CHAPTER 4

IMPLEMENTATION OF HETEROGENEOUS TIME VARIANT ACTIVITY PATTERN BASED NETWORK IMMUNE SYSTEM FOR INTRUSION DETECTION

4.1 PREVIEW

The real world is running over the Internet where the users access different services through the Internet, and everyone spends their almost time in the Internet. There exist various threats to the information of the users and the malicious users or nodes try to steal the information of others without knowing them. We have many securities enforced by the communication systems and the protocols used, but still the malicious users are capable of launching their attack in a successful manner. Generally, the intrusion-detection methods use different features to carry out the intrusion detection process, but we propose a novel approach which uses different attributes of the packets that flow through the network. The method uses time variant patterns of the user activity in heterogeneous nature. For every user, the method maintains the log that specifies the activity of them according to their access. The method maintains the log of the frequency of their access, number of hops a packet has traversed to reach the destination, the set of addresses of nodes through which the packet has traversed, the ttl value of the packet and payload size of the packet. Every user has a specific pattern of their access while they access the service in a genuine manner. Furthermore the frequency of accessing the service in a malicious manner of any genuinely user will be
less or zero. By using these, the intrusion detection can be performed based on these activity patterns. Based on these, the method may compute the weight, which specifies the genuine value of the packet. The legitimate value is computed by splitting the whole log into various time windows and for each time window the pattern of access, and frequency are computed. By using all these feature, a packet legitimate weight is computed. The legitimate weight shows the genuine value of the packet, and the method has produced efficient results in intrusion detection.

### 4.2 INTRODUCTION

The Internet has become more unavoidable entity in normal human life because the public use the Internet for various purposes and for everything. We have wireless services, which are provided by different service providers and could be accessed through the Internet. It is not necessary that the service has to be provided in a single server, but may be available at various nodes on the network. Each service has its own definition, design and requirement. In order to access the service the user has to provide the exact information with correct protocol. Otherwise, the service could not be completed properly. This may be called as malicious access or intrusion. For example, in a login service of user mail, the login service requires user name and password as two entities and this is the protocol of the login service. If the user provides the exact user name and password, then the user will be allowed to access the service and there is no problem otherwise the incorrect password can be considered as malicious act or access or threat.

There are many services available in different networks around the world that can be accessed by any user in the world through the Internet independent of their location. The user request generated from one location will be passing through various networks and nodes before it reaches the service point. Wherever the service available, the request produced from a
source node has to access the service point by entering into the network only through the router. The routers of the network have enforced with the firewalls that stop and restrict the illegal access of services and monitors the flow of packets for malicious threats. The logic available in the firewall plays an important role for the detection of intrusion and the performance of the network.

The security enforcement is the important factor for the resources and services available in the network. The efficiency and the performance of the overall network highly depend on the security measures on the network because the throughput on the network is based upon the security enforcement. The number of request of service completed successfully shows the throughput of the network and the frequency of identifying the malicious packet helps to increase the service availability of the network. Generally, there exist two different attacks, which can be classified as connection based and packet based attacks. In connection based attack, a malicious node keeps more connections and never used for data communication. This spoils the throughput of the server, and the genuine nodes will not get a connection to the server. Similarly, there are some nodes, which send an enormous amount of packets or data more than the allocated size. This generates congestion in the service point and increases the drop ratio and spoils the network performance. This is called packet flooding, and further these can be classified in many ways.

For example, any user within the organization may have distinct patterns of accessing the service available in their own network. Like an employee from audit department and account department has different behavior in accessing the company accounts and other accounts related to the cheque clearance and amount deposits. Similarly, the management people have a different kind of access behavior in accessing these accounts.
Moreover in any organization the worker’s access pattern of accessing the service will vary according to time windows. These patterns can be used to perform intrusion detection with the available logs of services.

The network is composed of a variety of services and has different constraints to access the service. There are different services available in different layers of the network and we considered about the service accessibility in the application layer where the service providers of any network provide various services. The services provided by service provider can be accessed by any user who is authorized, and every service has its own protocol.

For example, if we consider a login service, it has its own protocol where the service needs input of two values like username and password. Again, the values of username and the password can be of alphanumeric with allowed special characters. To access the login service the user has to provide two values, which match with the existing protocol otherwise the service cannot be accessed. Likely any service available in the network can be accessed by providing the exact input parameter with values.

Any user on the private network has a different behavioral approach of accessing the services which can be called as an access pattern. We consider different groups of users working for an organization. According to the profile of their work, a person working under stores department has a different pattern of accessing stock availability service than the user of marketing department or the top-level management.

The stock person has to visit the available stocks of any product to keep the minimum availability of any product in the organization. So, he will be frequently accessing the services in all his days of work. However, the marketing peoples will access the stock at a particular time once in a day.
The marketing person has to check out the possibility of immediate delivery of goods by accessing the service. So that the marketing person has less frequency of accessing the service than the store person.

Furthermore, the production manager or the general manager of the production company has to visit the stock available at the store for clear production planning. Similarly, many peoples of the manufacturing company will access the service frequently to know the status of the stock. But all the access are behind the frequency of the store person who will check the stock frequently and update the information at each production period.

Similarly the network services also have various range of access between different users of the network. Any network service has a provider who designs the protocol for the network service and the service can be accessible through the network, but the service has a unique protocol to access them. The protocol has to be followed to get access the service.

For example, if there is a service named product enquiry, which returns the information about the particular product like, the total stock available, and its prescription saying the product id, product name, size, weight and so on. All this information can be obtained by providing a single product id. The product id may be a combination of alphanumeric without any special character. To access this service the user has to provide valid product id so that the user will get the exact information about the product.

Similarly, there may be N number of services available in the network, but every service has different access pattern performed by different users. There are many attributes have to be considered to identify or access the service. All the services will not be accessed by a user at all the times. For example, a person may not check his account balance all the time in a working day. He may check the balance at the starting of the day or at the
end of the day. So there is no chance of access the same service at all the time window.

We can split the overall time of a day into N small units. For example, the overall time of the day may be split into 24 pieces so that we get 24 time windows. In that 24 time window, a user may access the balance enquiry service at the 10th window or in the 18th window. But the frequency of accessing the same service, in particular, time window will be less. This ideology can be applied to identify the malicious request. We have plans to adapt this approach to perform intrusion detection.

The time window can be created in different ways, which help to accurately identify the attacks. Time windows can be created in an hour basis, then we get 24 time windows for a single day. Similarly, the whole time window can be split into many pieces by splitting them based on minutes or seconds. If we increase the number of windows, then the value of traces obtained will be less.

**Table 4.1 Time window of a year**

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
</table>

The Table 4.1, displays the splitting the whole year into 12, as each month is considered as a window. The whole request comes in a single month is considered as a bundle and could be used for identifying the denial of service attacks.

**Table 4.2 Time window of week**

<table>
<thead>
<tr>
<th>Sun</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
</table>
The Table 4.2 displays the split up of the whole week where every single day is considered as a time slot. The packets or logs produced at each time window will be considered as different.

**Table 4.3 Split up for a day**

<table>
<thead>
<tr>
<th>Morning</th>
<th>Afternoon</th>
<th>Evening</th>
</tr>
</thead>
</table>

The Table 4.3 displays the split up of a single day. It has been split into three parts, namely morning, afternoon, forenoon.

Similarly, the logs produced from the traffic can be split into different ways, which are mentioned above. For each time window separated, the logs about a single service can be split into a number of groups according to the number of slots. From the split, we can identify a unique pattern of access, which shows the behavioral pattern of the service access.

This process can be further extended to pattern of each service, and we can produce a separate pattern of access to a single service. This can be performed for each service available.

The general features used to detect malicious packets are Source IP, Destination IP, Source port, and Destination port and TTL values, but we extend the feature set to time variant details like the number of connections at each time window, number of bytes received from each time window. This improves the performance of attack detection. The ultimate aim of our proposed method is to identify the network from where the attacks are initiated and stop serving them. There may be many nodes, which are supporting the attacks, so that the origin of the request could not be identified easier. We propose a new activity pattern based immune system for intrusion detection.
4.3 METHODS EXPLORED

For the immune system, various researchers proposed many methodologies and they each have their own merits and demerits. Now, we discuss few existing approaches here.

Network intrusion detection by an artificial immune system (Junyuan Shen 2011), is proposed which works based on feature selection and quantization algorithms. The feature selection algorithm uses various functions to select the features and the scheme has been tested with a different data set. The evaluation has been performed using KDD CUP 99 dataset and the method has produced higher accuracy result in intrusion detection.

Host Based intrusion Detection system presented by Vokorokos (2011) is an intrusion detection system which informs system administrator about potential intrusion incidence in a system. The designed architecture employs statistical method of data evaluation that allows detection based on the knowledge of user activity deviation in the computer system from the learned profile which representing the standard user behavior.

Macia Perez (2012) proposed a Network Intrusion Detection System Embedded on a Smart Sensor. In this system the NIDS is uploaded with a smart sensor device. This approach has been specified for service orient architecture. The method uses sensor method and NIDS functions to perform intrusion detection.

An Activity Pattern Based Wireless Intrusion Detection System has been proposed by Haldar (2012) which uses the activity patterns of the registered user and others. The method maintains the activity patterns of the users which are generated by accessing the service of the network. Also the method extracts various features of the service access and uses a threshold
value which is derived from the access records. The method identifies the intrusion detection when the threshold value exceeds the allowed limits.

Elhadi et al (2013) proposed and implemented a new intrusion-detection system named Enhanced Adaptive ACKnowledgment (EAACK) specially designed for MANETs. EAACK consists of three major parts, namely, ACK, secure ACK (S-ACK), and misbehavior report authentication (MRA). In order to distinguish different packet types in different schemes, it included a 2-b packet header in EAACK.

Security and cooperation in wireless network are discussed in Moti Geva et al (2014) by considering the fact that most routing protocols in MANETs assume that every node in the network behaves cooperatively with other nodes and presumably not malicious, attackers can easily compromise MANETs by inserting malicious or non cooperative nodes into the network.

An Artificial Immune System Based on Holland's Classifier as Network Intrusion Detection (Randiranosolo 2012), proposed as a new method for network intrusion detection, which is not aimed to provide a comparative study, but to give more understanding on the feasibility of combining Artificial Immune System and Hollandâs Classifier to detect network intrusion. This new Artificial Immune System, named AIS-CS, can attain higher than 90% intrusion detection with a false negative percentage below 10% and a fairly low false positive rate on a network composed of 50 regular nodes and 50 intruders.

All the above discussed approaches have the problem of accuracy in detection of intrusion. To overcome the problem of false negative ratio in intrusion detection, we propose an activity pattern based method to perform intrusion detection by considering time variant features.
4.4 PROPOSED METHOD

The heterogeneous activity pattern based approach has different functional components where the captured input packet feature is extracted first and then from the extracted feature the activity pattern is generated using the logs. Finally, with the activity pattern generated the intrusion detection is performed.

The proposed network forensics approach receives the incoming packet through the network interfaces and it extracts the features like IP Address, port, list of hops traversed, data length, payload details, TTL, time window based connections, data transfer and so on. The extracted feature is used to generate the activity pattern which is a heterogeneous one using which legitimate weight is computed for each packet. For any packet received on the network interface, the gateway of the network will compute the legitimate weight for both registered and un-registered user. The packet has to clear the trust code enforced by the security system. Once it clears the trust code, the packet will be sent to the destined port to be handled by the service.

![Figure 4.1 Network immune system based on Heterogeneous activity pattern architecture](image-url)
4.4.1 Feature Extraction

At the feature extraction phase, we extract various features of packets and each has distinct importance in the detection of intrusion. For example, there are some services where the packet has to be delivered on time, in this case the ttl of the packet has to be considered. In case of hop count the packet has to traverse within the limit of hop count. In case of pay load, the packet has to carry less than the limit of allowed pay load and in case of connection oriented network communication the number of connections hold by the particular process has to be considered. Here the network data packet received at the network interface port is handled at the feature extraction phase. The features of the packet cannot be extracted directly so that the packet will be converted in the form of Internet packet (IP) to get the address and port of the packet. Then the other packet features like payload, number of connections, payload in the form of streams and TTL values are extracted. All the features extracted are converted in the form of feature vector which is used for network forensic analysis.

Algorithm

**Input** : raw packet

**Output** : feature vector, Access History

**Step 1** : Read input packet.

**Step 2** : Perform conversion of raw packet into IP packets.

**Step 3** : Read Source IP.

**Step 4** : Read Source Port.

**Step 5** : Read TTL value.

**Step 6** : Compute payload details from packet frame.

**Step 7** : Generate the feature vector.

**Step 8** : Add the feature into the access history.
4.4.2 Activity Pattern Generation

Activity pattern is nothing but the style or behavior type identification of each process which sends packets to the network or services. For example, there are processes which sends packets through specific two or three routes and it has the style that the packet from that particular source always follows through a single path from the set of paths. And few nodes will send some limited packets with limited or particular type of data payload which has a style and also the nodes will follow the range in hops. We generate the activity pattern using extracted feature and previous access details available in the access history. We split the service log into different time windows and for each time window for a distinct service access history a pattern is computed based on the features of the log.

Algorithm

Step 1 : Read Input feature and access history.
Step 2 : Split overall time into different time window.
Step 3 : For each time window

Identify the service logs produced in the particular time window.

Compute total connections obtained.

Compute total number of packets being received.

Compute total payload produced in the time window.

Compute activity pattern by identifying the other features from logs.

end.
### Table 4.4 Sample activity pattern

<table>
<thead>
<tr>
<th>Time</th>
<th>Source IP address</th>
<th>Source port</th>
<th>Data stream</th>
<th>Number of connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 A.M</td>
<td>192.168.1.2</td>
<td>9898</td>
<td>12</td>
<td>148</td>
</tr>
<tr>
<td>5 A.M</td>
<td>192.168.1.2</td>
<td>9898</td>
<td>8</td>
<td>187</td>
</tr>
<tr>
<td>10 A.M</td>
<td>192.168.1.2</td>
<td>9898</td>
<td>10</td>
<td>128</td>
</tr>
<tr>
<td>3 P.M</td>
<td>192.168.1.2</td>
<td>9898</td>
<td>12</td>
<td>156</td>
</tr>
</tbody>
</table>

The above example shows the pattern generated from the network log and access history of users. It shows that there are four different windows exist and each has a different number of connections and different number or range of stream transfer. Using these details the newly arrived packet feature can be identified and a decision could be taken to perform intrusion detection.

#### 4.4.3 Intrusion Detection

From the extracted feature and the activity pattern of the packet arrived, the intrusion detection system reads the service history from the database. The service history has logs about previous malicious packets and their patterns, the proposed method compares with each of the patterns from the log.

**Algorithm**

- **Step 1**: Read feature vector, activity pattern and access history.
- **Step 2**: Split the overall time into different time window.
- **Step 3**: For each time window
  - Compute average number of connections requested.
  - Compute average stream produced.
- **Step 4**: If the stream greater than the threshold
  - Find out all the records of access from that source.
Compute average hop count (Ahp)
Compute average payload(Apl)
Compute average TTL(Attl)

**Step 5**: Compute legitimate weight.

\[ Lwp = \frac{\text{hop count}}{Ahp} \times \frac{\text{Payload}}{Apl} \times \frac{\text{TTL}}{Attl} \]

**Step 6**: If (Legitimate weight \( \geq \) Legitimate threshold)

- Forward packet to service handling

Else

- Reject the packet and add the packet feature to the malicious history.

End.

### 4.5 EXPERIMENTAL RESULTS AND PERFORMANCE COMPARISON

The proposed activity pattern based approach has been implemented and tested using the sonaNet Test bed (Figure 4.2). The test bed has number of routers and servers with large number of nodes. This test bed is used in the further research work. The test bed captures the incoming packets belongs to various services and generates number of logs according to the features of the incoming packet.

![Figure 4.2 SonaNET test bed](image_url)
4.5.1 Results of the Proposed Work

The extracted features of a packet have been converted into a pattern like purchase pattern in data mining. The pattern represents the activity of any user which represent the hop count of the packet, ttl value achieved by the packet, the payload size of the packet, hop counts crossed to reach the destination and so on. All of these features are used to identify the malicious packet. Once all this has been collected the method computes a legitimate weight for each of the packet using other features of the packet captured.

Figure 4.3 Snapshot of client interface in sending messages

Figure 4.4 Snapshot of data packets received at the destination
The Figure 4.6 shows the snapshot of the classification of features which is used for finding malicious packets.
The Figure 4.7 shows the snapshot of user interface which is captured at the time of sending the messages. Here we consider each process running on distinct port on a single machine as different nodes to form the network topology. So in this approach the packets will be routed through a set of ports instead of a set of IP addresses.

This screen shows two input, which specifies the message that is to be sent and the destination location address. Radio buttons are available to indicate whether the entered message has to be sent as a malicious packet or normal packet. After providing values for all essential parameters the send option is selected.

Figure 4.8 Snapshot of data packets received at the destination
The Figure 4.8 shows the snapshot of the destination interface where the packet has been received. The screen shot shows the details of the received packet such as source address, destination address, message, received time and Message type.

![Figure 4.8](image1.png)

**Figure 4.9** Snapshot of route reply and computed routes

The Figure 4.9 shows the snapshot of computed routes which the packets take to reach the destination and the route reply received from neighbor nodes.

![Figure 4.9](image2.png)

**Figure 4.10** Snapshot of computed values
The Figure 4.10 shows the snapshot which contains the details about the computed measures in this approach. We have computed average hop count, average connections, average payload, average TTL to compute the legitimate weight of the packet.

Figure 4.11  Result of heterogeneous activity pattern based approach

The Figure 4.11 shows the result of denial of service attacks detection where the two packets are shown as malicious because the packets has the legitimate weight of 38.0 which is less than the legitimate threshold value of 85.0. The legitimate threshold is not a constant factor and is a dynamic one which is computed based on the size of log or history available.

The Figure 4.12 shows the snapshot of the data set maintained by the proposed method which is produced by extracting the features of the incoming packet. This dataset is being used in the remaining work.
The method has been used all these features to compute the legitimate weight by splitting the record as different time window and we generated the activity pattern to compute the legitimate weight.

The method also displays the result of denial of service attack detection where the method has computed the legitimate value for each of the packets and the computed weight will be used to identify the malicious node to perform denial of service attack detection.

4.5.2 Performance Analysis with Existing System

The method has been evaluated with various time window results and the method produced efficient results in identifying the intrusion detection. The activity pattern based approach handles both the connection based and packet based attacks. By extracting the features of the packet being
captured and the method has generated the activity pattern for each user at each time window. Finally the method computes legitimate weight based on the computed average hop count, average payload and average ttl value. The legitimate weight decides the access flag for the incoming packet. The method has higher accuracy of intrusion detection and has produced efficient results.

Figure 4.13 The frequency of detection of malicious packet

Figure 4.13 shows the result of the proposed system in finding malicious packet, and if there are 100 packets which are malicious arrived on time, then the graph shows the frequency of detection of malicious packet. It is very clear that the proposed system identifies the more malicious packet compared to other host based and activity pattern based intrusion detection systems.

The Figure 4.13 also displays that the method has produced maximum detection accuracy than the other approaches. The efficiency of the algorithm is highly based on the number of samples. An algorithm can be concluded as efficient one, only if it can produce efficient results with less number of patterns used.
In our approach, the method has used time variant pattern mining technique to identify the denial of service attacks. The input data set or the traffic log is grouped according to various time windows and at each time window the service access of each service is grouped which will give you the exact pattern of access at each time window.

From generated pattern, we can generate the current pattern of access and verify the genuine value of the access before providing access to the user. Once we have identified the pattern or behavior of user in accessing the service we can easily identify the malicious requests so that the detection rate can be improved and the frequency of attack can be reduced.

Similarly from the computed pattern of accessing the network service, we can identify the pattern of malicious user so that they can be stopped before anyone access that service. In this approach we consider various attributes and methods to identify the malicious request and stop denial of service attacks.

![Time Complexity](image)

**Figure 4.14** The time complexity of the proposed system
Figure 4.14 shows the time complexity of the proposed system compared to other methodologies. It shows clearly that the proposed system takes only a little time compared to other methods for different number of packets. The other methods take more time to compare to the proposed system to analyze and detect the intrusion for number of packets.

4.5.3 Accuracy Analysis of Proposed System

The accuracy of intrusion detection can be measured with various parameters like precision, recall, entropy and so on. Here we consider the entropy measure by which the classification accuracy can be evaluated.

The Precision, Recall and Entropy has been computed for the proposed approach to evaluate how the classification has been performed with the following assumptions.

NNN – Number of classifications performed Normal as Normal
NNA – Number of classifications performed Normal as Attack
NAN – Number of classifications performed Attack as Normal
NAA – Number of classifications performed Attack as Attack.

The precision is computed as follows:

\[ \text{Precision} = \frac{NAA}{NAA + NNA} \]

The precision value shows the relevancy of packet classification.

Similarly the Recall measure is computed as follows:

\[ \text{Recall} = \frac{NAA}{NAA + NNN} \]
The recall value shows the relevancy of classification towards the condition provided.

The entropy measure is computed as follows:

\[ \text{Entropy } E = -\sum NAA \times \log(NAA) \] where the entropy measure shows the probabilistic measure of classification accuracy.

Similarly, we have evaluated the proposed method by computing Sibson distance which shows the closeness of patterns using which classification is performed. The basic Sibson distance is computed as follows:

\[ G(x,y) = \sum_{i=1}^{N} W \times f(x,y) \] Where \( W \) specifies the weight of packet which is computed using the function with the value of data packet at the point \( x \) and \( y \).

The accuracy of classification and intrusion detection has been evaluated using various parameters which are depicted in Table 4.5.

**Table 4.5 Classification accuracy**

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.96</td>
<td>0.92</td>
<td>0.93</td>
</tr>
</tbody>
</table>

The proposed approach performs intrusion detection based on legitimate weight, which is computed using the packet features. The decision is based on the weight and threshold. The threshold value of the proposed approach will be variable based on the flow of the packet and the signature of packets. In case of voice packets the threshold will be allowed up to 0.7 whereas for normal data packet the threshold will be allowed only up to 0.5 and so on.
4.6 CHAPTER CONCLUSION

The proposed heterogeneous time variant activity pattern based network immune system is to find the malicious packet and mitigation attacks. Unlike other methodologies, the new system uses various parameters to find the malicious packets. The proposed method finds the malicious nodes in an efficient way and increases the frequency of intrusion detection and reduces the threat. The proposed method increases the overall throughput of the network. The proposed method can be further improved by adapting various other features at different layers.