Hybrid Intrusion Detection System Using Anomalous Internet Episodes Rules With Weighted Signature Generation

[1]Pawar Bhakti, [2]Prof. Kalvadekar P.N
[1] ME II year, [2] PG Co-ordinator
[1] Computer Department,
[1] Sanjivani College of Engineering,Kopargaon,Kopargaon,India

***Abstract* - To provide security to network I use existing Intrusion Detection System(IDS) for identification of known attack with low false alarm,but it is not working when unknown attacks occurs so to identify unknown attacks I use Anomaly based IDS(ADS) with high false alarm. HIDS is the combination of IDS and ADS with their advantages for identification of known as well as unknown attack.IDS used signature based model to identify known attack and ADS used anomaly based model for identification of unknown attack. HIDS used internet episode rules for identify unknown attacks. Here the packets are send from client to server, then all attributes are extracted for each packet in a network. Then I am comparing attributes of each packet with kdd dataset. If the attributes match then Known attack found. After that for anomaly detection as it behavior based I am generating some normal profile and generating their rules. If the incoming packet sequence rules do not match with normal profile then we can say anomaly is detected. Finally the signature gets created for that anomaly so that if same type of attack will come then it directely identified by signature model so efficiency gets increases Here efficiency gets increases when I create more numbers of signature.**

***Index Terms* - Intrusion Detection System, anomaly detection, signature generation, internet episodes, Traffic data mining, Network security, false alarm**

I. INTRODUCTION

As we know that intrusions and anomalies are two different kinds of abnormal traffic events in an open network environment. An intrusion takes place when an unauthorized access of a host computer system is attempted. An anomaly is observed at the network connection level. Both attack types may compromise valuable hosts, disclose sensitive data. The intrusion detection system (IDS) offers intelligent protection of networked computers which is much better than using fixed-rule firewalls. These existing IDSs are built with either signature-based or anomaly-based systems.

Signature based IDS- A signature-based IDS employs a priori knowledge of attack signatures. The signatures are manually constructed by security experts analyzing previous attacks. The collected signatures are used to match with incoming traffic to detect intrusions. These are conventional systems that detect known attacks with low false alarms. However, the signature-based IDS cannot detect Unknown attacks without any precollected signatures or lack of attack classifiers. signature matching performs well only for single-connection attacks. With the sophistication of attackers, more attacks involve multiple connections. This limits the detection range by signature matching.

Anomaly based IDS- A network anomaly is revealed if the incoming traffic pattern deviates from the normal profiles significantly. Through a data mining approach, anomaly detection discovers temporal characteristics of network traffic. This system can detect unknown attacks and handles multiconnection attacks well. However, anomaly detection may result in higher false alarms. The newly proposed HIDS is designed to solve these problems with much enhanced performance.

Hybrid based IDS- Here a new hybrid intrusion detection system (HIDS). This system combines the positive features of both intrusion detection models to achieve higher detection accuracy, lower false alarms, thus, a raised level of cyber trust. An adaptive base support threshold is applied on selected axis attributes in mining the Internet episode rules. The episode rules are used to build the HIDS, which detects not only known intrusive attacks but also anomalous connection sequences.

II. LITERATURE SURVEY

It gives the different data mining techniques using association rules to build IDS.HIDS is combination of IDS used to identify known attacks and ADS is used to identify unknown attacks.[1]

This paper gives intrusion detection must be designed to monitor the connection features at the network, transport and application layers.[2]

This paper gives SNORT and Bro are two widely used IDS that are based on the misuse model.[3]

In this paper, the HIDS architecture and prove its effectiveness through simulation experiments. The HIDS integrates the flexibility of ADS with the accuracy of a signature-based IDS. ADS is designed by mining FERs over Internet connections.[5]

III. SYSTEM OVERVIEW

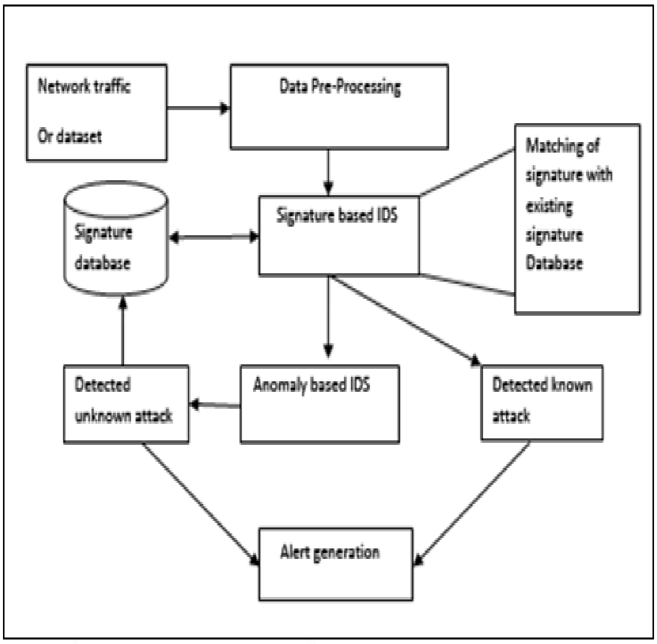

Fig .1: System Overview

In this paper I am implementing a Hybrid Intrusion Detection System used to identify both known as well as unknown attacks. This method combine IDS used to identify known attacks and ADS used to identify unknown attacks but with high false alarm. So to overcome this disadvantage HIDS used internet episode rules to identify both known and unknown attacks. After that it creates signature based on anomaly detected and stored in signature database.

It consists of following blocks-

- **Network traffic**- It is input for the HIDS. Network traffic is in the form of packets.
- **Data preprocessing-** The data information coming from multiple sources is usually incomplete, noisy and inconsistent. These raw data must be preprocessed and converted into ASCII network packet information forms or host the event data, and then build the connection records for the network connection or records for the host session data.
- **Signature database creation-** In this module signature database is created by using generating the signatures and store that signatures for detection of known attacks.

Objectives

1. The main purpose of HIDS is provide LAN security.
2. By implementing the HIDS, secure communication will be provided within the network.
3. Identification of different types of known attacks(like DDOS,U2R etc) and unknown attacks.

Scope

1. HIDS system is a scalable solution identify the known as well as unknown attacks for detection of intrusion and anomalies in network.
2. It is used to secure the network host and sensitive data.
3. This system show good performance when more signatures gets generated.

IV. BREAKDOWN STRUCTURE

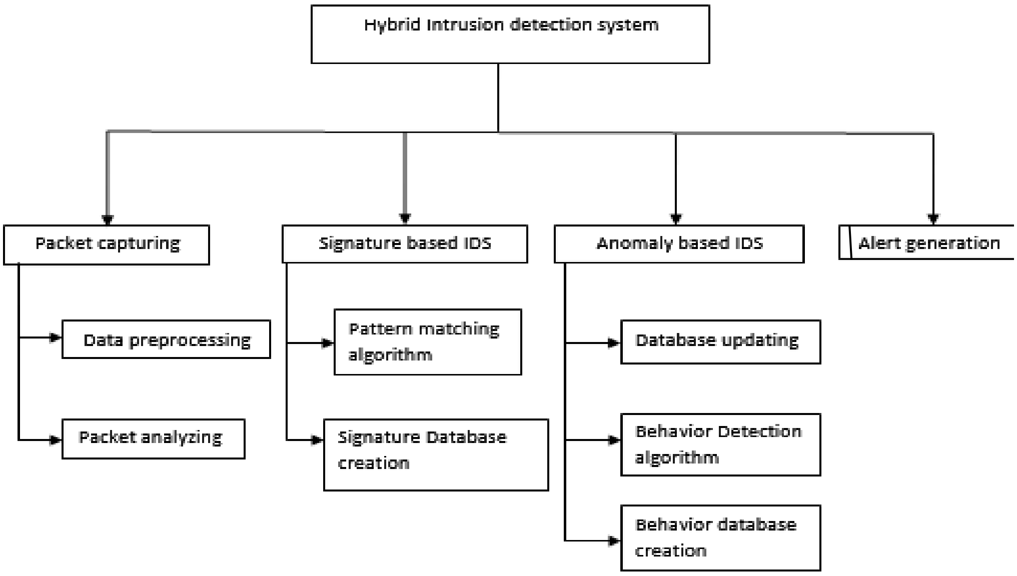
Hybrid Intrusion detection system
Packet capturing
Signature based IDS
Anomaly based IDS
Alert generation
Data preprocessing
Packet analyzing
Pattern matching algorithm
Signature Database creation
Database updating
Behavior Detection algorithm
Behavior database creation

Fig.2 Breakdown Structure of HIDS

Module1. Packet Capturing

1. In this module packet captured from client through LAN network.
2. Packet capturing is done by creating the socket between client and server.

Packet Analyzing:

It Extract attribute from incoming packet and store it in temporary buffer.
It the attributes like service,protocol_type,Flag,source_IP,destination_IP etc.

Data Pre-processing:

.Incomplete data is preprocessed**.**

Module 2: Signature based IDS:

a) Signature database creation:

KDD99 dataset: it is having many attributes/features about 41 attributes are present in KDD.

b) Pattern matching algorithm:

1. Pattern matching algorithm is mainly used for detecting known attack in signature IDS.
2. Pattern matching algorithm is performed on signature database and packet data.

Algorithm 1: Brute Force Single-Keyword Matching Algorithm:

```
1: procedure Brute Force(x,m, y, n)
Input:
x = array of m bytes representing the keyword
m = integer representing the keyword length
y = array of n bytes representing the text input
n = integer representing the text length
2: for j = 0 to n − m do. For every possible character in y
3: i= 0
4: while i< m and x[i] = y[i + j] do
5: i=i+ 1,    i = count of matching characters at and after y[j]
6: end while
7: if i>= m then
8: output j
9: end if
10: end for
11: end procedure
```

Module 3: Anomaly based IDS:
a)Behavior database creation:
FER Database:

an FER is specified by the following expression:
L1,L2,… Ln -> R1,..,Rm
e.g E2->E1,E3
Events are nothing but services like TCP,UDP, Athentication
Rule: for normal sequence E2->E1,E3
(service = authentication) →(services = smtp)(service = smtp)

Algorithm 3: Weighted Apriori Algorithm for Generating Signatures from Anomalies Detected

```
INPUT: A set of items I, a set of connections (that is Transactions) T, weight wt of connection t, and minimum weighted support min_wsup(threshold)
OUTPUT: Weighted frequent item sets X with Wsup(X)> min_wsup
W=summation t belongs to Twt;
k =1;
L1= {i|I belngs to I ^ Wsup(i)> min wsup};{ Find all weighted frequent 1 item sets }
repeat
k =k+1;
Ck=apriori_gen(Lk-1); { Generate candidate item sets }
for each connection t belongs to T, do
Ct=subset(Ck, t); { Candidates contained in t }
for each candidate item set c belongs to Ct, do
c:weight+=wt; { Add connection weight }
end for
```

Module 4: Alert Generation:

1.Alert is indication for detection of attack.
2.Alert is generated, when known or unknown attack found.
3.Attack message display on system if attack found.
4.Alert is in the form of text message.

V. DATASET

KDD Dataset - using KDD99 dataset for comparing incoming data signature with KDD99 to identify known attacks.
Signature Database - It stores number of signatures of detected anomalies.
Frequent Episode Rules Database -It stores all the rules generated in ADS.

VI. RESULT ANALYSIS

Following tables shows the efficiency of Signature based IDS,Anomaly based ADS and HIDS.

Table 1 Average Efficiency for Signature based IDS

Day	Number of packets	Number of packets contain attacks	Number of packets detects attacks	True Positive	False Positive
Day1	29	29	24	83%	17%
Day2	29	29	21	72.41%	27.59%
Day3	29	29	23	79.31%	20.69%
Day4	24	24	15	62.5%	37.5%
Day5	35	35	30	85.71%	14.29%
Average efficiency for Signature based IDS :76.2%					

True positive=No.of attacks detected/Total no. of packets
for day1=24/29
=83%
False Positive=100-true positive
=100-83
=17%

Table 2 Average Efficiency for Anomaly based ADS

Day	Rules generated for Normal Profile	Number of packet_ sequence send	Anomaly Detected	True positive	False Positive
Day1	5	10	1	10%	90%
Day2	5	10	2	20%	80%
Day3	5	10	1	10%	90%
Day4	5	10	2	20%	80%
Day5	5	10	1	10%	90%
Average Efficiency for Anomaly based ADS:14%					

True positive=No.of anomalies detected/Total no. of packets sequence send
for day1=1/10
=10%
False Positive=100-true positive
=100-10
=90%

Table 3 Average Efficiency for HIDS

Day	Anomaly Detected	Signature for detected anomaly created	Signature Match with KDD	Efficiency
Day1	1	1	1	100%
Day2	2	2	1	50%
Day3	1	1	0	0%
Day4	2	2	1	50%
Day5	1	1	1	100%
Average Efficiency of Signature generated by detected anomalies:60.10%				

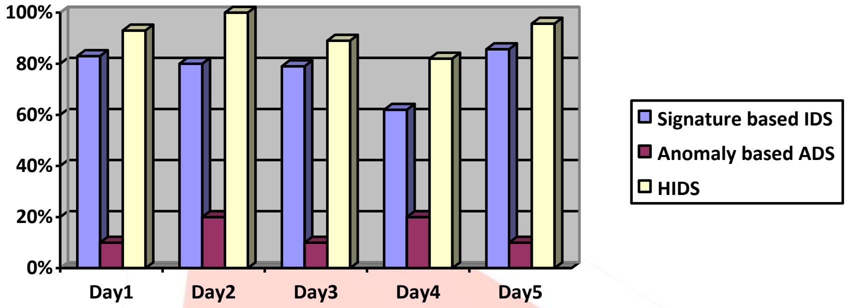

Fig. Chart5:Efficiency of signature based IDS ,anomaly based ADS and HIDS for 5 days

Table 4 Average Efficiency for HIDS

Day	Anomaly Detected	Signature for detected anomaly created	Signature Match with KDD	Efficiency
Day1	1	1	1	100%
Day2	2	2	1	50%
Day3	1	1	0	0%
Day4	2	2	1	50%
Day5	1	1	1	100%
Average Efficiency of Signature generated by detected anomalies:60.10%				

Efficiency=Signature match with KDD/signature for detected anomalies
(for day2)=1/2
Efficiency=50%

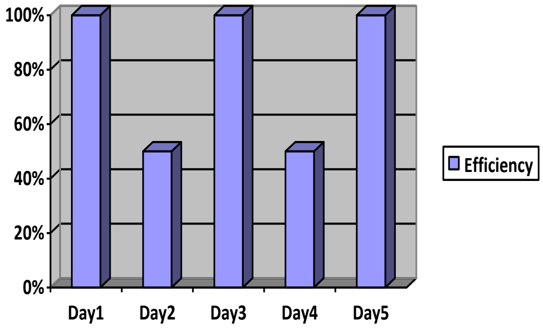

Fig. Chart 2:Efficiency for signature generated in detected anomalies for 5 days

Table 5 Average false rate for signature based IDS

Day	Number of packet send	Number of attacks detected	False Alarm	False Alarm Rate
Day1	29	14	1	7%
Day2	29	26	2	8%
Day3	29	29	1	4%
Day4	24	24	1	6%
Day5	35	35	2	6%
Average False Rate for Signature based IDS:6.2%				

False alarm rate=false alarm/No.of attack detected
=1/14
=7%

Table 6 Average False Alarm Rate for anomaly based ADS

Day	Number of file packet_ sequence send	No of Anomaly Detected	Wrong Anomaly detected	False Alarm Rate
Day1	10	1	0	0%
Day2	10	2	1	10%
Day3	10	1	0	0%
Day4	10	1	0	0%
Day5	10	2	1	10%
Average False Alarm Rate for anomaly based ADS:4%				

False alarm rate for day2=No of wrong anomaly detected/Total no. of packet send
=1/10
=10%

Table 7 Average False Alarm Rate for HIDS

Day	False alarm rate for Signature based IDS	False alarm rate for Anomaly based ADS	False Alarm rate for HIDS
Day1	7%	0%	7%
Day2	8%	10%	18%
Day3	4%	0%	4%
Day4	6%	0%	6%
Day5	6%	10%	16%
Average False Alarm Rate for HIDS:58.2%			

False Alarm rate of HIDS=False Alarm rate for IDS+False Alarm rate for ADS
(for DAY3) =4+0
=4%

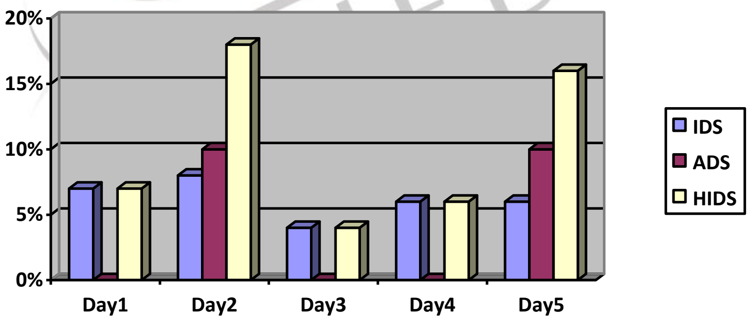

Fig. Chart5:False alarm rate for Day1,Day2,Day3,Day4 and Day5 of IDS,ADS and HIDS

Table 8:Number of Different categories of attacks found in 5days

Days	No. of different categories attacks found			
	Normal	DDOS	U2R	Other
Day1	6	10	1	12
Day2	3	3	1	14
Day3	6	11	0	13
Day4	10	6	0	7
Day5	4	7	0	24

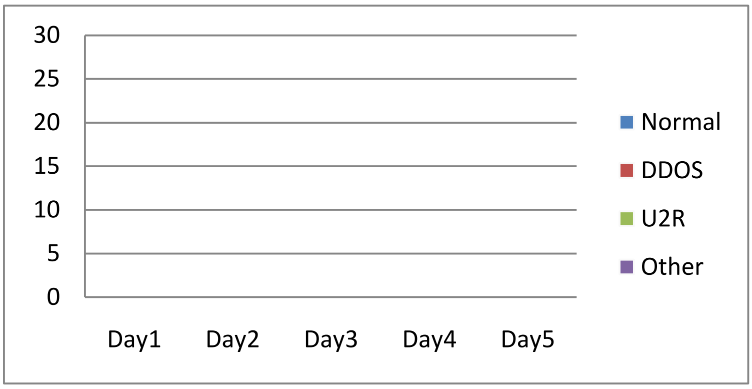

Fig. Chart 4 No. of Normal,DDOS,U2R and Other attacks found in 5 days.

VII. COMPARISON OF EXISTING SYSTEM WITH IMPLEMENTED SYSTEM

Following two table shows the comparison between existing system and implemented system. The detection rate of HIDS in implemented system is greater than existing system, but false alarm rate of HIDS in implemented system is higher than existing system.

Table 9 Detection rate and False alarm rate for Existing System

For Existing System	IDS	ADS	HIDS
Detection Rate	23%	25%	48%
False Alarm Rate	0.3%	3.2%	3.4%

Table 10 Detection rate and False alarm rate for Implemented System

For Implemented System	IDS	ADS	HIDS
Detection Rate	76.2%	14%	90.2%
False Alarm Rate	6.2%	4%	10.2%

VIII. CONCLUSION

In this system I have implemented HIDS which is used to identify known as well as unknown attacks. It uses algorithms as base support traffic data mining algorithm and Apriori algorithm. The detection rate of HIDS is better than Signature based IDS and Anomaly based ADS, here DDOS attacks found in large number. The False alarm rate for Signature based anomaly is less than ADS and HIDS. The efficiency of signature increases when we increase signature database.

IX. ACKNOWLEDGMENT

First and foremost, I would like to give thanks to God Almighty for His mercy and beautiful blessings in my life. Even in the hardest times, He has always been there to guide and comfort me and to make sure that I never stray too far from His narrow way.I would like to thank my family for their unconditional love and support. I would especially thanks my father and my grandmother. I also wish to thank my mother and my brother for being there for me, despite all the troubles and problems that grow in life.

I express my sincere gratitude to **Prof. D. B. Kshirsagar**, Head of Department (Computer), and **Prof. P.N. Kalavadekar**, (Guide/ PG Co-ordinator) SRESCOE for his unending support and encouragement during the years I have studied under his tutelage.

I am also grateful to my guide for this project, for his guidance and useful insight. This project would never have been possible without his instruction and teaching.

My sincere thanks go to all staff for their help and understanding.

X. REFERENCES

[1] D. Barbara, J. Couto, S. Jajodia, L. Popyack, and N.Wu, “ADAM:Detecting Intrusions by Data Mining”, Proc. IEEE Workshop Information Assurance and Security, 2001.

[2] S. Floyd and V. Paxson, “Dificulties in Simulating the Internet, IEEE/ACM Trans. Networking”, vol. 9, no. 4, pp. 392-403, Aug. 2001.

[3] B. Casewell and J. Beale, SNORT 2.1, “Intrusion Detection, second ed. Syngress”, May 2004.

[4] H. Mannila and H. Toivonen, “Discovering Generalized Episodes Using Minimal Occurrences”, Proc. Second Intl Conf. Knowledge Discovery and Data Mining, Aug. 1996.

[5] W. Lee, S.J. Stolfo, and K. Mok, “Adaptive Intrusion Detection: A Data Mining Approach, Articial Intelligence” Rev., vol. 14, no. 6,pp. 533-567, Kluwer Academic Publishers, Dec. 2000.

[6] W. Fan, M. Miller, S. Stolfo, W. Lee, and P. Chan, “Using Artificial Anomalies to Detect Unknown and Known Network Intrusions”, Proc. First IEEE Intl Conf. Data Mining, Nov. 2001.