Strengthening Intrusion Detection using Conditional Random Fields and Layered Approach

Miss Ketaki Mohan Patil[1] Prof. A.B.Chougule[2]

[1,2]Department of Computer Science and Engineering

[1,2]Shivaji University, Kolhapur Maharashtra, India

Abstract— *Intrusion detection systems provide way of detecting attacks on systems by monitoring network activities for malicious or abnormal behaviors. As the use of network has become a part of each and everyone's daily routine today's intrusion detection system faces number of challenges. Network intrusion detection system has become an important component in network security. In this paper we address two issues one conditional random field (CRFs) and second layered approach. we integrate both of them to improve overall accuracy and efficiency. The system will detect attacks like Probe layer attack, Dos attack, R2L attack and U2R attack.*

Keywords— Intrusion detection, Conditional Random fields, Layered approach, network security

I. INTRODUCTION

Intrusion detection is defined to be the process of monitoring the events occurring in a computer system or network and noticeably different from normal system activities and thus detectable [2].

In recent years, the security has become a critical part of any industrial and organizational information systems. The intrusion detection system is an effective approach to deal with the problems of networks and so used to detect the different kinds of attacks. In a network system the intruders will always take the benefits of the weaknesses of the system and an efficient network intrusion detection system is always needed to filter out all the attackers and hackers from legitimate users.

In this paper we have demonstrated College Management Website for which large numbers of login attempts are considered. Using a intrusion detection system it would be possible to classify those attempts into legitimate and illegitimate attempts to login. Then it would be possible to block IP addresses that are generating large number of attacks.

The system will describe the use of Conditional Random Fields (CRFs) [3] and Layered Approach fcor signature based intrusion detection systems. The proposed system can address both the issues of accuracy and efficiency using Conditional Random Fields and Layered Approach for signature based intrusion detection system.

II. BACKGROUND

A. Intrusion detection system

An intrusion detection system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions and misuse functions, and does following activities.

- Monitors and analyzes user and system activities
- Analyzes system configurations and vulnerabilities
- Access system and file integrity
- Recognizes patterns of typical attacks
- Analyses abnormal activity patterns
- Tracks user policy violations

B. Related Work

The field of intrusion detection and network security has been around since late 1980s after the influential paper from Anderson since then, a number of methods and frameworks have been proposed and many systems have been built to detect intrusions. Many techniques such as association rules, clustering, genetic algorithms, artificial neural networks, Baye's classifiers and others have been applied to detect intrusions. Researchers have used following techniques for intrusion detections.

- Data mining approaches for intrusion detection such as association's rules and frequent episodes based on building classifiers are used to discover relevant patterns that describe user behavior [4].
- Data clustering methods such as K-means and fuzzy C-means have also been applied for intrusion detection system [5].
- Naive Bayes Classifies and Bayesian network is also used for intrusion detection [6].
- Decision trees have also been used for intrusion detection. The decision trees select the best features for each decision nodes during the construction of the tree based on some well-defined criteria [7].
- Other approaches for detecting attacks include the use of genetic algorithm and autonomous and probabilistic agents for intrusion detection.

Observations from Literature Survey Show that

- The data mining approach for intrusion detection uses association rules and frequent episodes to learn the record pattern that expresses user behavior. These methods can deal only with symbolic data and the features can be defined in the form of packet and connection details.
- The clustering techniques are based on calculating numeric distance between the observations and hence the observations must be numeric.
- The Naive Bayes Classifiers make strict assumptions between the features in an observation resulting in lower attack detection accuracy when the features are correlated which is often the case for intrusion detection.
- Bayesian Network tend to attack specific and build a decision network based on special characteristics of individual attacks Thus, the size of Bayesian network increases rapidly as the number of features and type of attacks modeled by the network increases.
- Genetic algorithm methods are generally aim at developing a distributed intrusion detection system.

III. CONDITIONAL RANDOM FIELDS

Conditional random fields are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. Lafferty, McCallum and Pereira define a CRF on observations X and random variables Y as follows Let G=(V,E) be a graph such that :

$$Y = (Yv)_{v \in V}$$

So that Y is indexed by the vertices of G. Then(X, Y) is a conditional random field in case, when conditioned on X, the random variables Y v obey the markov property with respect to the graph is:

$$P(Yv|X, Yw, w \neq v) = P(Yv|X, Yw, w \approx v)$$

Where w ≈v means that w and v are neighbors in G.i.e a CRF is a partially directed graphical model whose nodes can be divided into exactly two disjoint sets X and Y, the observed and output variables respectively the conditional distribution is then modeled.

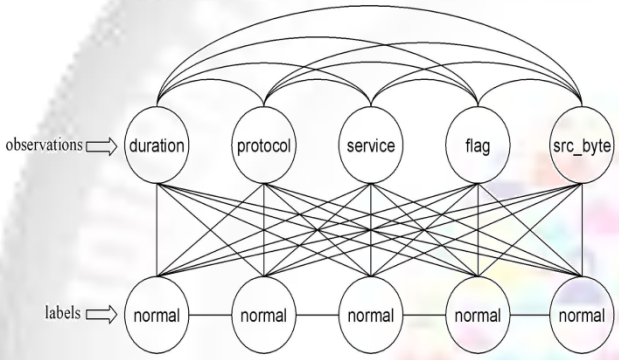

Fig. 1: Graphical representation of CRF

IV. LAYERED SECURITY MODEL FOR INTRUSION DETECTION

Layered security model is a sequential model in which number of security checks are performed one after other in sequence. The objective of using a layered security model is to reduce computation and the overall time required to detect anomalous events. In model every layer is trained separately. We define four layers that corresponds to the four attack groups. They are Probe layer, Dos layer,R2L layer and U2R layer. Every layer is separately trained with a small set of relevant features. The layers essentially act as filters that block any anomalous connection, thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion.

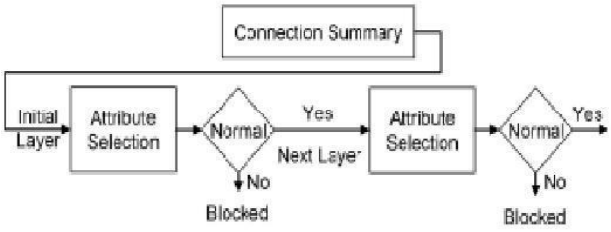

Fig. 2: Layered Approach representation.

V. FEATURE SELECTION FOR EACH LAYER

For the system every layer is separately trained to detect a single type of attack. We select features for each layer based upon the types of attacks that the layer is trained to detect.

Layer Name	Features Selected
Probe Layer	– Duration of connection – Service – Source Bytes – Flag
Dos Layer	– Percentage of connections having same destination host and same service – Total TCP synchronization count packets
R2L Layer	– Duration of connection – Number of failed login attempts – SQL injection – Cross site scripting
U2R Layer	– Number of file creations – Number of shell prompts invoked

Table I: Feature Selection For Each Layer

VI. FEATURE SELECTION FOR EACH LAYER

A. *Training:*

1) A real-time data is collected periodically.
2) Perform features selection for each layer.
3) Train a separate model with CRFs for each layer using the features selected from step2.
4) Pass on the connections labeled as normal to the next layer.

B. *Testing:*

1) For every (next) test instance perform steps 6 through 9.
2) Test the instance and label it either as attack or normal.
3) If the instance is labeled as attack, block it and identify it as an attack represented by the layer name at which it is detected and go to step 5. Else pass the sequence to the next layer.
4) If the current instance is not the last layer in the system, test the instance and go to step 7. Else go to step 9.
5) Test the instance and label it either as normal or as an attack. If the instance is labeled as an attack, block it and identify it as attack corresponding to the layer name.

VII. ATTACKS CATEGORIZED IN LAYERED APPROACH

A. *Probe Layer:*

The probe attacks acquire information about the target network from a source that is often external to the network. It is a class of attack in which an attacker scans a network of computers to gather information or find known vulnerabilities. Example port scanning, Nmap etc. This module will have rules relating to block all the attacks in which the user tries to access any service without actually being logged in. Rules will be developed in such a way that any user in no way would be able to access any service without being logged in.

For the implementation of probe layer the college management website is used. Intrusion Detection database tables called as Real time data table and trained data table

are maintained. Both the tables are analyzed and maintained so as to detect the attacks.

The probe layer attack will be detected under two services HTTP and FTP. For the HTTP service the user will be told to login, the login user can be a normal user or can be an attacker.

Rules are developed using the features such as source and destination port, Protocol, syn packet status, packet length, ack flag and finish flag to determine the attack. For FTP service FileZilla Server which acts as client and wing FTP which act as server are used to perform the file manipulation activities. Same features as HTTP are used to detect the attack.

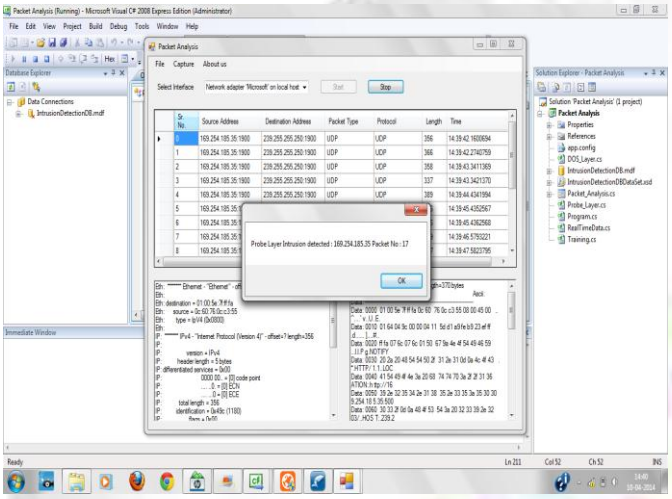

Fig. 3: Attack Detected at Probe Layer

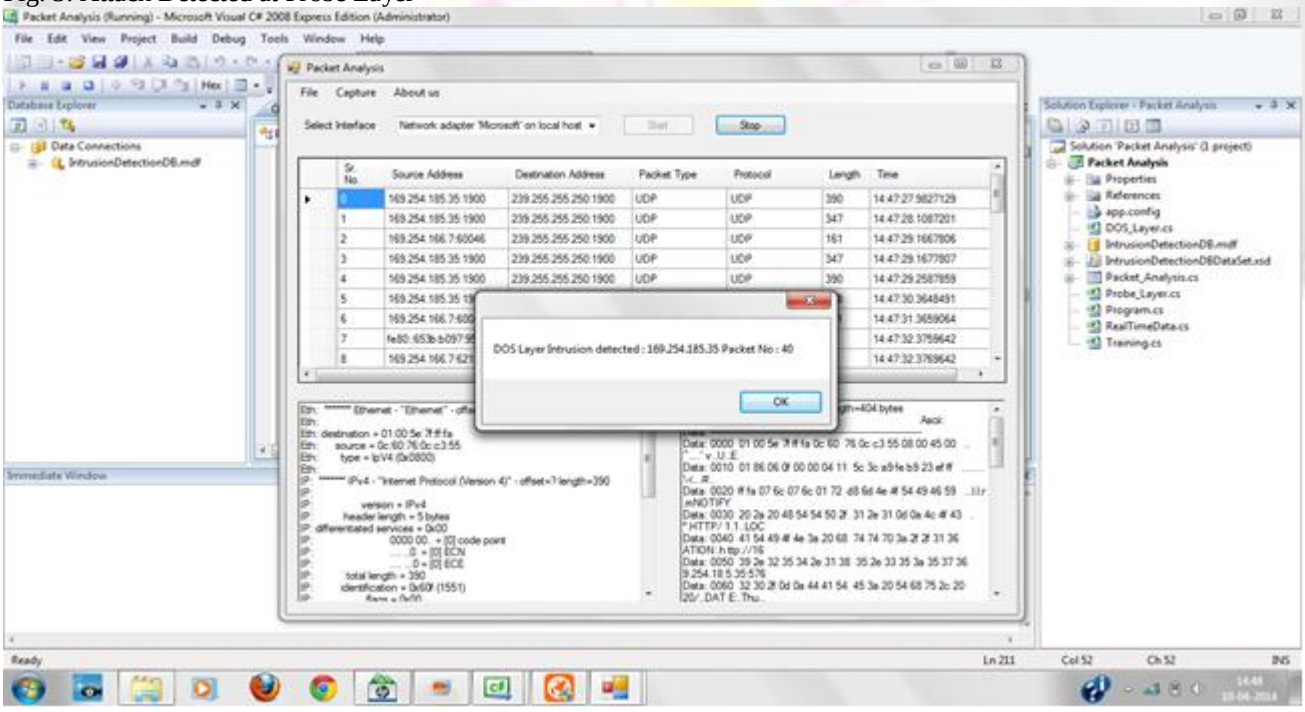

Fig. 4: Attack Detected at DOS Layer

Once the attack is detected for any of the service in probe layer the legitimate user will be blocked for the software level as well as through firewall. If the attack is not detected at this layer the data will be carried on to the next layer i.e, Dos layer for analysis.

B. Algorithm for probe layer

1) Step 1: Initialize service ports, flags and source bytes.
2) Step 2: Get server IP Address to track real time traffic received to server.
3) Step 3: Extract packet details.
4) Step 4: If packet destination address is not equal to server IP Address. Then move to step 10.
5) Step 5: Else check whether service port is valid. If no then move to step 9.
6) Step 6: Else if check flags and source bytes are valid.
7) Step 7: If no then move to step 9.
8) Step 8: Move to step 11.
9) Step 9: Intrusion detected - return IP address and packet number.
10) Step10: Discard packet to analyze.
11) Step 11: Go to step 3 if real time traffic is available.
12) Step 12: Exit Probe Layer

A. Dos Layer:

The DoS attacks are meant to force the target to stop the services that are provided by flooding it with illegitimate requests. In such attacks, an attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. Examples ping of death, tear drop etc. This module will have rules to block all the attacks which can keep the server busy for long enough that many valid and authenticated users won't be able to give the services.

For Denial of service SYN flood is used. A flood occurs when a host sends a flood of TCP/SYN packets, often with a forged sender address. Each of these packets is handled like a connection request, causing the server to spawn a half-open connection, by sending back a TCP/SYN-

ACK packet (Acknowledge), and waiting for a packet in response from the sender address (response to the ACK Packet). However, because the sender address is forged, the response never comes. These half-open connections saturate the number of available connections the server is able to make, keeping it from responding to legitimate requests until after the attack ends.

For analysis of these attacks features such as duration, protocol type, flag, source, bytes, count, dst_host_same_srv_rate,dst_host_serror_rate,dst_host_srv_serror_rate, dst_host_rerror_rate are used. The duration table is maintained to identify no of packet received from specific IP per minute. we also maintain a Train Dos field which will use no of TCPSYN packet received per minute.

In this module as a normal user, the user will get logged to the application of college management website and can register or can get login. As an attacker the user will just request for the connection of the application will not use any of the service but will just keep on flooding the messages so as to establish the connection. To do so an attacker can also use forged IP addresses which will keep on requesting the connection.

At the server end the server will keep on continuously sending the ack packets as a reply for syn packets and hence will fail to provide service to all its normal users.

To identify the attacker an array of no of IP addresses and the TCPSYN count i.e, syn packet received per minute is maintained. Also at server end the uploading and downloading limit is maintained by using Net limiter. Rules are generated by maintaining a threshold value for syn count received per minute by each and every user. If this value is exceeding to our threshold determines the attacker.
As an attacker use Engage Packet Builder which uses forged

IP addresses continuously to request the connection. Because of this the no of TCPSYN count packets will be increased tremendously. As the syn packet value and the uploading speed value is violating our rule an attacker will be blocked. Apart from engage packet creator application can also be used as an attacker. Once the IP address of the attacker is blocked the illegitimate user will not be allowed to access any services further.

B. Algorithm for Dos Layer

1) Step 1: Initialize number of time interval schedules depending upon load of traffic.
2) Step 2: Retrieve threshold counts from the database.
3) Step 3: Start processing packets and maintain the TCP SYN count for time interval.
4) Step 4: Check for time interval whether TCP SYN count exceeds threshold value for that schedule.
5) If yes return intrusion detected with IP Address and packet number.
6) Else Move to step 3 or step 5 if no real traffic is available.
7) Step 5: Exit DOS layer.

VIII. CONCLUSION

In this paper we discussed about layered approach and conditional random fields used for intrusion detection system. The system can help in identifying an attack once detected at a particular layer module which expedites the intrusion response mechanism. Thus, minimizing the impact of the attack. The further work can be developing and implementing an algorithm to detect attacks for R2L layer and U2R layer.

REFERENCES

[1] Kapin Kumar Gupta, Bai Kunth Nath, Senior member IEEE and Ramamohanraw Kotagiri, member IEEE "Layered Approach using Conditional Random Fields for Intrusion Detection". IEEE transactions on dependable and secure computing VOL 7 No1. Jan-March 2010.
[2] Srinoy S Kurutach W, Chimphlee, "Network anomaly detection using soft computing" proceedings of world academy of science, engineering and technology, VOL9 pp140-144,2005.
[3] Charles Sutton, "An introduction to Conditional Random Fields", University of Edinburgh, 17November2010
[4] Autonomous Agents for intrusion detection, http://www.cerias.purdue.edu/research/aafid/,2010
[5] S Devaraju, S Ramakrishnan, "Detection of accuracy for intrusion detection system using neural network classifier", International conference on Information Systems and computing (ICISC-2013), India.
[6] Neveen I Ghali, "Feature selection for effective anomaly - based intrusion detection", International journal of computer science and network security VOL9 No3, March 2009.
[7] Emma Ireland, "Intrusion Detection with genetic algorithms and fuzzy logic", Division of science and mathematics university of Minnesota, Morris, USA
[8] R.Graham,"FAQ:Network Intrusion Detection Systems".March 21,2000.