# A Novel Filtered Classification Algorithm for Network intrusion detection

**Bala Bhaskara Rao Emani[1], Vijaya Krishna Sonthi[2], Satya Srinivas Maddipati[3]**

*[123]Assistant professor, Dept.of computer science and Engineering, SASI Institute of technology and Engineering, AP, INDIA*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *The network problems have boosted the challenge of data mining approaches for discovering network attacks. In the traditional approach, various classification techniques for identifying various real-time network attacks have been proposed for identifying network attacks by using data mining. Yet the majority of the techniques do not succeed to separate the different kinds of network attacks since the absence of cooperative filtering procedure and better classification algorithms. In this particular proposed approach, a new cooperating selection classifier for filtering the network attacks and a hybrid classification technique for the classification of different DDOS attacks in KDDCUP99 dataset are proposed. Proposed algorithm is a statistical optimizer method is used for fine tuning of the characteristics where as improved decision tree an accurate classifier, is created to detect and classify kinds of DOS type of attacks.**Key Words**:* **Decision Trees, Principal component analysis, network intrusion detection, filtering, Normalization.**

## 1. Introduction:

Intrusion detection system mainly used for prevent the security Having the tremendous development of network-based services and sensitive on networks, the contact information plus the severity of network-based computer attacks have significantly increased. Completely preventing breaches of security is unrealistic by security technologies namely information encryption, control access, and intrusion prevention. Thus, Intrusion Detection Systems (IDSs) play an important function in network Security Network Intrusion Detection Systems (NIDSs) detect attacks by observing various network activities, while Host-based Intrusion Detection Systems (HIDSs) detect intrusions within individual host.

### 1.1 Denial of service attack (DoS):

In computing, a denial-of-service (DoS) attack is an attempt to make a machine or network resource unavailable to its intended users, such as to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet.

**1.2 User –to-Root (U2R):** User to Root attack is a class of attack in which the attacker starts from access to a normal user account on the system and is able to gain root access to the system.

**1.3 Remote to User (R2L):** Remote to user attack is a type of attack the unauthorized person access typically from a remote machine.

**1.4 Probing:** A probing is a attack against the confidentiality information in association with other probing attacks like ,port-scan, ping-sweep,et

## 2. Overview of the Framework

The frame work to applies a novel filtered classification algorithm to identifying or recognize the network intrusion .The frame work in shown in figure 2.1 .The network Ids(NIDS) catches the network traffic and dataset pre-processing in the existing system the random based service patterns. Proposed system the collaborative filtered approach on your probe attacks after filtering is matched, the result is targeted against both classifiers in that case results are when compared to traditional results .The proposed approach to identify the probe and various kind of attacks.

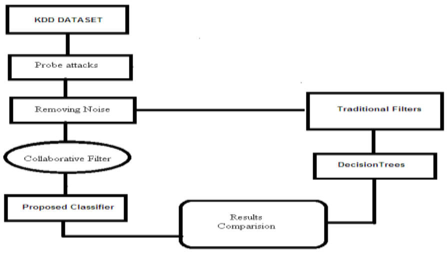
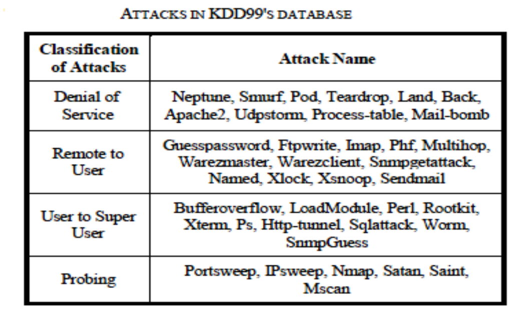


Figure 2.1: Proposed Framework

## 3. Dataset and preprocessing:

The DARPA dataset is commonly to evaluate all most of IDSs. The KDD99 (Knowledge Discovery and Dataming)dataset is a subsection ,subdivision, subgroup, subclass associated with DARPA dataset prepared by sal Stofo and wenke lee. This dataset is a pre-processed dataset containing 41 attributes and extracted from the tcp dump information inside the 1999 DARPA dataset. In this dataset containing 4,898,431 connections with attacks shown in Table 3.1 can be used for experimentation.



ATTACKS IN KDD99'S DATABASE

| Classification of Attacks | Attack Name |
|---|---|
| Denial of Service | Neptune, Smurf, Pod, Teardrop, Land, Back, Apache2, Udpstorm, Process-table, Mail-bomb |
| Remote to User | Guesspassword, Ftpwrite, Imap, Phf, Multihop, Warezmaster, Warezclient, Snmpgetattack, Named, Xlock, Xsnoop, Sendmail |
| User to Super User | Bufferoverflow, LoadModule, Perl, Rootkit, Xterm, Ps, Http-tunnel, Sqlattack, Worm, SnmpGuess |
| Probing | Portsweep, IPsweep, Nmap, Satan, Saint, Mscan |

Table: 3.1 Attacks in KDD99's Database

A pre-processing is a technique feature selection to indentify and remove irrelevant attributes that do not play any role in the classification task .Several feature selection methods are available with different search techniques to produce a reduced dataset. In this reduced data set improve the accuracy when compared with original dataset. The feature selection methods are categorized as filter. The result of these methods varies time and accuracy.

## 4. Algorithm:

The existing standard filtering algorithms are not adaptive fashion to currently the conditions when KDD99 dataset is large and generate the false recommendations. In this new method collaborating filter can be used to indentify active probe attacks dynamically based on the network feature.

Step 1: Determining the actual load W(i,j) this action is used to select the best attribute for probe type attack.

Step 2: Using approved pearson's correlation to the similarity between two vector of ratings.

$$sim(i,j)=\frac{\sum_{c\in I_{i,j}}(R_{i,c}-A_i)(R_{j,c}-A_j)}{\sqrt{\sum_{c\in I_{i,j}}(R_{i,c}-A_i)^2*\sum_{c\in I_{i,j}}(R_{j,c}-A_j)^2}}$$

Where $R_{i,c}$ is the of the probe kind of attack c by system protocol i, $A_i$ is the average rating of network protocol i for all the system attributes, and $I_{i,j}$ is the probe attack set both ratings by system protocol i and protocol j.

IMPROVED BINARY DECISION TREE USING NAIVE BAYES AND CLUSTERING TREE ALGORITHM

```
--------------------------------------------------------------------
flag = SF
| service = eco_i
| | diff_srv_rate = 0: anomaly (194.0/9.0)
| | diff_srv_rate != 0: normal (9.0)
| service != eco_i
| | service = ecr_i: anomaly (123.0/11.0)
| | service != ecr_i
| | | service = private
| | | | dst_bytes = 0: anomaly (104.0/6.0)
| | | | dst_bytes != 0: normal (46.0/2.0)
| | | service != private
| | | | dst_bytes = 377: back (3.0)
| | | | dst_bytes != 377
| | | | | dst_bytes = 1367: back (2.0)
| | | | | dst_bytes != 1367
| | | | | | dst_bytes = 8377: teardrop (2.0)
| | | | | | dst_bytes != 8377
| | | | | | | dst_bytes = 18666: land (2.0)
| | | | | | | dst_bytes != 18666
| | | | | | |   dst_bytes = 8314: anomaly (29.0/1.0)
| | | | | | | | dst_bytes != 8314
| | | | | | | | | service = ftp
| | | | | | | dst_bytes = 3081: land (4.0/1.0)
| | | | | | | | | | dst_bytes != 3081
| | | | | | | dst_bytes = 2451: anomaly (5.0)
| | | | | | | | | | | dst_bytes != 2451
| | | | | | | dst_bytes = 2449: anomaly (4.0)
| | | | | | | | | | | | dst_bytes != 2449
| | | | | | | | dst_bytes = 2447: anomaly (3.0)
| | | | | |   | dst_bytes != 2447: normal (29.0/4.0)
| | | | | | | | | service != ftp
| | | | |  | | | | dst_bytes = 25260: anomaly (2.0)
| | | | | | | | | | dst_bytes != 25260
| | | | | | | | | | | dst_bytes = 1141: smurf (2.0)
| | | | | | | | | | | dst_bytes != 1141
| | | |  | | | | | | | dst_bytes = 3490: smurf (2.0)
| | | | | | dst_bytes != 3490: normal (2573.0/98.0)
flag != SF
| service = http
| | diff_srv_rate = 0
| | | srv_serror_rate = 1
| | | | flag = S0: anomaly (3.0)
| | | | flag != S0: normal (2.0)
| | | srv_serror_rate != 1: normal (118.0/3.0)
| | diff_srv_rate != 0: anomaly (54.0/1.0)
| service != http
| | logged_in = 0
| | | service = smtp
| | | | diff_srv_rate = 0: normal (10.0)
| | | | diff_srv_rate != 0: anomaly (6.0)
| | | service != smtp
| | | | dst_bytes = 15: normal (2.0/1.0)
| | | | dst_bytes != 15: anomaly (1939.0/60.0)
| | logged_in != 0: normal (19.0)
Number of Leaves  : 28
Size of the tree : 55
Correctly Classified Instances       5070           95.8231 %
Incorrectly Classified Instances      221             4.1769 %
```

## 5. Experiment Results:

Performance of the system can be calculated on the number of decision trees constructed during the training phase. More number of trees constructed more the amount of accuracy. Figure 5.1 shown the comparison of proposed algorithm with other existing algorithms .The number of decision trees increases, the false positive rate decreases to determine the attacks. Figure 5.2 shows the comparison of different improved classifier algorithms with several different models the number decision trees increases, the execution for given test set increases show the Figure 5.3. In this proposed approach using new collaborating filters has only higher performance in DOS, probe, R2L, U2R.

| Datasize | ProposedAccuracy | ExistingAccuracy |
|---|---|---|
| 1000# | 94.56 | 84 |
| 2000# | 95 | 85.2 |
| 5000# | 95.76 | 89.4 |
| 10000# | 95.3 | 85 |
| 5000# | 95.7 | 89.6 |

Table: 5.1 Data accuracy for proposed and existing system

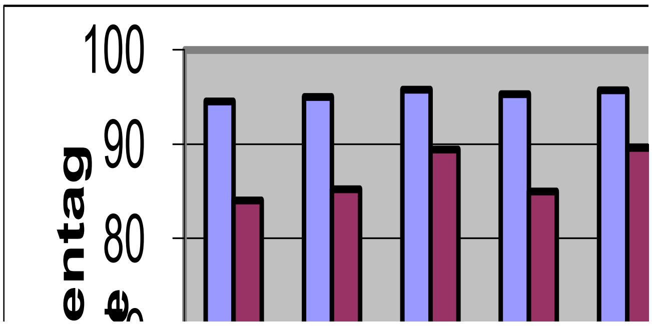


Figure 5.1: Data accuracy for proposed and existing system

| Datasize | Error |
|---|---|
| 1000# | 5 |
| 2000# | 5 |
| 5000# | 4 |
| 10000# | 4.2 |
| 5000# | 4.3 |

Table: 5.1 proposed error rate for different data sizes

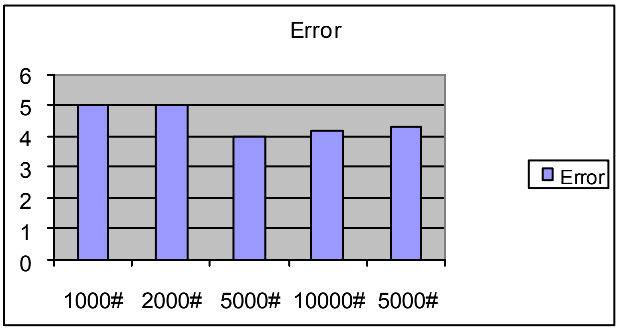


Figure 5.2: Proposed error rate for different data sizes

| Datasize | Error |
|---|---|
| 1000# | 5 |
| 2000# | 5 |
| 5000# | 4 |
| 10000# | 4.2 |
| 5000# | 4.3 |

Table: 5.3 Datasize virsus Treesize

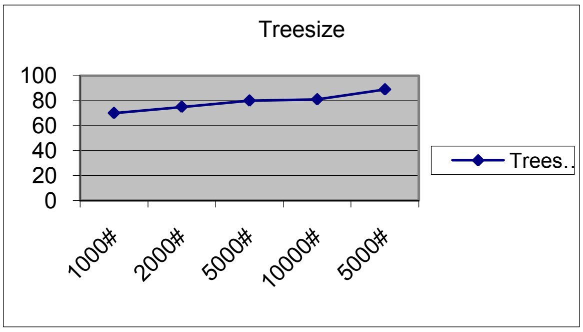


Figure: 5.3 Datasize virsus Treesize

## 6. Conclusion:

In this particular proposed approach, a new cooperating selection classifier for filtering the network attacks and a hybrid classification technique for the classification of different DDOS attacks in KDDCUP99 dataset are proposed. Proposed algorithm is a statistical optimizer method is used for fine tuning of the characteristics where as improved decision tree an accurate classifier, is created to detect and classify kinds of DOS,Probe,U2R and R2L type of attacks.

## References:

[1] Kanwal Garg, Rshma Chawla [2011], "Detection of DDOS attacks using Data Mining", Dated 07-01-2013, International Journal of Computing and Business Research (IJCBR), ISSN(Online):2229-6166, Volume 2, Issue 1.

[2] H. Patel & J. Sarvakar [2011], "Analysis of Data Mining Algorithm in Intrusion Detection", Dated 02-09-2012, International Journal of Emerging Technology and Advanced Engineering (IJETAE), ISSN 2250-2459, Volume 1, Issue 2, U.V.

[3] D. Denning, "An intrusion-detection model", In IEEE computer society symposium on research in security and privacy, 1986, pp. 118- 131.

[4] Zhang Jianpei, Liu Jiandong, Yang Jing. Data Preprocessing Method Research for Web Usage Mining[J]. Computer Engineering and Applications, 2003, (10):191-193(In Chinese)

[5] Yu kai,Xu Xiao-wei,Martin Ester,et al.Collaborative Filtering and Algorithms:Selecting Relevant Instances for Efficient and Accurate Collaborative Filtering[C]//Proceedings of the Tenth International Conference on Information and Knowledge Management. 2001:239-246

[6] Huang Z,Chen H,Zeng D.Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering [J].ACM Transactions on Information Systems, 2004, 22(1):116-142

[7] Herlocker J,Konstan J,Terveen L,et al.Evaluating Collaborative Filtering Recommender Systems.ACM Trans.on Information Systems(TOIS),2004,22(1):5-53

[8]KDD'9datasets,The UCI KDD Archive,http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, Irvine, CA, USA, 1999.