Implementation of Secured Network Based Intrusion Detection System Using SVM Algorithm

S.M.Poonkuzhali[1] , M.Santhana Joyce[2], R.Jayashree[3], S.Ramya[4], K.Shalini[5]

[1,2]*Assistant Professor, Dept. of Computer Science and Engineering, Panimalar Institute of Technology, Tamil Nadu, India.*

[3,4,5] *Student, Dept. Of Computer Science and Engineering, Panimalar Institute of Technology, Tamil Nadu, India.*

---***---

Abstract - *This paper encompasses an outline that hardens the distinctive executions and proposals of Intrusion Detection. NB is one of the classification methods applied in intrusion detection system which is an effective probabilistic classifier employing the Bayes' theorem with naive feature independence assumptions. Each data transmitted is captured and analyzed for malicious content, reveal the layers in which the impact of the malicious data is visible. Furthermore, as a preventive measure the data transmission from corrupt host is blocked. NB classifier is that it only requires a small amount of training data to estimate the parameters of a classification model. Ability to incorporate flow correlation information in to the classification process. IDNB (Intrusion Detection using Naive Bayes) demonstrates higher malicious behavior detection rates in certain circumstances while does not greatly affect the network performances.*

Key Words: ***IDS- Intrusion Detection System ; NIDS - Network intrusion detection system ;*** **TCP-** ***Transmission control Protocol ; UDP-User Datagram Protocol.***

1.INTRODUCTION

IDS is a defensive mechanism whose primary purpose is to keep work going on considering all possible attacks on a system[1].An IDS system is a defense system, which detects hostile activities or exploits in a network[3].Focusing on network based, as it is more cost effective and only a small number of IdSs could monitor a larger coverage of network, it further could be based on anomaly or signature based.Signature and anomaly-based systems are similar in terms of conceptual operation and composition. The main differences between these methodologies are inherent in the concepts of "attack" and "anomaly". An attack can be defined as "a sequence of operations that puts the security of a system at risk". An anomaly is just "an event that is suspicious from the perspective of security" [7].Anomaly detection has two phases: learning phase and detection phase. In the learning phase, we construct a profile or a model of the normal system behaviour. While in the detection phase, we compare the actual system behaviour with ones in the normal system[10].There are some IDS devices that detect attacks based on comparing traffic patterns against a baseline and then looks for anomalies[2].For detecting the cyber attacks, intrusion detection is one of the popular technique. So, signatures based NIDS are somehow unable to detect the unknown attacks[8].The motivation for using the hybrid approach is to improve the accuracy of the intrusion detection system when compared to using individual approaches. The hybrid approach combines the best results from the different individual systems resulting in more accuracy[6].Several machine-learning paradigms including neural networks (Mukkamala et al., 2003), linear genetic programming (LGP) (Mukkamala et al., 2004a), support vector machines (SVM), Bayesian networks, multivariate adaptive regression splines (MARS) (Mukkamala et al., 2004b) fuzzy inference systems (FISs) (Shah et al., 2004), etc[4].Decision tree techniques are put into action in the field of intrusion detection. This piece of writing attempt to apply C4.5 decision tree with pruning technique into intrusion detection, at the same time this method is prove to give good results[9]. To overcome the limitations of intrusion detection, a broader perspective is introduced, saying that in addition to detecting attacks, countermeasures to these successful attacks should be planned and deployed in advance[5].

2. RELATED WORK

Some research and project works have been presented in the literature for the proposed system. A comprehension of some research and project work is presented here. B.Ben Sujitha , R.Roja Ramani and Parameswari[1] modelled knowledge base as a fuzzy rule such as "if-then" and improved by a genetic algorithm. The method is tested on the benchmark KDD'99 intrusion dataset but the computation time is greater. DikshantGupta, SuhaniSinghal, Shamita Malik and Archana Singh[2] have different data mining algorithms. A comparative analysis of these techniques to detect intrusions has also been made. Ozgur Depren, Murat Topallar, Emin Anarim and M. Kemal Ciliz[3] arrived on rulebased Decision Support System (DSS) is also developed for interpreting the results of both anomaly and misuse detection modules. Temporal relations between successive connections were used for intrusion detection. Sandhya Peddabachigaria , Ajith Abrahamb,, Crina Grosanc and Johnson Thomas[4] combined the individual base classifiers and other hybrid machine learning paradigms to maximize detection accuracy and minimize computational complexity resulting in a hybrid intrusion detection model. ULF T. MATTSSON[5] proposed a solution that offers the ability to detect misuse and subversion through the direct monitoring of database operations inside the database host, providing an important complement to hostbased and network-based surveillance. Suites of the proposed solution deployed throughout a network, and their alarms man-aged, correlated, and acted on by remote or local subscribing security services, thus helping to address issues of decentralized management. While the overall complexity of the security program has dramatically increased, enterprises can still implement effective security solutions by integrating sound external protection and internal security controls with appropriate security audit procedures. Vahid Golmah[6] gave a hybrid method of C5.0 and SVM and investigate and evaluate the performance with DARPA dataset. The hybrid C5.0–SVM approach gave the best performance for probe, U2R and R2L attacks. It gives 100% accuracy for probe, U2R and R2L attacks. P. Garcı´a-Teodoroa, J. Dı´az-Verdejoa , G. Macia´-Ferna´ndeza and E. Va´zquezb[7] embraces review of the most well-known anomaly-based intrusion detection techniques. outline the main challenges to be dealt with for the wide scale deployment of anomaly-based intrusion detectors, with special emphasis on assessment issues. Starting point for addressing R&D in the field of IDS. B M Mehtre[8] utilized Decision Tree (J48) algorithm to classify the network packet that can be used for NIDS and 134665 network instances for training and testing, but used 10 fold cross validation. Neha G. Relan Prof. Dharmaraj R. Patil[9] proposed two techniques, C4.5 Decision tree algorithm and C4.5 Decision tree with Pruning, using feature selection. In C4.5 Decision tree with pruning considering only discrete value attributes for classification. Exploits KDDCup'99 and NSL_KDD dataset to train and test the classifier. Whereas a random sampling of training and testing partition may generate diverse results in different runs. ELKHADIR Zyad[10] enhanced PCA by L1-norm maximization, which is more robust to outliers, instead of the Euclidean norm in the classical PCA. Extensive experiments on the well-known KDDcup99 dataset are exploited for testing the effectiveness of the proposed approach. Obtained results confirm the superiority of L1-norm PCA over the traditional PCA in terms of network attacks detection and false alarms reduction. However, this approach consumes more CPU time compared to classical PCA due to the greediness nature of the algorithm.

3. IMPLEMENTATION

3.1 Data Collection

Streaming massive statistics is an analytic computing platform this is targeted on velocity. That is because those programs require a non-stop stream of frequently unstructured data to be processed. Therefore, facts is continuously analyzed and transformed in reminiscence before it is saved on a disk. Massive information is an all-encompassing term for any collection of statistics units so large and complex that it turns into tough to technique using conventional records processing programs. Data streams values may be numbers, such as real numbers or integers, for example representing a person's height in centimeters, but may also be nominal data (i.e., not consisting of numerical values), for example representing a person's ethnicity. More generally, values may be of any of the kinds described as a level of measurement. For each variable, the values will normally all be of the same kind. However, there may also be "missing values", which need to be indicated in some way.

3.2 Pre Processing

On this module we're going to get hold of the community packet and extract attributes the use of the WinPcap and JPCap. In information generation, a packet is a formatted unit of information carried by using a packet mode laptop network. computer communications hyperlinks that do not help packets, together with conventional point-to-point telecommunications links, absolutely transmit information as a series of bytes, characters, or bits alone. when information is formatted into packets, the bit price of the communication medium can higher be shared amongst customers than if the community have been circuit switched. by means of the use of packet switched networking it's also more difficult to assure a lowest possible bitrate.

A packet includes two varieties of information: control statistics and consumer facts (also referred to as payload). The control information gives statistics the community needs to deliver the consumer information, for instance: source and vacation spot addresses, error detection codes like checksums, and sequencing facts. typically, control facts is found in packet headers and trailers, with consumer statistics in between.Distinct communications protocols use distinctive conventions for distinguishing among the factors and for formatting the data. In Binary Synchronous Transmission, the packet is formatted in eight-bit bytes, and special characters are used to delimit the specific elements. other protocols, like Ethernet, establish the start of the header and facts factors via their place relative to the start of the packet. a few protocols layout the information at a chunk level in preference to a byte level.A good analogy is to recollect a packet to be like a letter: the header is just like the envelope, and the records location is regardless of the character places inside the envelope. A difference, however, is that some networks can ruin a bigger packet into smaller packets whilst vital (observe that those smaller facts factors are nonetheless formatted as packets).A network design can reap fundamental consequences through using packets: errors detection and more than one hosts addressing. A nominal traffic profile consists of single and joint distributions of various packet attributes that are considered unique for a site. Candidate packet attributes from

IP headers are:

1. packet size,
2. Time-to-Live (TTL) values,
3. protocol-type values, and
4. source IP prefixes.

Those from TCP headers are:

5. TCP flag patterns and
6. server port numbers, i.e., the smaller of the source port number and the destination port number.

Server port number is more stable than sort/destination port numbers because most of the well-known port numbers are small numbers (e.g., below 1,024) and a large portion of Internet traffic uses the well-known port numbers. To increase the number of attributes, we can employ joint distributions of the fraction of packets having various combinations, such as:

7. <packet-size and protocol-type>,
8. <server port number and protocol-type>, and
9. <source IP prefix, TCP flags and packet size>, etc.

Joint distributions often better represent the uniqueness of the traffic distribution for a site, and are harder to guess for the attackers. As many different combinations of single attributes as needed may be used while the storage space permits.

WinPcap is an open source library for packet capture and network analysis for the Win32 platforms.Most networking applications access the network through widely used operating system primitives such as sockets. It is easy to access data on the network with this approach since the operating system copes with the low level details (protocol handling, packet reassembly, etc.) and provides a familiar interface that is similar to the one used to read and write files. The purpose of WinPcap is to give this kind of access to Win32 applications; it provides facilities to:

- Capture raw packets.
- Transmit raw packets to the network.
- Gather statistical information on the network traffic.

Jpcap is a Java class package that allows Java applications to capture and/or send packets to the network.Jpcap is based on libpcap / winpcap and Raw Socket API. Therefore, Jpcap is supposed to work on any OS on which libpcap / winpcap has been implemented. Currently, Jpcap has been tested on FreeBSD 3.x, Linux RedHat 6.1, Fedora Core 4, Solaris, and Microsoft Windows 2000/XP.

3.3 Feature extraction

The gadget captures IP packets crossing a goal network and constructs site visitor's flows by checking the headers of IP packets it's far waft-level site visitor's class. A drift includes successive. IP packets with the identical five-tuple: source IP, source port, vacation spot IP, destination port, and shipping layer protocol. It makes use of heuristic manner to determine the correlated flows and version them. If the flows located in a positive time period share the equal vacation spot IP, destination port, and transport layer protocol, they may be determined as correlated flows and shape a network drift. For the category cause, a fixed of float statistical features are extracted and discredited to symbolize network flows.

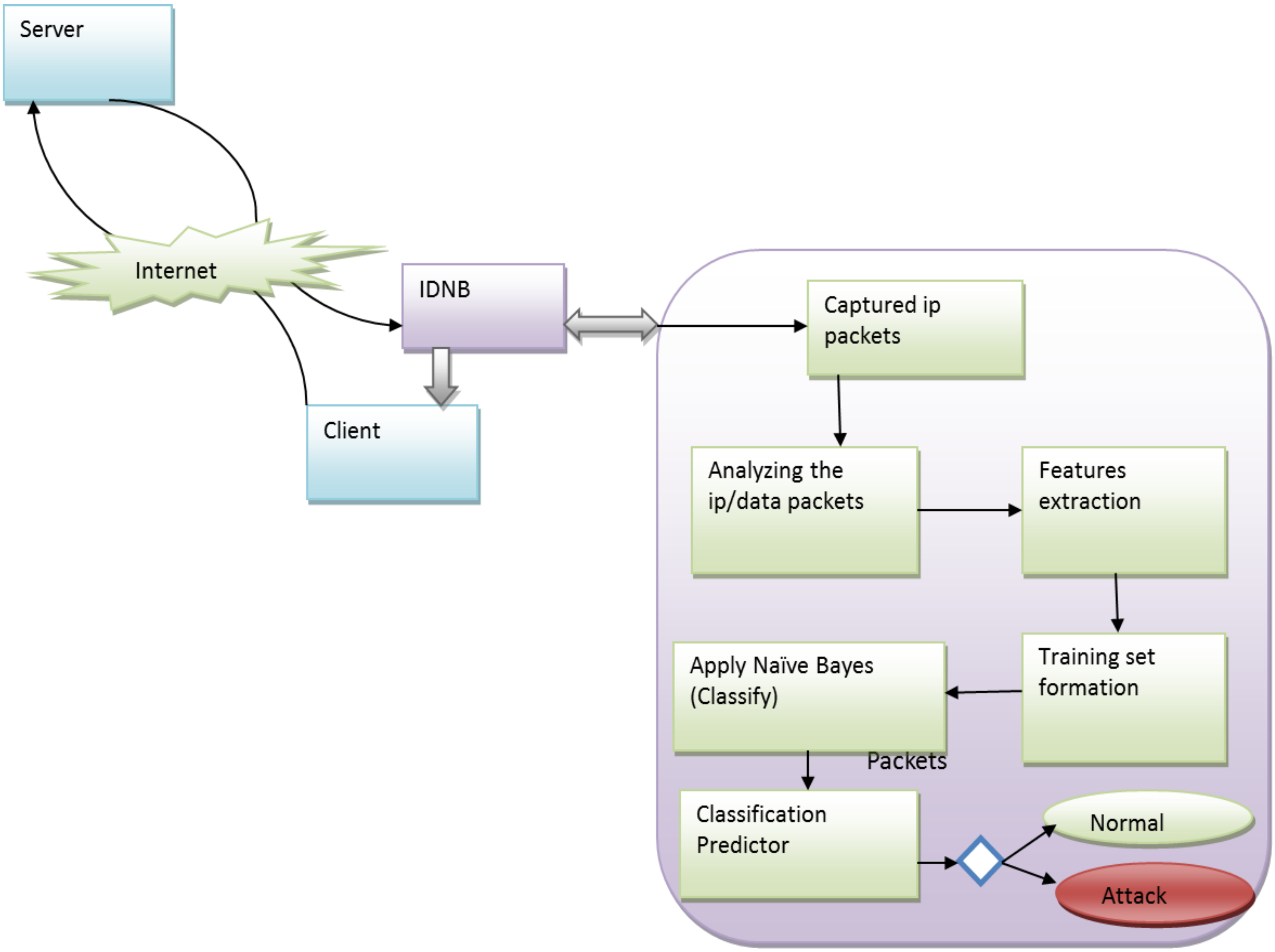


Fig -1: Architecture Diagram

3.4 Classification

Binary classifiers are generated for each class of occasion the use of applicable functions for the magnificence and class set of rules. All of the samples inside the list belong to the identical elegance. Whilst this takes place, it surely creates a leaf node for the selection tree saying to pick out that elegance. None of the features provide any statistics gain. In this situation, C4.five creates a selection node better up the tree the usage of the anticipated value of the elegance. Instance of formerly-unseen elegance encountered. Once more, C4.five creates a choice node better up the tree the use of the anticipated price. Moreover, applying the information advantage or advantage ratio will return all of the capabilities that incorporate more information for keeping apart the contemporary magnificence from all other training. The output of this ensemble of binary classifiers may be decided the use of arbitration characteristic based on the self belief stage of the output of character binary classifiers. C4.five builds choice bushes from a fixed of training facts in the equal manner as ID3, using the concept of facts entropy. The training facts is a set S = s1,s2,... of already labeled samples. every sample si = x1,x2,... is a vector in which x1,x2,... constitute attributes or functions of the pattern. The training statistics is augmented with a vector C = c1,c2,... where c1,c2,... constitute the elegance to which every sample belongs.

At every node of the tree, C4.5 chooses one attribute of the facts that most successfully splits its set of samples into subsets enriched in one magnificence or the alternative. Its criterion is the normalized information gain (distinction in entropy) that outcomes from deciding on an characteristic for splitting the statistics. The characteristic with the very best normalized facts advantage is selected to make the choice. The C4.5 algorithm then recourse at the smaller sub lists.

3.5 Efficiency Calculation – Weka tool

The impact of mixing unique classifiers may be defined with the theory of bias-variance decomposition. One method of deriving a single prediction (for new observations) is to use all trees observed inside the exclusive samples, and to use some easy vote casting: The very last category is the only most usually predicted by the exclusive timber. note that some weighted aggregate of predictions (weighted vote, weighted common) is also feasible, and typically used. In the flow diagram Fig:2 Illustrates the technical details that a data packet over the LAN faces as it's capture, naive bayesian classification etc,.

The concept of bagging (voting for classification, averaging for regression-type problems with continuous dependent variables of interest) applies to the area of predictive data mining, to combine the predicted classifications (prediction) from multiple models, or from the same type of model for different learning data. It is also used to address the inherent instability of results when applying complex models to relatively small data sets. Suppose your data mining task is to build a model for predictive classification, and the dataset from which to train the model (learning data set, which contains observed classifications) is relatively small. It could repeatedly sub-sample (with replacement) from the dataset, and apply, for example, a tree classifier to the successive samples.

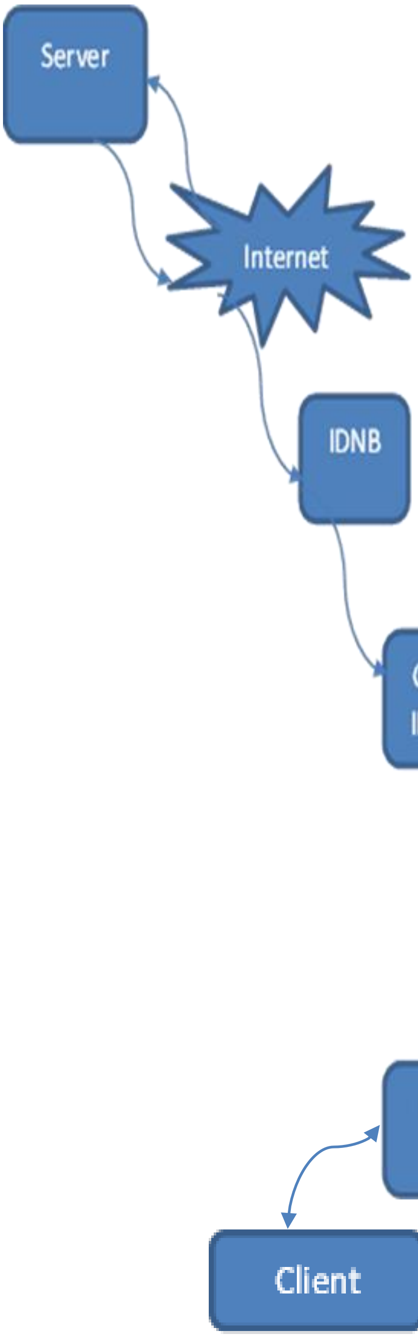
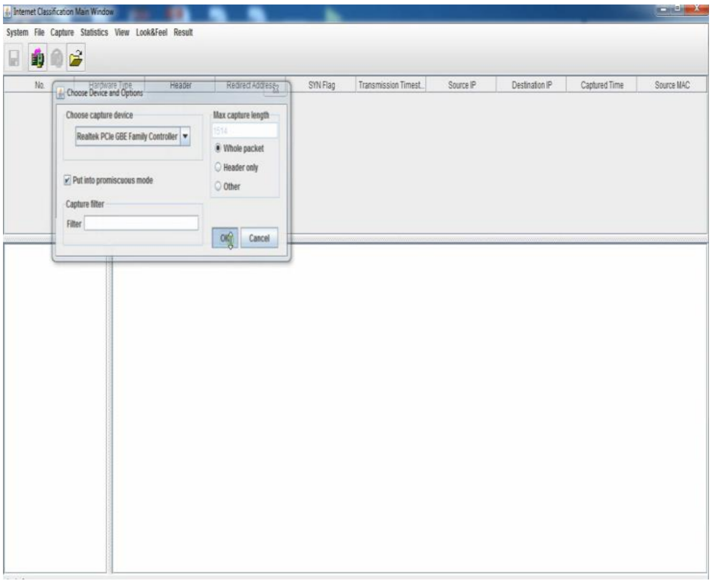


Fig -2: Flowchart

4. EXPERIMENTAL RESULTS

4.1 Initiation


Internet Classification Main Window
System File Capture Statistics View Look&Feel Result
No. Hardware Type Header Redirect Address SYN Flag Transmission Timest... Source IP Destination IP Captured Time Source MAC
Choose Device and Options
OK
Cancel

Choose capture device: Realtek PCIe GBE Family Controller

☑ Put into promiscuous mode

Max capture length: 1514

● Whole packet
○ Header only
○ Other

Capture filter
Filter:

Fig -3: Start to capture packets over the LAN

4.2 Layer chooser

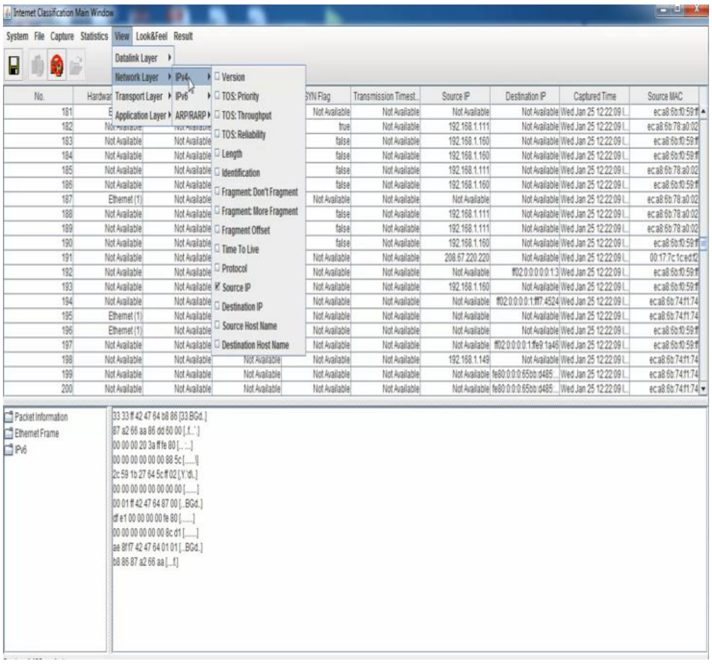
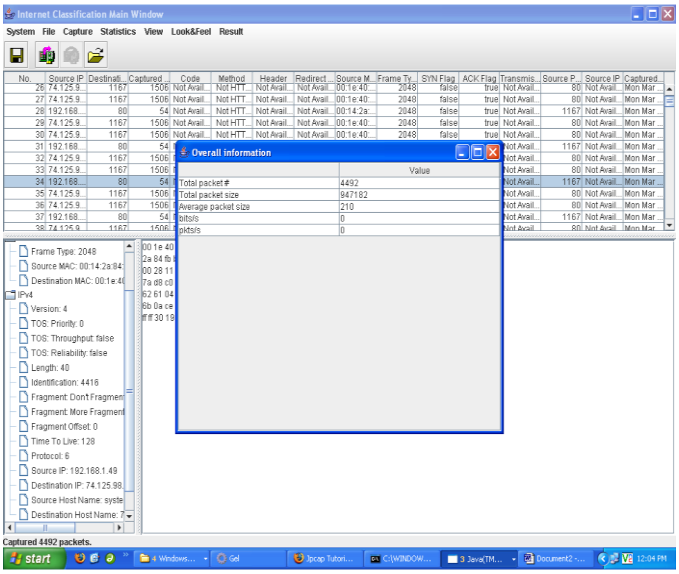


Fig -4: Displays all the OSI layers and their respective protocols

4.3 Overall information, packet details

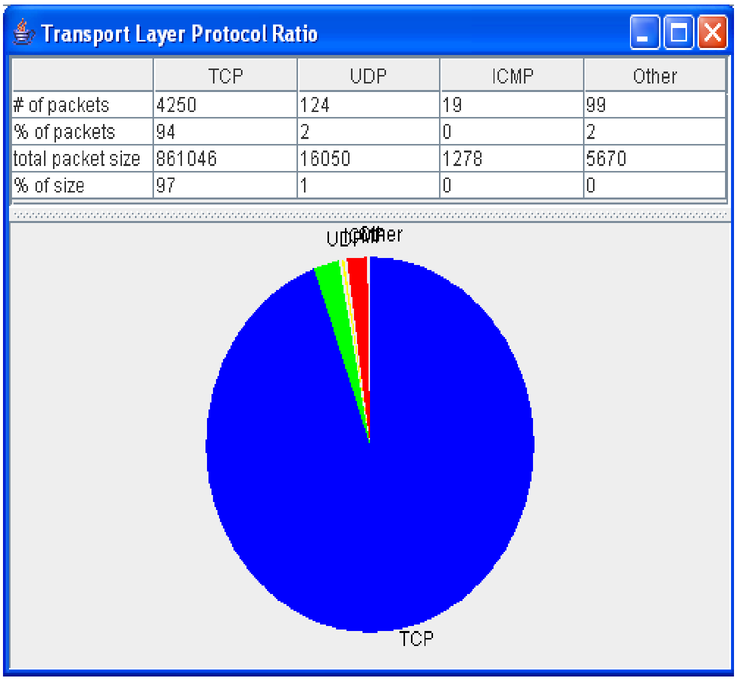


Overall information

	Value
Total packet #	4492
Total packet size	947182
Average packet size	210
bits/s	0
pkts/s	0

Fig -5:Displays the total packet size, number of packets in transmission in the LAN.

4.4 Transport Layer Protocol Ratio

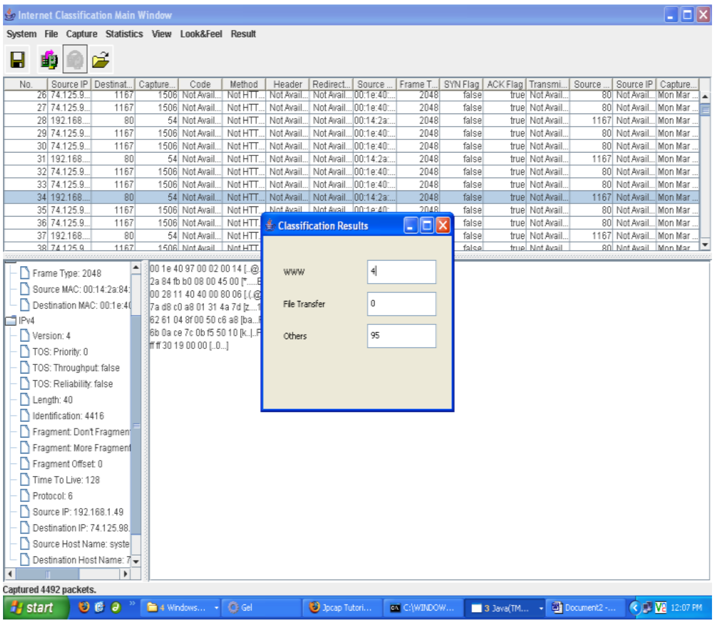


Transport Layer Protocol Ratio

	TCP	UDP	ICMP	Other
# of packets	4250	124	19	99
% of packets	94	2	0	2
total packet size	861046	16050	1278	5670
% of size	97	1	0	0

Fig -6:Displays the ratio of the packets in accordance with the transport layer protocol

4.5 Classifiaction Results


Internet Classification Main Window
System File Capture Statistics View Look&Feel Result
No. Source IP Destinat... Capture Code Method Header Redirect... Source Frame T... SYN Flag ACK Flag Transmi... Source Source IP Capture
start

Classification Results

WWW: 4

File Transfer: 0

Others: 95

Frame Type: 2048
Source MAC: 00:14:2a:84:
Destination MAC: 00:1e:40
IPv4
Version: 4
TOS: Priority: 0
TOS: Throughput: false
TOS: Reliability: false
Length: 40
Identification: 4416
Fragment: Don't Fragmen
Fragment: More Fragment
Fragment Offset: 0
Time To Live: 128
Protocol: 6
Source IP: 192.168.1.49
Destination IP: 74.125.98.
Source Host Name: syste
Destination Host Name: 7

```
00 1e 40 97 00 02 00 14 [..@.
2a 84 fb b0 08 00 45 00 [*.....E
00 28 11 40 40 00 80 06 [.(.@
7a d8 c0 a8 01 31 4a 7d [z....1
62 61 04 8f 00 50 c6 a8 [ba...F
6b 0a ce 7c 0b f5 50 10 [k..|..F
ff ff 30 19 00 00 [..0...]
```

Captured 4492 packets.

Fig -7:Outputs the number of packets over the LAN in WWW, File Transfer and others.

4.6 Analysis of network layer

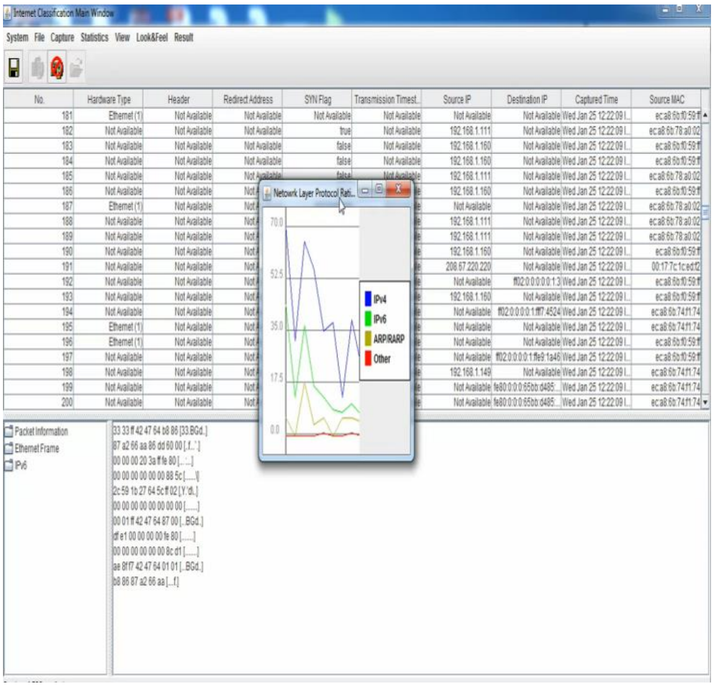

Internet Classification Main Window
System File Capture Statistics View Look&Feel Result
No. Hardware Type Header Redirect Address SYN Flag Transmission Timest... Source IP Destination IP Captured Time Source MAC
Network Layer Protocol Rati...
70.0
52.5
35.0
17.5
0.0
IPv4
IPv6
ARP/RARP
Other
Packet Information
Ethernet Frame
IPv6

Fig-8: The ratio of the protocols being used over the LAN such as IPv4,IPv6

4.7 Analysis of Transport layer

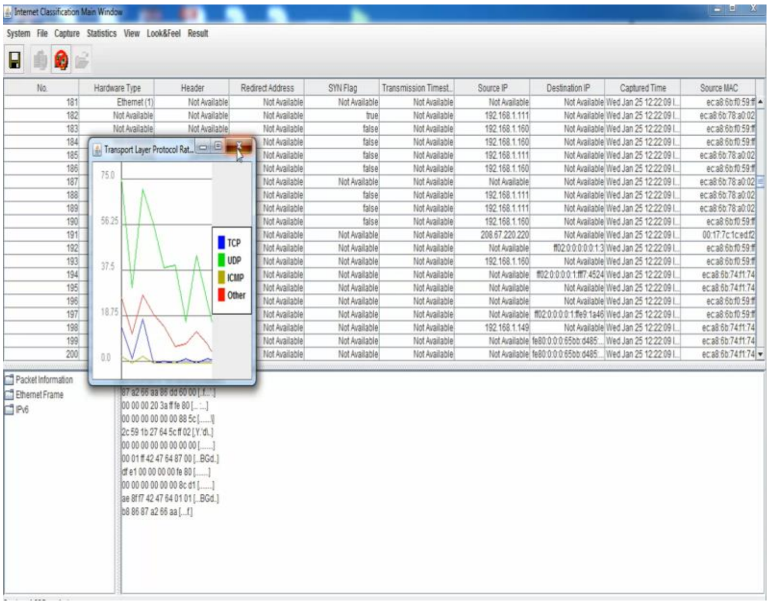

Internet Classification Main Window
System File Capture Statistics View Look&Feel Result
No. Hardware Type Header Redirect Address SYN Flag Transmission Timest... Source IP Destination IP Captured Time Source MAC
Transport Layer Protocol Rat...
75.0
56.25
37.5
18.75
0.0
TCP
UDP
ICMP
Other
Packet Information
Ethernet Frame
IPv6

Fig -9: The ratio of the protocols being used over the LAN such asTCP,UDP

4.8 Analysis of Application layer

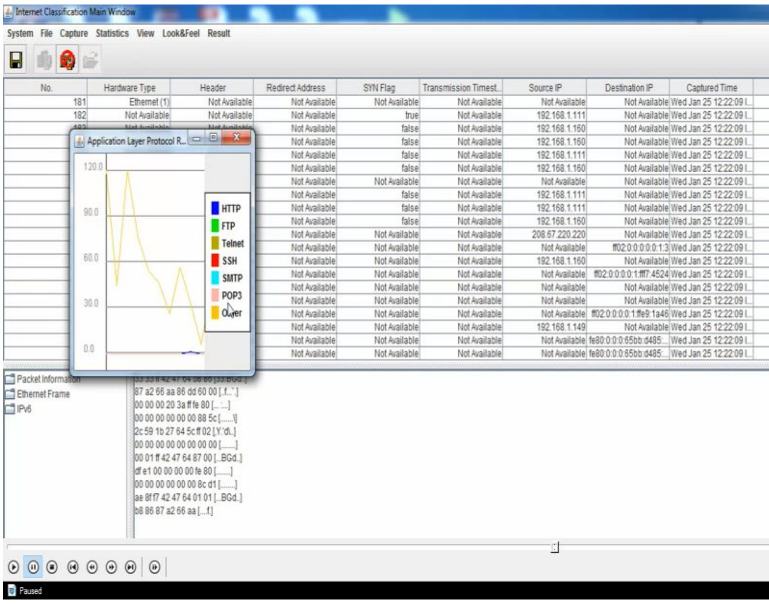
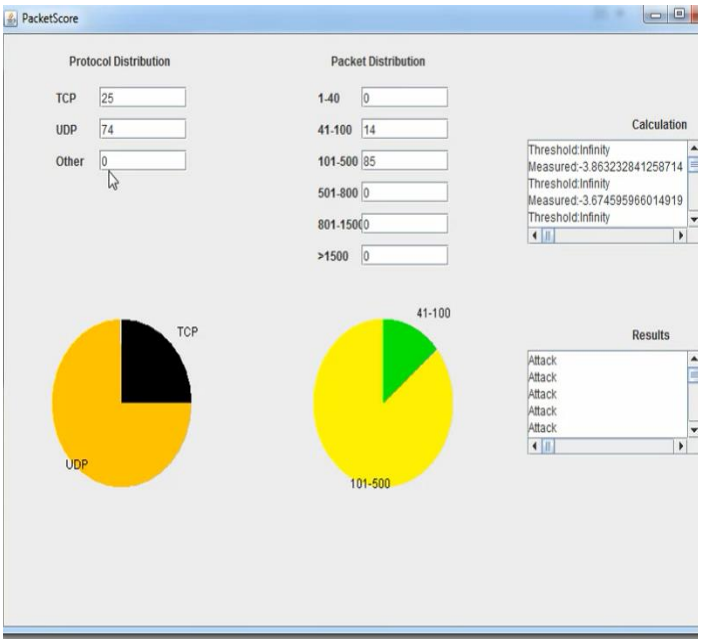

Internet Classification Main Window
System File Capture Statistics View Look&Feel Result
No. Hardware Type Header Redirect Address SYN Flag Transmission Timest... Source IP Destination IP Captured Time
Application Layer Protocol R...
120.0
90.0
60.0
30.0
0.0
HTTP
FTP
Telnet
SSH
SMTP
POP3
Other
Packet Information
Ethernet Frame
IPv6
Paused

Fig -10 : The ratio of the protocols being used over the LAN such as HTTP,SMTP,POP3

4.9 Packet Score



PacketScore

Protocol Distribution

TCP 25

UDP 74

Other 0

Packet Distribution

1-40 0

41-100 14

101-500 85

501-800 0

801-1500 0

>1500 0

Calculation

Threshold:Infinity
Measured:-3.863232841258714
Threshold:Infinity
Measured:-3.674595966014919
Threshold:Infinity

Results

Attack
Attack
Attack
Attack
Attack

Fig-11: Determines attack or not and represents graphically.

5. CONCLUSION

In this paper, a new data-mining based approach with combination of hybrid classification algorithms: SVM, Naïve Bayesian on Network IDS for improved and efficient detection of intrusion with probabilistic approach. The detail of the extent to which corruption is on each layer of the network is diagramed in a user friendly manner.

REFERENCES

[1] *International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 10, December 2012 Copyright to IJARCCE www.ijarcce.com 827 Intrusion Detection System using Fuzzy Genetic Approach B.Ben Sujitha1 , R.Roja Ramani2 , Parameswari3.*

[2] *International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07, 2016, R. L. Jalappa Institute of Technology, Doddaballapur, Bangalore, India Network Intrusion Detection System Using various data mining techniques DikshantGupta1 SuhaniSinghal2 Shamita Malik3 Archana Singh4.*

[3] *Expert Systems with Applications 29 (2005) 713–722An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks Ozgur Depren, Murat Topallar, Emin Anarim, M. Kemal Ciliz*.*

[4] *Journal of Network and Computer Applications 30 (2007) 114–132Modeling intrusion detection system using hybrid intelligent systems Sandhya Peddabachigaria , Ajith Abrahamb,, Crina Grosanc , Johnson Thomas.*

[5] *A Practical Implementation of a Real-time Intrusion Prevention System for Commercial Enterprise Databases ULF T. MATTSSON Chief Technology Officer Protegrity ulf.mattsson@protegrity.se http://www.protegrity.com.*

[6] *International Journal of Database Theory and Application Vol.7, No.2 (2014), pp.59-70 http://dx.doi.org/10.14257/ijdta.2014.7.2.06 ISSN: 2005-4270 IJDTA Copyright ⓒ2014 SERSC An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM Vahid Golmah.*

[7] *computers & security 28 (2009) 18–28 Anomaly-based network intrusion detection: Techniques, systems and challenges P. Garcı´a-Teodoroa, *, J. Dı´az-Verdejoa , G. Macia´-Ferna´ndeza , E. Va´zquezb.*

[8] *978-1-4799-8792-4/15/$31.00 c 2015 IEEE Network Intrusion Detection System Using J48 Decision Tree Shailendra Sahu School of Computer and Information Science University of Hyderabad CIAM Lab IDRBT Hyderabad, India shailendrasahu668@gmail.com B M Mehtre CIAM Lab IDRBT Hyderabad, India bmmehtre@idrbt.ac.in.*

[9] *2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE-2015) 978-1-4799-7263-0/15/$31.00 ©2015 IEEE Implementation of Network Intrusion Detection System using Variant of Decision Tree Algorithm Neha G. Relan Prof. Dharmaraj R. Patil.*

[10] *Network Intrusion Detection System Using L1-norm PCA CHOUGDALI Khalid GEST Research group National School of Applied Sciences (ENSA) Ibn Tofail University, Kenitra Email: chougdali@gmail.com ELKHADIR Zyad RLCST Research Laboratory Ibn Tofail University, Kenitra Email: zyad.elkhadir@gmail.com BENATTOU Mohammed.*