Network Intrusion Detection using Layered Approach and Hidden Markov Model

Archana I. Patil
Research Scholar
SSBT's College of Engineering & Technology, Bambhori, Jalgaon, Maharashtra, India

Girish Kumar Patnaik, PhD
Professor
SSBT's College of Engineering & Technology, Bambhori, Jalgaon, Maharashtra, India

Ashish T. Bhole
Associate Professor
SSBT's College of Engineering & Technology, Bambhori, Jalgaon, Maharashtra, India

ABSTRACT
Traditional intrusion detection systems uses either anomaly based or signature based technique. Both of these techniques have some problems. In anomaly based intrusion detection, the strategy is to suspect an unusual activity and thereby to continue further investigation. This approach is particularly effective against novel attacks. Signature based intrusion detection system detects known attacks timely and efficiently. For this approach, it is important to know the attack.

The proposed system introduces a hybrid of anomaly based and signature based technique. The proposed system uses layered approach to get the results faster. Each layer in the layered approach is independent to detect and block an attack. Four different layers Probe, U2R, R2L and DOS are assigned with different features. The proposed hybrid technique with Hidden Markov Model can give better results compared to signature based and anomaly based intrusion detection techniques alone.

Keywords
Intrusion detection, Layered approach, Hidden Markov Model, Network security, Decision trees, Naive Bayes.

1. INTRODUCTION
Intrusion detection is defined as the problem of identifying individuals who are using a computer system without authorization and those who have legitimate access to the system but are abusing their privileges. Optionally intrusion detection is the identification of attempts to use a computer system without authorization or to abuse existing privileges [7]. According to Heady et al. where an intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource, disregarding the success or failure of those actions [9]. The definition of an intrusion detection system does not include preventing the intrusion from occurring, only detecting it and reporting it to an operator [9].

Depending on the way of distributing components, two types of intrusion detections are :

The centralized intrusion detection system is one where the analysis of the data is performed in a fixed number of locations, independent of how many hosts are being monitored[3]. We do not consider the location of the data collection components, only the location of the analysis components such as IDES, IDIOT.

The distributed intrusion detection system is one where the analysis of the data is performed in a number of locations proportional to the number of hosts that are being monitored. Again, we only consider the locations and number of the data analysis components, not the data collection components such as DIDS, GrIDS.

The intrusion detection can also be classified as following:

1.1. Anomaly detection
The strategy is to suspect of what is considered an unusual activity for the subject (users, processes, etc.) and carry on further investigation. This approach is particularly effective against novel (i.e. previously unknown) attacks. Its main drawback is the high rate of false positives, because any legitimate but new activity can raise an alert.

1.2 Signature detection
The strategy is to look for some special activity (signature) of previously known attacks. Signature based detection systems detect previously known attack in a timely and efficient way. The main issue of this approach is that in order to detect an intrusion this must to be previously detected [14] [15].

1.3 Hybrid detection
Until today, only one technique is used at a time. The proposed approach uses both signatures based and anomaly based hybrid technique as shown in figure 1. That is we are developing hybrid system using HMM based layered approach for NIDS. We also integrate the Layered Approach with the HMMs to gain the benefits of computational efficiency and high accuracy of detection in a single system [16]. By using this we get fast result because we are using layered approach .Layered approach means we have different four layers as PROBE , DOS , U2R, R2L and for every layer different features are assigned and whenever we got some malicious attack that attack must be detected at that moment, the attack is not allowed to go further. Due to this technique, speed of the operation increases [5] [11].

A hidden Markov model (HMM) is a statistical generative model in which the system being modelled is assumed to be a Markov process with unobserved state. An HMM can be considered as the simplest dynamic Bayesian network. An HMM is like a finite state machine in which not only transitions are probabilistic but also output. An HMM is a doubly stochastic process with an underlying stochastic process that is not observable, and can only be observed through another set of stochastic processes that produce the sequence of observed symbols[8][10] . HMM is a useful tool to model sequence information. This model can be thought of as a graph with N nodes called 'state' and edges representing transitions between those states. Each state node contains

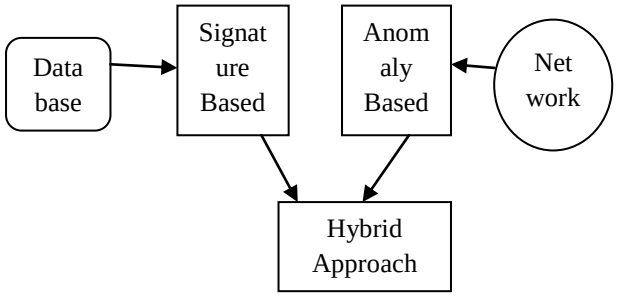

Fig 1: Overview of Hybrid Approach for Intrusion Detection

Initial state has distribution and observation probabilities at which a given symbol is to be observed. An edge maintains a transition probability with which a state transition from one state to another state is made.

2. RELATED WORK

The history of intrusion detection starts since 1980s. Since then, a number of methods and frameworks have been proposed and many systems have been built to detect intrusions.

Data Mining Approach: Data mining approaches for detecting intrusions in W. Lee and S. Stolfo, "Data Mining Approaches for Intrusion Detection [1], W. Lee, S. Stolfo, and K. Mok, "Mining Audit Data to Build Intrusion Detection Models[2], and W. Lee, S. Stolfo, and K. Mok, "A Data Mining Framework for Building Intrusion Detection Mode[3].Data mining approaches for intrusion detection include association rules and frequent episodes, which are based on building classifiers by discovering relevant patterns of program and user behavior. Association rules and frequent episodes are used to learn the record patterns that describe user behavior. These methods can deal with symbolic data.

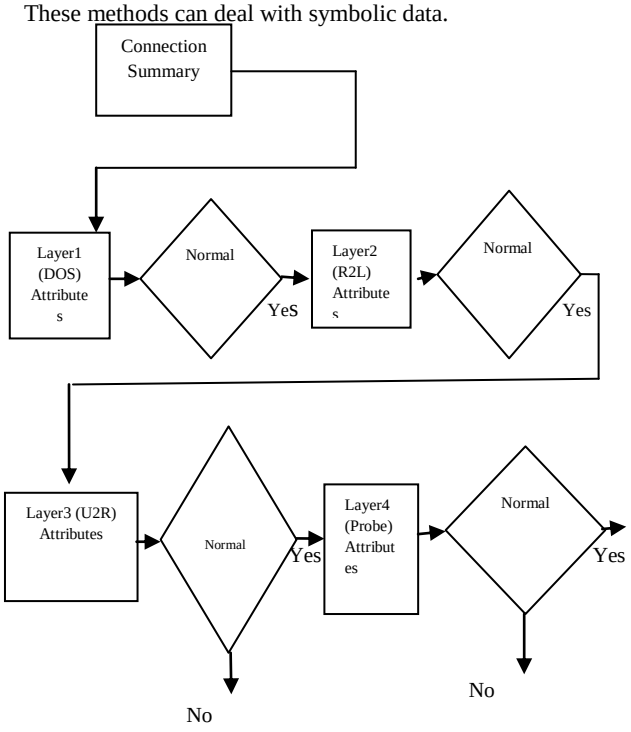

Fig 2: Representation of Layered Approach

and the features can be defined in the form of packet and connection details[13]. Data Clustering Methods: Data clustering methods in intrusion detection includes the k-means and the fuzzy c-means clustering methods used in L. Portnoy, E. Eskin, and S. Stolfo, "Intrusion Detection with Unlabeled Data Using Clustering"[4] and H. Shah, J. Undercoffer, and A. Joshi, "Fuzzy Clustering For Intrusion Detection . The main drawbacks in these clustering technique is that it is based on determining the numeric distance between the observations resulting that, the observations must be numeric .Hence observations with symbolic features cannot be easily identified and used for clustering, resulting in inaccuracy for finding the attacks.

Naive Baye's classifiers: The next approach discussed here in intrusion detection is Naive Bayes classifiers in N.B. Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs. Decision Trees in Intrusion Detection Systems" [6]. Those approaches will make strict independent assumption between the features in an observation records which resulting in very low detection accuracy when the features in the observation are having correlation between them. However, those networks may tend to be attack specific and construct a decision.

Network based on special characteristics about individual attack groups [24]. Thus, the size of a Bayesian network increases rapidly whenever the number of features and the type of attacks are increases.

Decision Trees: The intrusion detection also performed using Decision trees approach in N.B. Amor, S. Benferhat and Z. Elouedi "Naive Bayes vs. Decision Trees in Intrusion Detection Systems" [6]. This approach presents decision tree techniques that are used to automatically learn intrusion signatures and perform the classification activities in computer network systems as normal or intrusive. This approach usually has very high speed of its operation and high accuracy of attack detection.

Neural Networks: The neural network components also used for finding the intrusive events in the network [5]. The neural network in intrusion will work well with correlated kind of data in the observation (noisy data) .However, the neural networks require large amount of data for training the observations and also it is often difficult to select the best possible architecture for a neural networks.

3. PROPOSED WORK

The following sections describe the proposed layered approach for attack detection in computer network, the types of sub-attacks and algorithm for the approach.

3.1 Layered Approach

The goal of using a layered model is to reduce computation and the overall time required to detect anomalous events. Figure 2 shows the representation of layered approach. The time required to detect an intrusive event is significant and can be reduced by eliminating the communication overhead among different layers. This can be achieved by making the layers autonomous and self-sufficient to block an attack without the need of a central decision-maker. Each layer in the LIDS framework is trained separately and then deployed sequentially. We define four layers that correspond to the four attack groups mentioned in the data set. They are Probe layer, DoS layer, R2L layer and U2R layer. Each layer is then separately trained with a small set of relevant features [12]. Feature selection is significant for Layered Approach and discussed in the next section. In order to make the layers independent, some features may be present in more than one layer. The layers essentially act as filters that block any

anomalous connection. We have four different attacks probe attack, dos attack, u2r attack and r2l attack corresponding to four different layers Probe layer, R2L layer,U2R layer and DOS layer.

3.1.1 Probe Layer

The probe attacks are aimed at acquiring information about the target network from a source that is often external to the network.

3.1.2 DoS Layer

The DoS attacks are meant to force the target to stop the service(s) that is (are) provided by flooding it with illegitimate requests.

3.1.3 R2L Layer

R2L attacks are one of the most difficult to detect as they involve the network level and the host level features.

3.1.4 U2R Layer

The U2R attacks involve the semantic details that are very difficult to capture at an early stage.

3.2 Sub-attacks in each layer

Each attack has sub-attacks as shown in Table 1.

Table 1. Sub-attacks in each Layer

Layer type	Sub-attacks in the layer
DOS Layer	Smurf, Teardrop, Neptune, Back, Land
R2L Layer	Ftp_write, Warezclient, Phf, Multihop, Spy,Warezmaster
U2R Layer	Buffer_overflow, Loadmodule, Perl, Rootkit
Probe Layer	Satan, Portsweep, Ipsweep, nmap

3.3 Algorithm of Proposed System

i] Input is taken as a KDD cup dataset (1999).
ii] Perform data labeling on input.
(in this step +1 is assign if record is normal , -1 if record present attacks)
iii] Perform training on it.(find attacks)
iv] Perform testing (find sub-attacks from record)
v] Find attack detection rate
attack detection rate = attacks detected automatically/attacks detected manually*100.
vi] For same input file by using HMM and WEKA tool find attack detection rate for each technique.
vii] Find time for each technique,
i) Find start time, convert it into seconds
ii) Find end time, convert it into seconds
iii) Perform end time - start time = time required for Layered approach
iv) Perform same for HMM and WEKA tool
viii] Find attack detection rate of HMM
go to step no. v
after finding attack detection rate goto step no. ix
ix] Find attack detection rate of WEKA
go to step no. v
after finding attack detection rate go to step no. x
x] Find time required to detect attacks
go to step no. vii
after finding time goto step no. x
xi] Find time required to detect attacks
go to step no. vii
after finding time goto step no. xii
xii] Compare attack detection rate of layered approach, HMM, WEKA from step no. v, viii and ix
xiii] Compare time required to detect attacks from step no. vi, x and xi.

4. EXPERIMENTAL RESULTS

Table 2 summarizes the number of attacks present in input file. Input file is nothing but KDD cup dataset 1999 and number of attacks detected by proposed layered approach. There are four different attacks as DOS, R2L, U2R, Probe.

Table 2. Attacks found automatically and manually

Name of attack	Number of attacks present in Input file	Number of attacks detected by proposed Layered Approach
DOS	8172	5824
R2L	1354	1192
U2R	60	36
PROBE	11026	9533

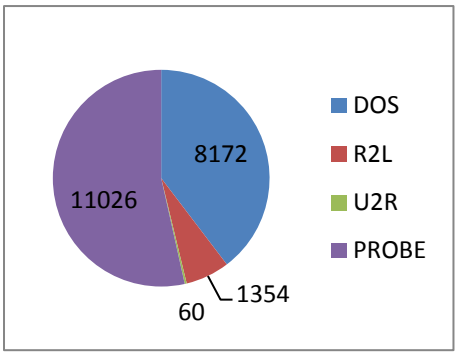

Fig 3: Attacks found manually and automatically

The first type of attack is DOS attack. Table 3 summarizes analysis of DOS attack. There are sub attacks of DOS as smurf, teardrop, Neptune, back, land. Figure 4 shows number of sub attacks present and number of sub attacks detected by proposed layered approach.

Table 3. Analysis of DOS attacks

Name of sub-attacks	Number of attacks present	Number of attacks detected by proposed LA
Smurf	527	340
Teardrop	183	49
Neptune	7212	5190
Back	249	245
Land	1	0

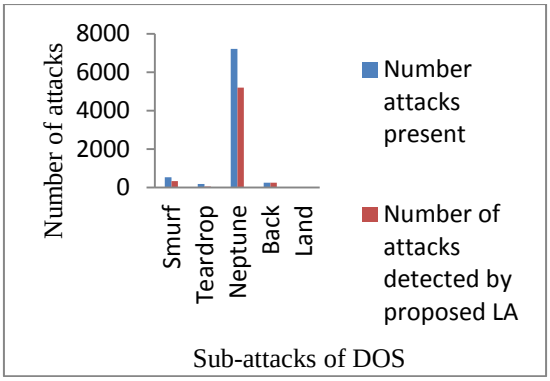

Fig 4 : Analysis of DOS attacks

Second type of attack is R2L attack. Table 4 summarizes analysis of R2L attack. There are sub attacks of R2L as Ftp_write, Warezclient, Phf, Multihop, Spy, Warezmaster. Figure 5 shows number of sub attacks present and number of sub attacks detected by proposed layered approach.

Table 4. Analysis of R2L attacks

Name of sub-attacks	Number of attacks present	Number of attacks detected by proposed LA
ftp_write	341	335
Warezclient	505	456
Phf	456	370
Multihop	16	15
Spy	13	3
Warezmaster	15	13

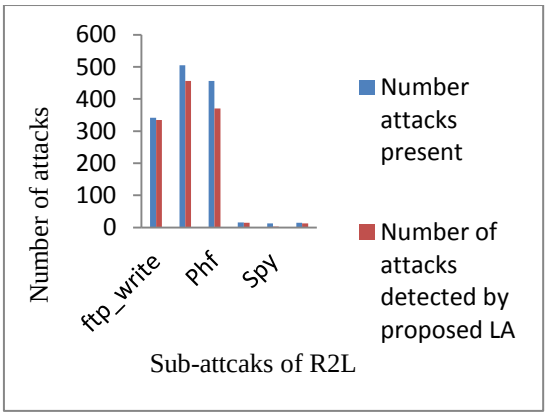

Fig 5: Analysis of R2L attacks

Third type of attack is U2R attack. Table 5 summarizes analysis of U2R attack. There are sub attacks of U2R as Buffer_overflow, Loadmodule, Perl, Rootkit. Figure 6 shows number of sub attacks present and number of sub attacks detected by proposed layered approach.

Table 5. Analysis of U2R attacks

Name of Sub-attacks	Number of attacks present	Number of attacks detected by proposed LA
Buffer_overflow	40	23
Loadmodule	1	0
Perl	0	0
Rootkit	19	13

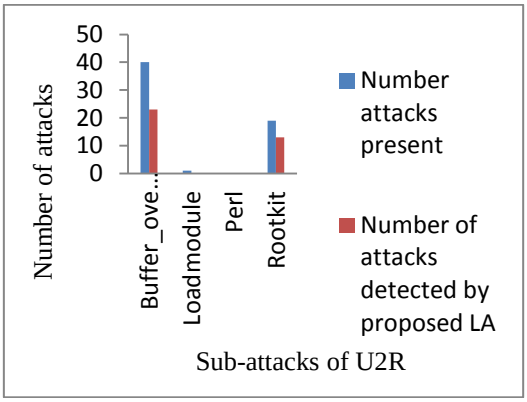

Fig 6: Analysis of U2R attacks

Fourth type of attack is Probe attack. Table 6 summarizes analysis of Probe attack. There are sub attacks of Probe as Satan, Portsweep, Ipsweep, nmap. Figure 7 shows number of sub attacks present and number of sub attacks detected by proposed layered approach.

Table 6. Analysis of Probe attacks

Name of sub-attacks	Number of attacks present	Number of attacks detected by proposed LA
Satan	2119	2089
Portsweep	4302	4293
Ipsweep	700	336
Namp	3905	2815

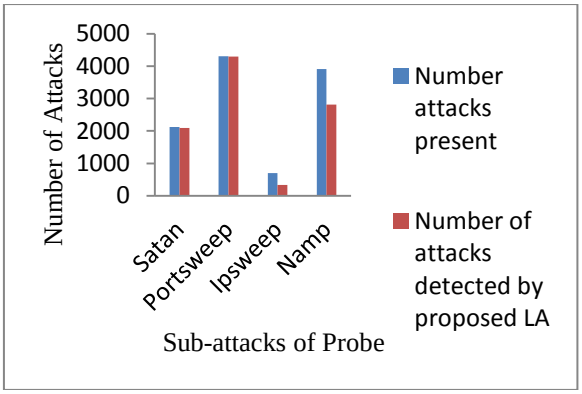

Fig 7: Analysis of Probe attacks

Attack detection rate is calculated as
Attack detection rate =

$$\frac{Number\ of\ attacks\ detected\ by\ Proposed\ Layered\ Approach}{Number\ of\ attacks\ present\ in\ Input\ File} * 100 \quad (1)$$

From equation (1) Attack detection rate of proposed Layered Approach = 97.51%.

Table 7 summarizes the attack detection rate and time required to detect the attacks of Hidden Markov Model, Weka with decision Tree & Naïve Bayes and proposed layered approach.

Table 7. Comparison of attack detection with different techniques

Name of Technique	Attack detection rate in %	Time required in seconds
HMM	95.054321	4
WEKA with DT	95.054392	3
WEKA with NB	93.51329	4
PLA	97.53031	≤1

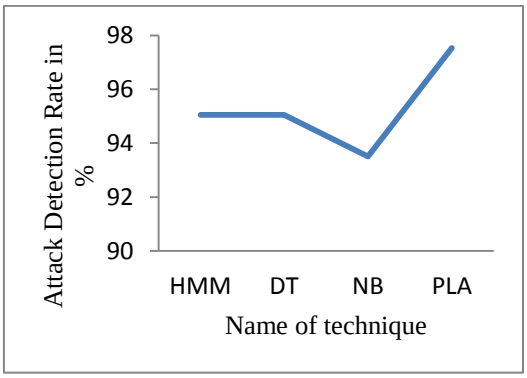

Fig 8: Comparison of rate of attack detection with different techniques

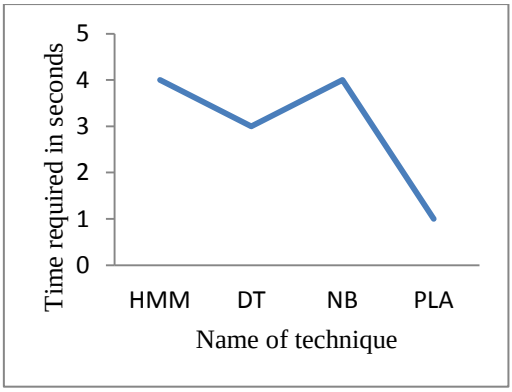

Fig 9: Comparison of time required for attack detection

HMM - Hidden Markov Model
DT - Decision Tree
NB - Naive Bayes
PLA - Proposed Layered Approach

Figure 8 shows that the proposed system gives attack detection rate of 97.5 % which is more as compared to HMM and NB. Also time required to detect these attacks is very less as shown in figure 9.

5. CONCLUSION

Network Intrusion Detection is becoming very challenging day by day. Hidden Markov Model and WEKA Tool can detect attacks in network but the efficiency is less. These methods are not efficiently able to detect R2L and U2R attack. The proposed layered approach helps in detecting and identifying an attack at a particular layer which expedites the intrusion response mechanism, thus minimizing the impact of an attack. The proposed layered approach has experimentally proven to be more efficient for attack detection in network than traditional Hidden Markov Model and WEKA Tool with Decision Tree & Naive Bayes.
In future, the pipelined layer approach with multicore processor can be used to get higher rate for attack detection.

6. REFERENCES

[1] Kapil Kumar Gupta, Baikunth Nath, "Layered Approach Using Conditional Random Fields for Intrusion Detection", *IEEE Transaction On Dependable and Secure Computing*, Vol. 7, No. 1, January –March 2010.

[2] T. Abraham, "IDDM: Intrusion Detection Using Data Mining Techniques", *International Journal of Network Security & Its Applications (IJNSA)*, Vol. 2, No. 2, 2008.

[3] N.B. Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs. Decision Trees in Intrusion Detection Systems", *Proc. ACM Symp. Applied Computing*, 2004.

[4] Gupta, Kapil Kumar, Baikunth Nath, and Kotagiri Ramamohanarao, "Conditional random fields for intrusion detection", In *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st IEEE International Conference on*, vol. 1, pp. 203-208, 2007.

[5] Yusufovna, Sattarova Feruza, "Integrating intrusion detection system and data mining", In *Ubiquitous Multimedia Computing, 2008. UMC'08. IEEE International Symposium on*, pp. 256-259, 2008.

[6] Christopher Kruegel, Darren Mutz, William Robertson, Fredrik Valeu, "Bayesian Event Classification for Intrusion Detection", In *Computer Security Applications Conference, 2003. Proceedings. 19th IEEE Annual*, pp. 14-23, 2003.

[7] Portnoy, Leonid, "Intrusion detection with unlabeled data using clustering" 2000.

[8] Wu, Yu-Sung, Bingrui Foo, Yongguo Mei, and Saurabh Bagchi, "Collaborative intrusion detection system (CIDS): a framework for accurate and efficient IDS", In *Computer Security Applications Conference, 2003. Proceedings. 19th IEEE Annual*, pp. 234-244, 2003.

[9] Autonomous Agents for Intrusion Detection, 2010. http://www.cerias.purdue.edu/research/aafid/

[10] Probabilistic Agent Based Intrusion Detection, 2010. http://www.cse.sc.edu/research/isl/agentIDS.shtml

[11] Akbar, Shaik, K. Nageswara Rao, and J. A. Chandulal, "Intrusion detection system methodologies based on data analysis", *International Journal of Computer Applications*, Vol. 5, No. 2, 2010.

[12] Wang, Wei, Xiaohong Guan, Xiangliang Zhang, and Liwei Yang, "Profiling program behavior for anomaly intrusion detection based on the transition and frequency property of computer audit data", *computers & security*, Vol. 25, No. 7, 2006.

[13] Landwehr, Carl E., Alan R. Bull, John P. McDermott, and William S. Choi, "A taxonomy of computer program security flaws", *ACM Computing Surveys (CSUR)*, Vol. 26, No. 3, 1994.

[14] Nicholas Pappas, "Network IDS & IPS Deployment Strategies", 2008.

[15] Bouzida, Yacine, and Sylvain Gombault, "Eigenconnections to intrusion detection", In *Security and Protection in Information Processing Systems*, pp. 241-258. Springer US, 2004.

[16] Kim, Dong Seong, and Jong Sou Park, "Network-based intrusion detection with support vector machines" , In *Information Networking*, pp. 747-756. Springer Berlin Heidelberg, 2003.