behaviors and patterns. This is nearly the exact opposite of AD systems.When you think of a signature detection IDS, think of it as an antivirus scanner for network traffic. Signature inspection engine can query any portion of a network packet or look for a specific series of data bytes.

## III. Signature Detection Rules

Rules are usually contains the following information as:

- Unique signature byte sequence
- Protocol to examine(TCP,UDP)
- IP port requested
- IP addresses to inspect

AD system are great at detecting a sudden high value for some metricAD IDS fail horribly is in establishing an initial baseline and in detecting malicious activity that does not violate an accepted behavioral norm, especially in the realm of malicious content.It defining the baseline norm in a chaotic changing world can be difficult.

**3.1 Advantages of Signature Detection:**

Signature detection IDS are proficient at recognizing known threats. Once a good signature is created, signature detector are great at finding patterns, and because signature detection IDS are popular, a signature to catch a new popular attack usually exists within hours of it first being reported. This applies to most open source and commercial vendors.

It will specifically identify the threat, whereas an AD engine can only point out a generally. An AD engine can only point out a generality. An AD IDS might alert you that a new TCP port opened on your file server, but signature detection IDS will tell you what exploit was used. Because a signature detection engine can better identify specific threats, it has a better chance at providing the correct countermeasure for intrusion prevention.

**3.2 Disadvantages of Signature Detection:**
Signature detection IDS are the most popular types of IDS.

- Cannot recognize unknown attacks
- Performance suffers as signature rules

## IV. Hidden Markov Model

HMM is a generative model that can model data which is sequential in nature. It is used to model data where the Assumption

*Markav property:* Consider a system with N states and at discrete time intervals, there is transition among states. Let these instances be t, t = 1, 2, 3, · · · .Any process is Markovian if the conditional probability of future states, given the present state and past states, depend only upon the present state

## V. INDENTATIONS AND EQUATIONS

Definition of a HMM:
HMM [λ] is a five tuple, i.e., _ = [N, M, A, B,$\pi$].

The parameters of the model are
N, number of states in the model, Q = {Q1, Q2, · · · , QN}.
M, number of observation symbols, V= {V1, V2, · · ·, VM}.

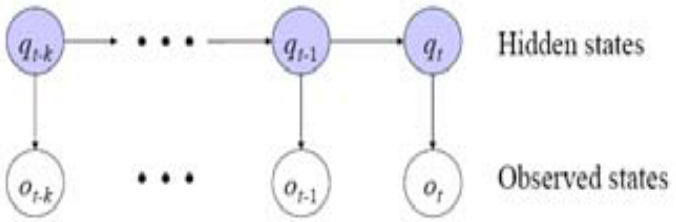


Fig.1
HMM Architecture

- Graphical Model
- Arrows indcate probabilistic dependencies
- Circles indicate states

**5.1 Algorithm**
Algorithm for hmm*:* The following algorithm used to model

1. Baum-Welch algorithm is used to learn the parameters of the model, {A, B, $\pi$}, from input data.
2. .Forward-Backward algorithm is used to learn the probability of occurrence of an observation sequence given the model, P[O| λ

## VI. FIGURES AND TABLES

### 6.1 Single Model:

In this, stream of traffic is coming towards the server. The classifier is made to profile over clean traffic, i.e., traffic stream which is devoid of any malicious traffic stream. The classifier flags any traffic that deviates from clean traffic profile as suspicious. The intuition behind this approach is that clean traffic and malicious traffic are not generated from the same distribution.

traffic is separated according to source/destination IP pair and trained with a single HMM model.All packets between a unique source/destination IP pair constitute a stream. Each stream consist of series of TCP flags that were used in the packets throughout the connection. Then a single HMM model is used to learn the characteristics of all streams to the server.

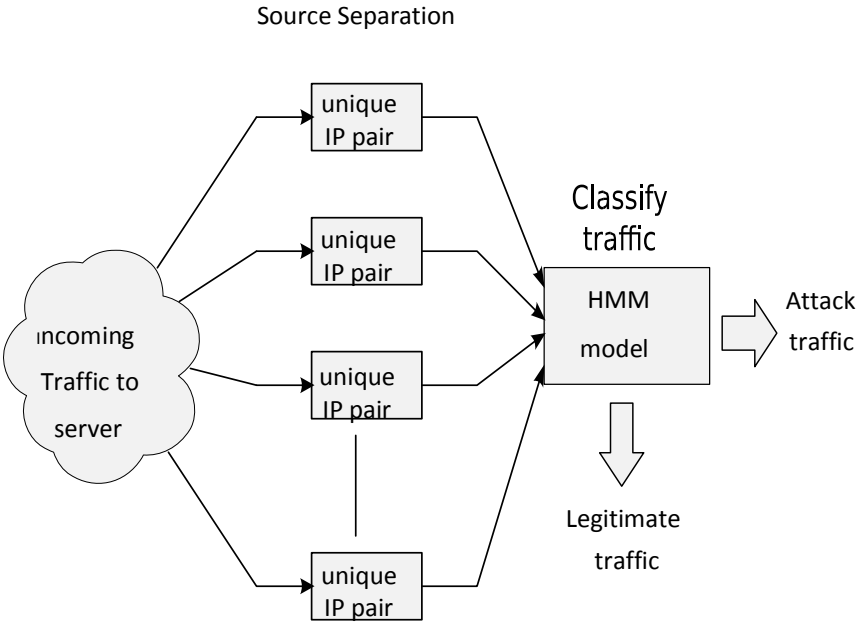


Fig.2
Single Model

## 6.2 Layered Model

To overcome from the performance of the single model, we performed source separation of the stream of traffic according to destination ports of the server and then upon source/destination IP address. For instance, different traffic streams belonging to a particular port numbers, say port 25 (SMTP), port 23(TELNET),port 20(FTP). This approach improved the results. By using multiple models, this model had higher accuracy and lower false positive rate as compared to the single HMM approach

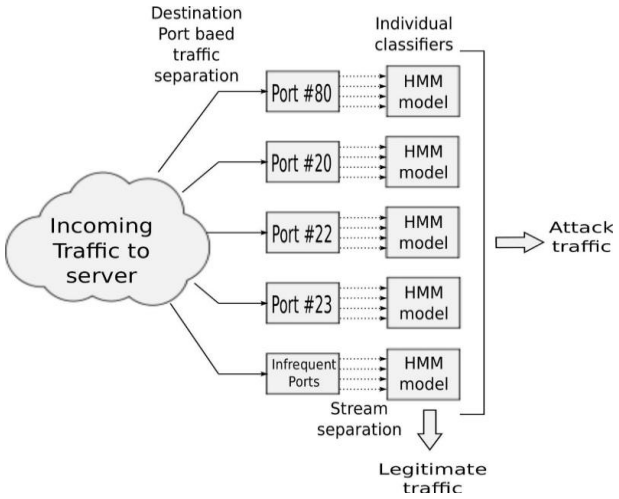


Fig. 3
Layered Model

Protocols with large amount of incoming traffic were trained separately. This approach reduced the false positive Rate. In this paper, we detect the fraggle attack and smurf attack.

## VII. Preliminary Result

HMM profiles this data and uses this information to test incoming traffic. During the testing phase, traffic that were not used for training are tested against the model learnt.

**7.1 Experiments**

The experiments that were conducted are described as follow

| Src IP | Dest IP | Src Port | Dest Port | Protocol | Capture ... | Pac Len | Conn Time | Syn#Ack#... |
|---|---|---|---|---|---|---|---|---|
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |
| 192.168.... | 129.55.1... | 3114 | 80 | TCP | 54 | 40 | 128 | false#tru... |
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |
| 192.168.... | 129.55.1... | 3114 | 80 | TCP | 54 | 40 | 128 | false#tru... |
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |
| 192.168.... | 129.55.1... | 3114 | 80 | TCP | 54 | 40 | 128 | false#tru... |
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |
| 192.168.... | 129.55.1... | 3114 | 80 | TCP | 54 | 40 | 128 | false#tru... |
| 129.55.1... | 192.168.... | 80 | 3114 | TCP | 1454 | 1440 | 49 | false#tru... |

Table
Results on Data set

## VIII. Conclusion

In this paper, we have proposed a HMM model for intrusion detection.HMM perform well well on MySQL Database .The difficulties aries when implementing HMM model for anomaly based intrusion detection. The proposed model also performed well in detecting intrusions. In this paper we detect the signature attack in to the system.

## Acknowledgements

It is pleasure for we are present this paper where guidance plays an invaluable key and provides concrete platform for completion of the paper.We also like to express our sincere thanks to internal guide Prof.Mr. T.J Parvat,.Department of computer Engineering, for his unfaltering encouragement and constant scrutiny without which we have not looked deeper into our work and realized our shortcomings and our feats. This work would not have possible without him.

## REFERENCES

### Ieee Papers:

[1 ] R Rangadurai Karthick, Vipul P. Hattiwale, Balaraman ,Ravindran,*Adaptive Network Intrusion Detection System using an Hybrid Approach*, 2012.

[2] Jiankun Hu and Xinghuo Yu, Hsiao-Hwa Chen,*A Simple and Efficient Hidden Markov Model Scheme for Host-Based Anomaly Intrusion Detection*,2009.

[3] Lawrence R. Rabiner, *A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition*, Proceedings of the IEEE, 1989.

[4] Wenke Lee and Salvatore J. Stolfo and Philip K. Chan and Eleazar Eskin and Wei Fan and Matthew Miller and Shlomo Hershkop and Junxin Zhang, *Real Time Data Mining-based Intrusion Detection*, IEEE, 2001.

[5] Ourston, Dirk and Matzner, Sara and Stump,William and Hopkins, Bryan, *Applications of Hidden Markov Models to Detecting Multi-stage Network Attacks*, HICSS, 2003

### Books

[1] The complete reference Network Security by Roger A.Grimes

### Chapters in Books:

[1] Intrusion-Detection System