Artificial Intelligence based Intrusion Detection System to detect Flooding Attack in VANETs

Thesis *submitted in partial fulfillment of the requirements for the award of degree of*

Master of Engineering

in

Information Security

Submitted By

Mannat Jot Singh Aneja

(801433016)

Under the supervision of:

Ms. Tarunpreet Bhatia

Lecturer

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

July 2016

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*Artificial Intelligence based Intrusion Detection System to detect Flooding Attack in VANETs*", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Information Security* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Ms. Tarunpreet Bhatia* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

(Mannat Jot Singh Aneja

This is to certify that the above statement made by the candidate is correct and true to th best of my knowledge.

(Ms Tarunpreet Bha

Lecturer, CSED

Countersigned by

(Dr. Maninder Singh)

Head

Computer Science and Engineering Department

Thapar University

Patiala

(Dr. S. S. Bhatia)

Dean (Academic Af

Thapar University

Patiala

ACKNOWLEDGEMENT

I have been waiting long for this moment to acknowledge all those who helped, motivated and believed in me during my thesis. I would like to thank my mentor **Ms Tarunpreet Bhatia** who gave me guidance at each and every step, listened to my doubts and gave her valuable time for my success. I thank you for your extremely helpful suggestions and motivation at each step.

I express my profound gratitude to our director **Prof. Prakash Gopalan** for his inspiration and providing research environment at our campus. I extend my thanks to **Dr. Maninder Singh** Head, Computer Science and Engineering Department, Thapar University for constant guidance during my research work and imparting self confidence in me.

I also want to thanks to my research committee members, non-teaching staff of the institute for their help and support and my fellow M.E. scholars Mr. Ishan Singh Thakur and Mr. Manish Kumar for providing help during my thesis documentation.

At last, I also want to thanks to my family and friends for their emotional support

ABSTRACT

Vehicular Ad hoc Networks (VANETs) are classes of ad hoc network that provides communication among various vehicles and roadside units. VANETs are decentralized and due to this, these are susceptible to many security attacks. These attacks mainly affects the five requirements- the availability, confidentiality, integrity, non-repudiation and authenticity of the system. Intrusion Detection System (IDS) are used to combat these attacks. Flooding attack is one of the major security threats to VANET environment. The current thesis proposes a hybrid Intrusion Detection System which improves accuracy and other performance metrics using Artificial Neural Networks as classification engine and Genetic algorithm as optimization engine for feature subset selection. These performance metrics has been calculated in two scenarios namely misuse and anomaly. Various performance metrics are calculated and compared with other researchers work. The results obtained indicate high accuracy and precision and negligible false alarm rate. These performance metric are used to evaluate intrusion system and compared with other existing algorithms. The classifier works well for multiple malicious nodes. Apart from machine learning techniques, effect of the network parameters like throughput and packet delivery ratio are observed.

Keywords: RREQ Flooding, Intrusion Detection System, Artificial Neural Network, Genetic Algorithm, Feature Subset Selection, VANETs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
AODV	Ad hoc on Demand Distance Vector
AU	Application Unit
BPNN	Back Propagation Neural Network
CA	Certificate Authority
DDoS	Distributed Denial of Service
DoS	Denial of Service
DR	Detection Rate
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GA	Genetic Algorithm
IDS	Intrusion Detection System
MANET	Mobile ad hoc networks
NAM	Network Animator
OBU	On Board Unit
RREP	Route Reply
RREQ	Route Request
RSU	Road Side Unit
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
V2I	Vehicular to Infrastructure
V2V	Vehicular to Vehicular
VANET	Vehicular ad hoc networks

CHAPTER-1

INTRODUCTION

1.1 Introduction to Vehicular ad hoc networks (VANETs)

Vehicular ad hoc networks (VANETs) are special category of MANETs. These are almost MANETs but differ in its movement. Like MANETs, VANETs are also self organized, lacks the infrastructure, mobile in nature. The only difference lies in its movement. In MANETs the node can move randomly whereas in VANETs the node does not follow random movement. The nodes simulate like vehicle and move on the road in straight line direction as that of road. With the discovery of VANETs concept, the new age of vehicles has changed. Nowadays the vehicles are equipped with lot of sensors. Due to increase in population, there has been exponential increase in number of vehicles. This increase in vehicles tends to increase the chance of road accidents. According to the survey, there has been 12 lakhs life are lost daily worldwide [1]. With the advancement of technology the life of user has become easy but it has few shortcomings as well.

The new age of vehicles is network of sensors. The vehicles are also known as computers running on road. As mentioned above that there has been huge amount of life being lost on roads. So with increase in road traffic, there needs to be increase in road safety as well. We need to have a mechanism by virtue of which the vehicles can be made smart enough so that they are able to handle the road safety at their own. This concept was the laid under VANETs. VANETs allow the vehicles to communicate to each other and come to know about if there is any diversion ahead or any other information required. So VANETs support communication among the vehicle. Actually VANETs support four types of communication [2]:

- **Vehicle to vehicle (V2V):** This communication facilitates the interaction in between the vehicles to ensure safety on road. The communication could be information exchange about some accident ahead or diversion.
- **Vehicle to Infrastructure (V2I):** This communication facilitates the interaction to take place between vehicles and road side units. The communication could be information exchange like information from road side garage or filling station.

- **Hybrid communication:** This communication is heuristic communication of above communications. It includes inter vehicle and vehicle and infrastructure communication.
- **Intra-vehicular communication:** This communication is vital for the vehicle as it comes to know the information about itself like fatigue detection of driver, GPS navigation etc

The following figure shows the communication in VANET environment.

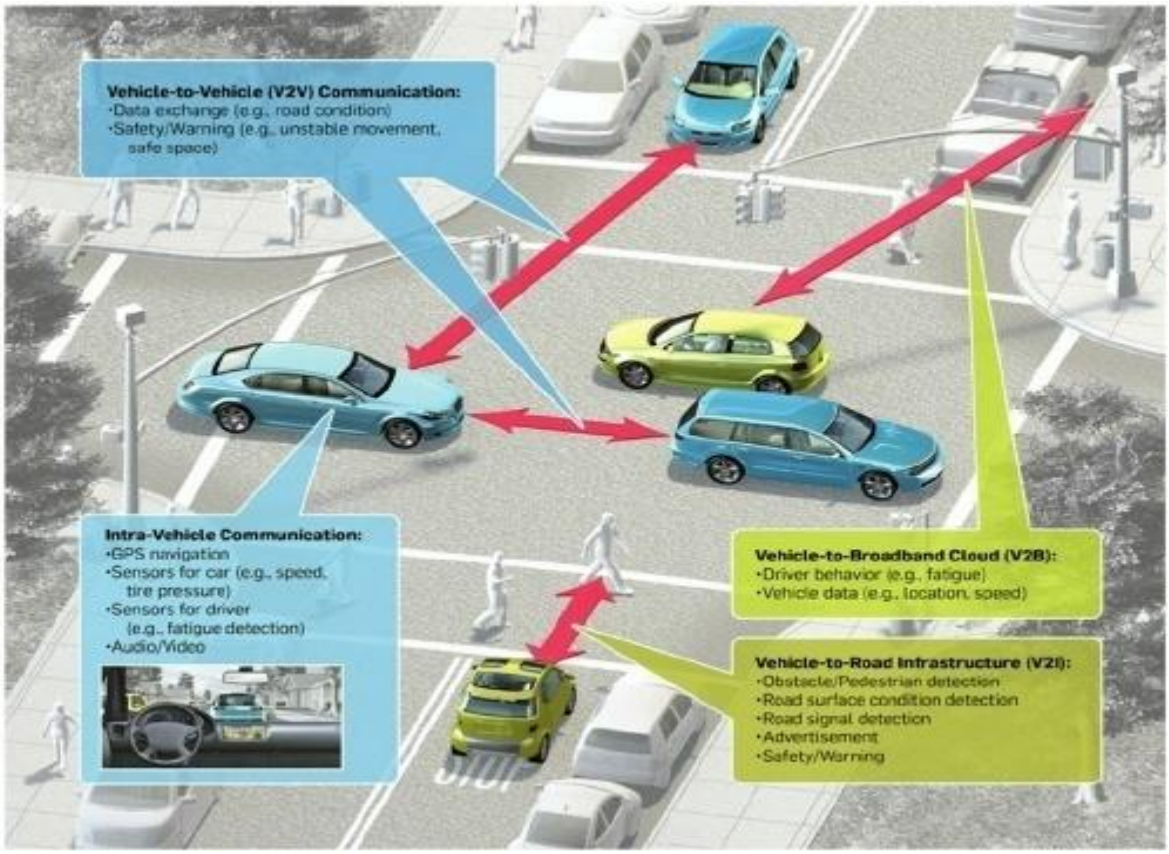


Fig 1.1 Communications in VANETs [3]

1.2 Motivation of Thesis

Vehicular ad hoc networks are the ad hoc networks by virtue of which the vehicles communicate with each other. These messages could be emergency messages like indicating the vehicle at the back to slow down as there is traffic jam ahead or sending the information about changing the lane due to accident ahead. Apart from these, there could be exchange of media among the vehicles. So as we move towards modernization the driverless car becomes the trend. However even after having so many applications, it has some shortcomings as well. Since the structure of this

network is decentralized and lacks infrastructure, it is vulnerable to many security threats. The one of the most serious attack is flooding in which the attacker tries to flood the network with request for resources to the node which does not even exist, thus making the channel unavailable to be used by legitimate node. The resource which were to be used by legitimate node are unavailable as they are busy serving the intruder node. This is denial of the service (DoS). Research is continuous process, so it is very crucial to have the research focus on flooding attack as it forms basis of various other attacks like distributed denial of service (DDos). The packet drop attack also has the similar concept.

Security is indispensable component in any industry or in any field. We need security as it gives sense of surety of wellness. The vehicular networks also need to be secured. The ill effects could also lead to loss of life. There has been a lot of research in the field of security but still there are lots of demerits, so there is a need to have the research go on in the field of network security especially vehicular networks as driverless cars is trending to become the hot topic in near future.

1.3 Thesis Structure

The thesis is organized into 7 chapters. Chapter 1 discusses about VANETs and the motivation for choosing this topic. Chapter 2 gives in detail about applications and security issues in VANETs, its solutions using cryptography and IDS and also discusses classification of IDS in details. Chapter 3 discusses about various researcher's work in the field of security in VANETs. Chapter 4 gives detail about the problem that is considered in this thesis, the objectives aimed to be accomplished and so on. Chapter 5 discusses the proposed algorithm and implementation and also gives details about the simulators used. Chapter 6 mentions result on the basis of various performance metrics and later compare our results with already existing protocols. Lastly Chapter 7 provides the conclusion and scope of improvement in terms of future scope of our thesis

.

CHAPTER-2

SECURITY IN VANETs

Now days due to exponential increase in vehicles on road, the number of road hazards has grown. To provide secure and reliable driving environment we need a mechanism by virtue of which the vehicles can interact with each other. This is accomplished by Vehicular ad hoc networks (VANETs). VANET is subgroup of MANETs

2.1 Architecture of VANETs

The VANET architecture is composed of three domains: in vehicle domain, ad hoc domain and infrastructure domain.

- **In vehicle domain**: This domain comprises of On Board Unit (OBU) and Application Unit (AU). The AU should be atleast one. However it can contain more than one AU. The network between them is generally wired, however it can be wireless as well
- **Ad hoc domain**: This domain is composed of vehicles with OBU and RSU (road side units).
- **Infrastructure domain:** It is composed of RSU and HS (hot spots). These infrastructure domain facilitate the OBU to communicate with Internet. Apart from this OBU can also communicate using GPS, GPRS and other radio networks. The following figure shows these domains.

The communication in VANETs takes place with the help of Dedicated Short Range Communication (DSRC) standard. This standard supports IEEE 802.11p for wireless communication. This DSRC 8012.11p is also known as Wireless Access in Vehicular environment (WAVE). In VANETs the OBU is responsible for interacting with outside network which includes other vehicles and roadside unit infrastructure

2.2 Characteristics of VANETs

The characteristics of VANET are distinct as compared to the other ad hoc networks. The system is efficient but it has serious issues as well. We first present with the features followed by the issues it might have.

- **Dynamic in nature**: Due to variable amount of speed of the vehicle and exponential increase in vehicle, the vehicle might join a particular network and might leave and join another network. It makes VANET environment dynamic.
- **Change in information:** Due to huge number of movement in VANET environment, there is change in information at very instant. The information a vehicle receives is from either a vehicle or RSU. Hence there is huge amount of change in information in the network.
- **Efficient:** Since VANET uses sensors to communicate with vehicles and infrastructure, the sensors keep utilizing the energy. Since VANET uses short range communication, the energy is not consumed that much thus making the whole environment efficient.
- **Frequent Disconnection:** The dynamic nature of VANETs results in frequent disconnections as the vehicle jumps from one network to another in not time.
- **Unpredictable:** The frequent change in network topology and irregular speed of vehicle makes the prediction of vehicle a bit difficult. Since the prediction is difficult it is equally difficult to mischief a vehicle.
- **Better physical security:** It is hard to make the nodes promiscuous physically as it has pool of sensors attached to it which improves the physical security.

2.3 Applications of VANETs

There has been ample amount of applications available in VANETs. These could be the safety applications, commercial applications or based on convenience applications.

- **Safety Applications:** The safety application is responsible for the applications that enable the safety of the vehicle and the people in vehicle. Some of them are listed below:
  - **Real time traffic:** The RSU can be made to know the traffic analysis ahead. This includes any traffic jams or roadside accidents or any halt ahead. With this knowledge the user can take diversion and avoid the congestion.
  - **Post Crash notification:** The vehicles which indulge in accident can send this message to other vehicles by virtue of which they can take decision about the route. Moreover this vehicle can contact some road side garage as well.

- **Collision Avoidance:** The collision can be avoided if the user gets an alert in time.
- **Traffic Signal Violation:** The RSU can be used to broadcast the messages about any traffic violation by any other vehicle.
- **Fatigue Detection Warning:** These messages alerts the driver if he/she is falling asleep.

- **Commercial Applications:** The commercial applications include internet services and value added services. These are listed as:
  - **Internet:** The vehicles can connect to internet through RSU if and only if RSU acts as router.
  - **VAS:** The value added service advertisements can be made by roadside restaurants or any other outlets to the vehicles in range. This is smart way of providing the offers and doing the business.
- **Convenience Application:** These applications are responsible to provide comfort to the drivers. Following are some of the applications listed
  - **Parking availability:** The driver can get to know about the parking situation ahead of reaching the place which eases the life of driver especially in big cities.
  - **Toll collection:** The electronic toll collection can be done which saves both the cost and time of driver. The cost is saved by saving the petrol which gets wasted waiting at toll plaza. The OBU can be used for this purpose.

2.4 Challenging Issues in VANETs

There are few challenging issues that arose with the advancement in VANETs. Some of them are with respect to security while some are with respect to driver. Some of them can impact economically while some of them can have physical impact. A few of these have been discussed in this section.

- **High Mobility:** As the vehicles are mobile in Vehicular Ad Hoc Networks, they remain in particular network for a very short duration. This results in disruption in the communication. There could be chance that those two vehicles may never meet again. It also becomes impossible to guess or to know the exact location of vehicle.

- **Fault Tolerance:** As the network topology changes rapidly, there are lots of vehicles which enter and leave the network at particular instant. During the communication if a vehicle leaves the network an alternative route has to be established by the routing protocol. This requires exchange of heavy amount of information.
- **Security:** Security is indispensable aspect in VANETs. The vehicles communicate with each other over a wireless medium. The security is essential for the same as without the security if a malicious vehicle enters and gain access to the information it is not supposed to know. This could result in misuse of information.
- **Scalability:** The network is growing at rapid pace and is still growing on every day. With the exponential increase in number of vehicles leading to huge network on one side there is acute shortage of uniform standard that monitors these networks for example the DSRC in North America differs from that in Europe.
- **Real Time Transmission**: If there is a sudden incident on road the driver needs to inform about the emergency situation. These type of critical and life saving information has to be transmitted without delay.

2.5 Routing In VANETs

There are various routing protocols in VANETs. These are topology based, position based, geocast based, broadcast based and cluster based [13]. Following is the tree view of the routing protocols of VANETs:

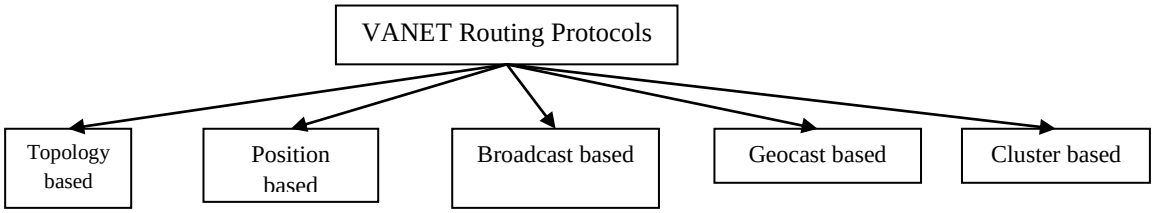


Fig 2.1 Routing Protocols in VANETs

- **Topology based:** These routing protocols are further divided into reactive and proactive. In proactive routing protocols the route is not prefixed. The route is established on the go. In reactive routing protocol, the route is

maintained for currently in use. Some of the reactive routing protocols are AODV, DSR and TORA.

- **Position based:** These protocols choose the position information for selecting next forwarding hop. There is no need to create global route from source to destination. Delay tolerant networks belong to this category
- **Broadcast based:** These routing protocols broadcasts the information in the network. These are essential in VANETs for sending emergency or any other critical information
- **Geocasting based:** These routing protocols sends the packet from source to all other nodes among the geographic region. These are similar to zone based routing
- **Cluster based routing:** In this routing collection of some nodes based on common traits lie in one cluster. The cluster head is responsible for sending the packet among other nodes.

Following is the summary of these routing protocols in form of table

Table 2.1 Comparison of routing protocols in VANETs

Protocol form	Topology based	Position based	Cluster based	Geocast based	Broadcast based
Forwarding	Wireless multi hop	Heuristic	Same as topology based	Same as topology based	Same as topology based
Recovery technique	Multi hop forwarding	Carry and forward	Same as position based	Flooding	Same as position based
Map requirement	No	No	Yes	No	No
Virtual infrastructure requirement	No	No	Yes	No	No
Scenario	Urban	Urban	Urban	Highway	Highway

2.5.1 AODV Routing Protocol

It is a reactive routing protocol. The route is established when there is a need to send the packets. This protocol is used for both wired as well as wireless channels. It has the ability of unicasting and multicasting routing. AODV uses two important messages: RREQ (route request) and RREP (route reply). The source node transmits data to destination node and if the destination node is not in its routing table, then the source node sends the RREQ message to its neighbours which is then forwarded to their neighbours until it reaches the destination. Upon reaching the destination, the

destination node sends the RREP message in the same path as that of RREQ message. Consider a source node S wants to transmit a packet to destination node D. Since D is not in range of S, S transmits the packet to its neighbouring nodes as Route Request (RREQ). The neigbouring node transmits the RREQ message to all its neighbours. This process is repeated until the packet reaches the destination. At the destination, Route Reply (RREP) message is prepared and is sent to the source as a reply to source. The path followed by RREP is reverse path of RREQ. This whole scenario is depicted in figure 2.2.

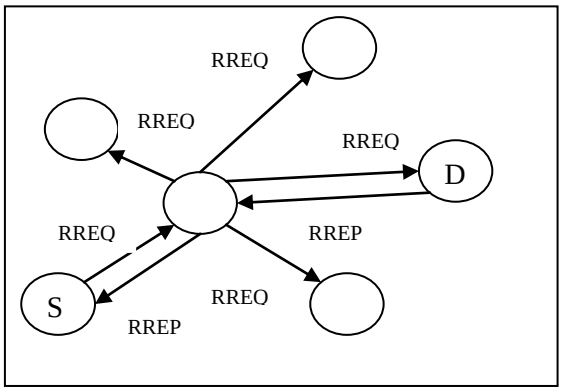


Fig 2.2 AODV Route Discovery

2.6 Adversaries in VANET Environment

- **Selfish Driver:** These drivers have an intention of greed. A selfish driver may distribute wrong information about traffic ahead or congestion on road. This leads to diversion of route by other vehicles. The selfish driver then in turn can place its vehicle with ease in parking area.
- **Malicious Driver:** These types of vehicular drivers try to cause the damage with the help of the applications available on the vehicular ad hoc networks. These drivers use the network resources for causing the damage.
- **Prankster:** These types of drivers do not have an intention to ruin the network or misuse the network resources or information. They just do it without knowing the ill effects it can have. For example they can tell the vehicle to slow down due to congestion and the other vehicle behind it to increase the speed and then enjoy but this could result in ill effects.

2.7 Security Requirements

Security is indispensable aspect in any area especially in VANETS. In previous section we discussed about the loopholes that exist. These loopholes can be exploited by an intruder or a malicious driver. So the security is absolute necessary in VANETS. The most important security requirement comprises of CIA triad. The CIA triad stands for Confidentiality, Integrity and Availability. Apart from these requirements there are two more requirements namely authentication and non-repudiation.

- **Confidentiality:** It aims to provide the guarantee that no malicious driver or vehicle is able to access the private information of any other vehicle or driver. The private information could be number plate details, license information or the vehicle status etc. This security requirement is established by using encryption and keys which have expiry time as well
- **Integrity:** This security requirement guarantees that the message is not changed from source to destination. Whatever the type or format of message is sent by sender the same message is being received by the receiver. There is not even a single minute change in message. This requirement is indispensable as it may lead to wrong or incomplete information. In some life saving applications this is very necessary requirement. The group signature enables this requirement.
- **Availability:** This security requirement means the network must be available all the time. As VANETS have time critical applications, the first most important requirement is the full time availability of ad hoc network. Sometimes the application requires VANETS to reply as early as possible. This lag in reply for particular request could make the service useless. To solve the availability problem group signature technique is proposed. The DOS and DDOS attack could affect if this security requirement is not established.
- **Non-Repudiation:** The non-repudiation means that the sender or receiver cannot say no to the messages being sent or received by them. The non repudiation means auditing. There could be information regarding accident or information regarding car parking. The sender later cannot decline the message. The bogus information attack could be launched if there would have been no security requirement like this. This requirement gives the guarantee

that intruder could be caught even after the attack is launched. In some cases non-repudiation is also known as auditability. This security requirement allows us to identify the attacker even after the attack has been launched.

- **Authentication:** This is utmost important requirement as it ensures the sender to prove that it is a legal sender. This is accomplished by public or private keys and CA. The sender sends the message with the key and certificate which is checked by the receiver. However signing the information makes the system complex. To make this task simple an efficient Elliptical Curve Cryptography is used.

2.8 Attacks in VANETs

There are various security attacks to which the VANET networks are vulnerable to. These attacks have huge impact not only on the network but this could also lead to loss of life as well. Following are the some of the security attacks which can be launched on VANETs.

- **Denial of Service Attack**: The Denial of Service (DoS) attack is the attack by virtue of which an ad-hoc network is unavailable. It can be launched either by flooding the network with unusual and useless request so that the resources of the network are kept in use and the legitimate request will not be able to have access to that particular resource. The other way of launching this attack is by crashing the communication channel [18].
- **Distributed Denial of Service**: This is a variant to the above attack in which more than one attacker tries to launch the Denial of Service attack on the victim node. In this attack the attack is launched with the usage of multiple computers and resources are acquired by multiple computers located at multiple positions as well. The main objective of this attack is to deny the availability as a security requirement.
- **Replay Attacks**: In this type of attack, the intruder replays the transmission of previous messages and tries to gain the access of the system and other resources available at the time of sending the message [18].
- **Sybil Attack**: In this type of attack duplicate nodes are formed using illegal identities and a node sends the information to different nodes using different identities. As a result different nodes have different impression about the

same node. This type of attack depends on how easy it is to form identities, and whether the system considers all the nodes similar or not. There are various techniques to combat this attack such as statistical and probability approach [18].

- **Alteration Attack**: This attack is launched when an intruder changes the existing data. This changed data is then forwarded to the network. The other way to launch this attack is delaying the information that has to be sent in the network [18].
- **Fabrication Attack**: In fabrication attack the attacker sends the untrue information into the network. The information could be wrong or the transmitter could claim it to be someone else [18]
- **Black hole attack**: This attack is formed when the node denies participating in the network. Another alternative could be that the node drops out of the network. In this attack. In this attack the whole data is sent to a node which is not existing in the network resulting in huge loss of data.
- **Malwares**: In VANETs malwares may lead to deviate from normal operation of the network. This might happen during the software update. This type of attack is generally caused by insider as an intruder.
- **Masquerading attack**: In this type of attack the attacker actively participates in the network. The attacker tries to pretend like other vehicles using false identity. Message fabrication, replay attack and alteration attacks could be used towards masquerading.
- **Tunneling attack**: The intruder tries to establish a network between two distant ad hoc networks using an extra channel between them. This channel is known as tunnel. The nodes in two distant networks have an impression of being neighbors and send the data through the tunnel. The attacker has an access to tunnel can misuse the data then [19]
- **ID Disclosure Attack**: In this attack, the ability of the node to let know its identity is exploited and as a result its location also becomes transparent to the whole network. The intruder sends malware to the neighbors of the target node. These malwares are replicating in nature and hence targets its neighbors. When the malware reaches the neighbor of the intruder, it notes the location of target node as well as its identity is noted. One of the solutions

proposed is to encrypt the confidential information of vehicle like number plate and location and later uses Public Key Infrastructure [19]

- **Node Impersonation Attack**: It is an attack in which a node sends fabricated message and gives an impression that it came from the originator. It can be launched in 2 ways either in form of Sybil attack or in form of false attribute possession. One of the methods to combat this attack by using trust value. The identity of vehicle is noticed by CA based on trust value threshold.
- **Sending Deceptive Messages Attack**: In this attack, a node intentionally sends fake messages to another node in the network in order to create havoc. Normally this type of attack is launched by a selfish driver which sends deceptive information about the traffic jam or road accident
- **Flooding Attack**: In this type of attack, the attacker tries to flood the network with request for resources to the node which does not even exist, thus making the channel unavailable to be used by legitimate node. This type of attack can be launched in various ways. The control packet flooding and data flooding are one of the common methods to launch flooding attack. It is a subset of Denial of Service attack. One of the way of control flooding is RREQ Flooding. According to RREQ Flooding attack, the attacker broadcast multiple number of RREQ messages to the node which does not exist in the network. Figure 2.4 shows RREQ flooding attack in which attacker sends multiple number of RREQ message to non-existing node by virtue of which the legitimate node source is unable to access the channel to destination node via attacker node.

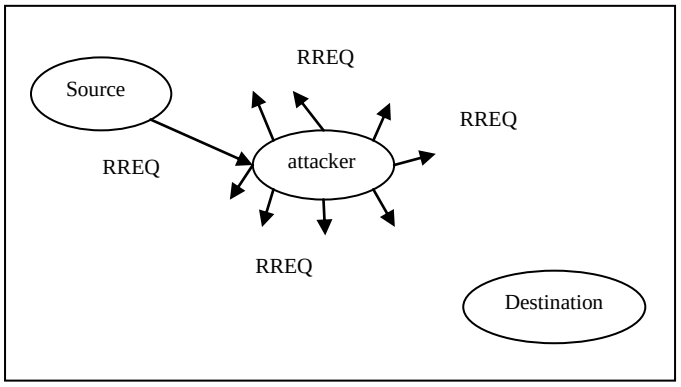


Fig 2.4: RREQ Flooding attack

- **Wormhole attack**: In this type of attack the two legitimate node which are not in each other's range wants to communicate through the tunnel. The intruder node lies in communication range of both the legitimate nodes. The legitimate

nodes communicate through the intruder inside the tunnel or having access to tunnel. This intruder can thereafter drop the packet, replay or launch any type of attack and disrupt [19]

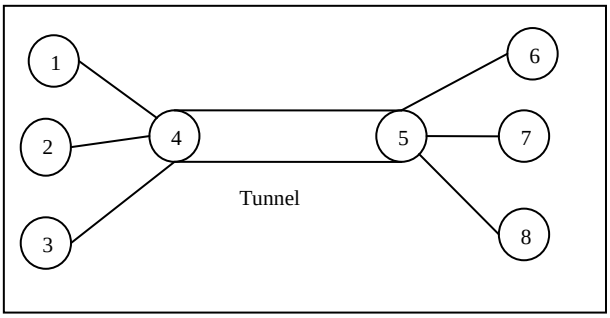


Fig 2.5 Wormhole attack

- **Gray hole attack**: This type of attack is one of the type of packet drop attack with variant. The attacker drops the packet with some variations. The packet drop could be for particular destination or at particular time. There could be another variant such as dropping the packet after some particular number of packets has been transmitted or dropping the subpart of the packet.
- **Bogus Information Attack**: In this type of attack, the intruder sends the untrue on invalid or non existing information in the network. This leads to disturbance in the network and inconsistent data. This intruder can be either inside the network or outside the network. The intruder might act as a legitimate node by sending false identity as bogus information. The intruder could also launch this attack for personal benefit.

Fig 2.3 shows attacks on various security requirements.

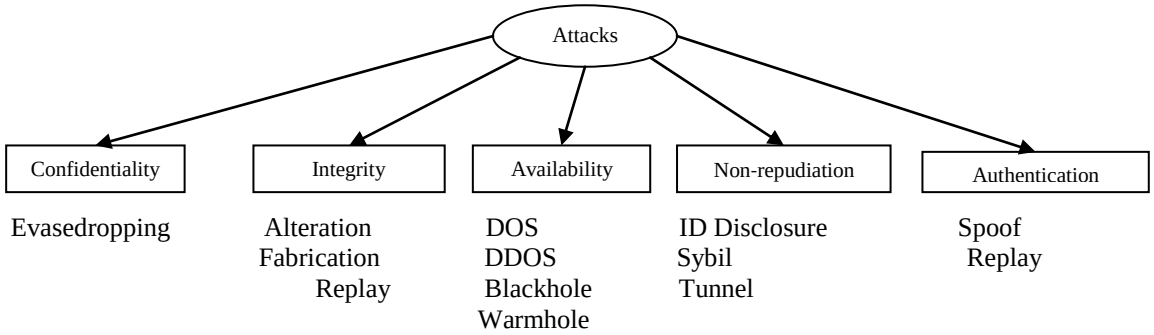


Fig 2.3 Attacks on various security requirements

2.9 Intrusion Detection System

As we saw VANETS are vulnerable to many security attacks, so there need to be some measures to combat these attacks. We saw some of the solutions like group keys, encryption policies etc. These solutions do not prove to be successful when a new attack is to be examined. These solutions can only be used as entry level of protection. After these solution there need to be another layer. This layer is of Intrusion Detection System (IDS). IDS is hardware or software that tries to detect any abnormal behavior in the network.IDS functions in 3 phases as depicted in following figure. First one is event monitoring which includes aggregation of data for any abnormal behavior and the second one is analysis process which includes various techniques like statistics, pattern matching, machine learning etc. The last phase is response generation which alarms about the abnormal behavior and report to admin. Three factors are used to determine the performance of IDS. These are Detection rate (DR), false positive rate (FPR) and false negative rate (FNR). False Positive (FP) takes place when the IDS wrongly identifies normal node as intruder. False Negative (FN) takes place when the IDS wrongly identifies abnormal node as normal. The detection occurs when the IDS identifies the intruder node.

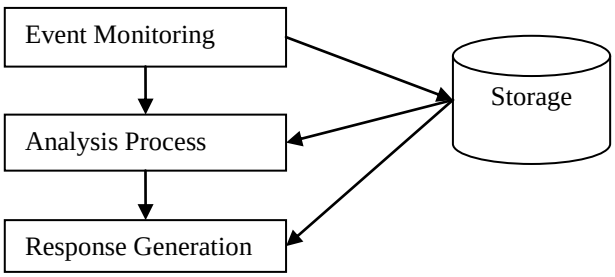


Fig 2.6 Intrusion Detection System

2.10 Classification Of IDS

The IDS is classified on the basis of architecture, techniques, data collection and detection methodology. Following is the flowchart diagram explaining the classification of IDS. We will explore these in this section.

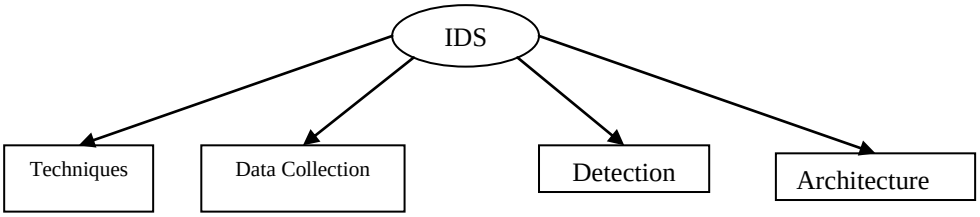


Fig 2.7 Classification of IDS

2.10.1 IDS on the basis of Architecture

On the basis of architecture the IDS can be classified as either standalone IDS, cooperative IDS and Hybrid IDS [4]. Let us discuss about each.

- **Standalone IDS**: In these IDS every node behaves like IDS and monitors the other node. Here there is no transmission of information from one node to other. Each node is individual IDS. Since there is no information about other node there are no alerts that get generated in the system.
- **Cooperative and Distributed IDS**: In these types of IDS there is a cooperation among IDS as there is exchange of information among themselves which makes the detection much more efficient as compared to standalone IDS. However there is a disadvantage as well as the system becomes complex due to frequent exchange of information. The distributed IDS are the ones where there is deployment of multiple numbers of IDS across the network at multiple points. It detects both known and unknown attacks.
- **Hybrid IDS:** To overcome the drawbacks of previous architecture based IDS, hybrid IDS is used. In these IDS, the network is subdivided into group of clusters each having a cluster head. The cooperation is between the cluster head and the members of that cluster. This reduces total exchange messages in the network. The members detect an abnormal behavior and in case the members are not able to detect an alert is sent to cluster head which monitors the cluster to detect attack and hence act as administrator. The detection is distributed to all nodes in network but the alert follows hierarchical manner.

2.10.2 On the basis of techniques

The IDS are classified on the basis of techniques as well. These techniques are implemented to detect the attack on the network. A few of them are discussed below.

- **Fuzzy Logic**: This technique is derived from rules based system. We had a binary system which had two values true or false or existing and non-existing. The real world is much more complex to this. In real world we cannot simply label a class as true or false. There is much more to it. For example the class is 90% belonging to class 1. The temperature of body is hot or cold is irrelevant in real world. Instead the temperature of body is 70% hot is what make sense in real world. The values are not constrained to true or false but have partial result as well. The rules are composed of IF-THEN statements. Following figure shows the black box model of fuzzy logic [10]. The first phase is fuzzify which decides the value of the input. Next is inference engine that contains fuzzy rules. The last phase is defuzzify which merges all suitable rules for implementation. The fuzzy logic is used to for both misuse and anomaly detection. The major advantage of fuzzy logic is that it is easy to modify the rules. However this technique is difficuit to develop.

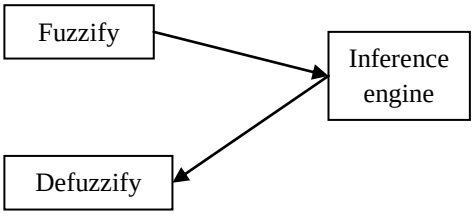


Fig 2.8 Fuzzy Logic phases

- **Support Vector Machines (SVM):** SVM are learning method for solving classification and regression problems. The hyperplane are used as classification medium between various classes. In SVM, classification is done by forming linearly separating hyperplane. The kernel function is used to accomplish this task and converts linear model to non linear [11]. The classification is fast and dynamic updating of training inputs can be done at ease.
- **Bayesian Classifier**: It is a statistical approach based on Baye's Theorem which is stated as follows. Along with other statistics techniques this classifier is used to code interdependencies between dependent and interdependent variables. For each variable the classifier maintains the conditional probability table. The advantage of this approach is high accuracy and handling huge database.

$$P(A|B)=P(B|A).\ P(A)/\ P(B)$$

- **K-Nearest Neighbour**: It stores a training record and waits until it gets a test record. It is an object based classifier that classifies the object on the basis of closest neighbours. The vote of neighbours is used to classify the object. This is also known as lazy learning approach. This approach is easy, simple and easily implemented. However it is slow in classifying and requires huge memory.
- **Decision Trees**: It is a classification model for predictive model. The decision tree is in form of tree based flowchart where the root determines the attribute, the branch represents the outcome, the internal nodes represent test on attribute and leaf represents the class label [9]. The data item is partitioned. This is iterative process and is suitable for large dataset as well. It works well for detecting misuse detection.
- **Pattern Matching**: It is also known as misuse or signature based IDS. It raises an alarm if found something abnormal. It is used to match the pattern of the event recorded with the database available and raises and alarm if found suspicious. It has various variants like exact matching of pattern or approximate matching of the patterns. It functions in two stages- the preprocessing phase and the searching phase. It is easy to add signatures and has huge storage capacity.
- **Clustering**: It is a mechanism by virtue of which the objects that are similar in their traits lie in one group. This group is known as cluster. The similarity is determined from an algorithm [8]. These algorithms vary in the logic. For example the simplest is to calculate centroid of cluster and Euclidian distance is set as logic. Another algorithm is based on density. Another approach is graph based where in any two nodes connected directly with an edge is considered to form a cluster
- **Artificial Immune System**: This technique forms the basis of human body immunity system such as self and non self discrimination. This technique is used to solve diverse list of problems ranging from pattern matching, fault detection and optimization. There are various algorithms useful for Immune system. One such algorithm is negative selection algorithm [7]. The major

objective of this algorithm is to form detector cells which in turn are used for anomaly detection.

- **Artificial Neural Networks:** ANN is a branch of machine learning techniques which are influenced by the central nervous system of animals specifically the brain. ANN is the system of connected neurons where in each neuron is connected to all other neurons. The neurons are the generally the most important processing unit.ANN can be described in two ways- single layer perceptrons (or single layer neural networks) and multi layer perceptron. The neural network consists of minimum of three layers. The first layer consist of input layer which contains as many neurons as features, the next layer is hidden layer where processing takes place and the last layer is output layer which classifies the class of an object. There could be multiple number of hidden layers and variable number of neurons in each hidden layer. Following is pictorial view of architecture of ANN with two neurons in input layer, one hidden layer having three neurons and an output layer with one neuron.

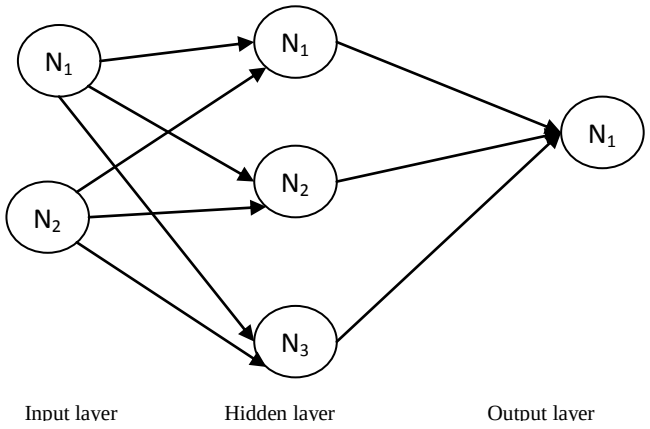


Fig 2.9 Artificial Neural Network block diagram

ANN can be described in two ways- single layer perceptrons (or single layer neural networks) and multi layer perceptron [6]. The single layer perceptron is a type of ANN which consists of two layers – input and output. There is no hidden layer in single layer perceptron. This type of ANN model is used as classification model in which the inputs are linearly separable. This means that inputs belonging to one particular type of class say class1 lies on one side and inputs belonging to other class say class2 lies on other side. The multi layer

consists of an extra hidden layer which is atleast one in number. Following figure shows linearly and non linearly seperable inputs.

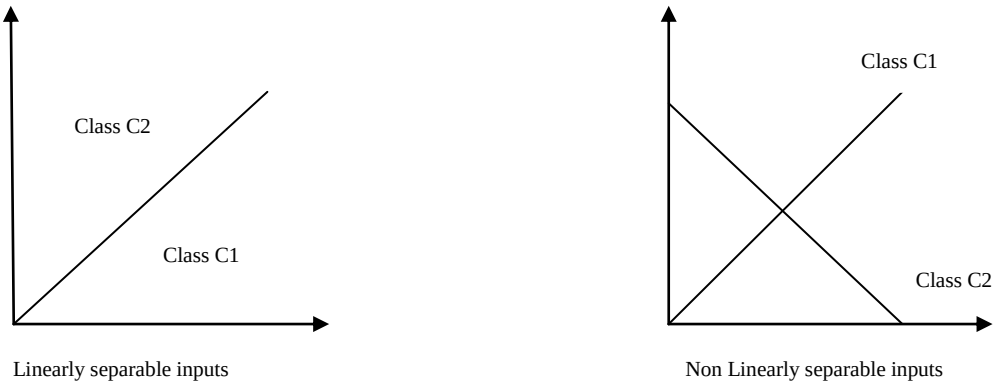


Fig 2.10 Linearly and non linearly seperable inputs

- **Genetic Algorithm**: It is a technique specifically used for improving the efficiency. It uses chromosomes data structure and consist of 3 phases- Selections which is evident from the name to select among the whole population in order to get best results. Next is crossover in which two chromosomes exchange some part of it to form a new one. The last is mutation which is similar to biological mutation. It ensures new chromosomes by changing one or more genes of chromosomes. After mutation a comparison is made of the new chromosome. If the chromosomes are optimized then we provide the solution else the recursive process continues until we find the best solution.
- **Hybrid**: These are combination of two or more of the above or any other new technique to improve the above mentioned techniques in order to detect the malicious activities inside the network. Some of the examples of hybrid techniques include genetic and k-means clustering, neural network and genetic algorithms, Fuzzy system and genetic algorithm etc.

2.10.3 On the basis of data collection

On the basis of data collection, the IDS can be classified as traffic based or behavior based [13]

- **Behavior based IDS**: These IDS are deployed on individual devices in the network to observe the incoming and outgoing packets. These IDS collect logs and determine the malicious activity. The major advantage of this IDS

is decentralization and hence suitable for ad hoc networks. However each node has an extra task of collecting data.

- **Traffic based IDS:** These IDS are deployed on particular points in the network and observes the traffic passively in its subgroup of network and compares with known attacks to detect attacks. The major advantage is that each node need not maintain logs. These are easy to implement however the major drawback lies in hard to detect the encoded data.

2.10.4 On the basis of detection

On the basis of detection, the IDS are classified as follows.

- **Anomaly based IDS**: This system monitors the network and on the basis of behavior, it either passes the nodes (or network) if found normal else raises an alarm if it finds something abnormal. The normal behavior of the network is studied by these IDS known as training data sets. The study of normal behavior of network is known as profiling. There are various ways to train the data set. Some of them are semi supervised learning and unsupervised learning. The merit of this IDS is detection of new attacks for which signature is not present. But it has some demerits as well. Following figure shows anomaly based IDS.

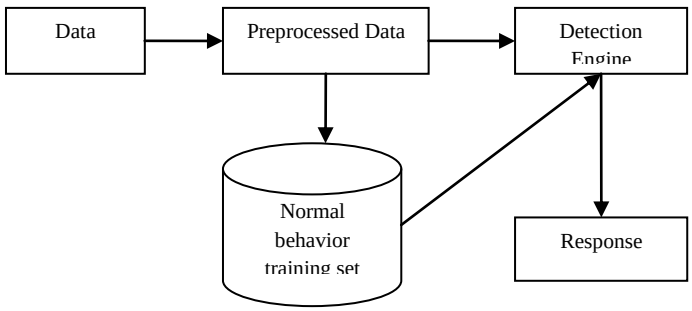


Fig 2.11 Anomaly based IDS

- **Signature based IDS**: These IDS monitors the network and look for abnormal patterns. These abnormal patterns are also known as signature. So these IDS compare the signature of acquired event against known signatures to detect any attack. These types of IDS are effective to detect known attacks and are suitable for outsider attacks. The problem with these IDS is complexity to update signatures and its inefficiency to detect new attacks. Following figure shows signature based IDS.

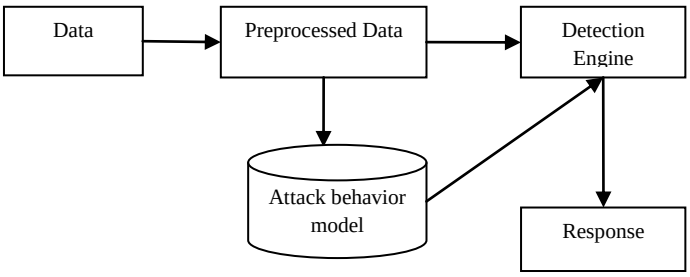


Fig 2.12 Signature based IDS

- **Specification based IDS**: These IDS system look for abnormal behavior at the system level. It is similar to anomaly IDS except that it requires no training phase, instead the IDS learns about normal behavior of the system from individual rather than training phase. These IDS are effective against insider attack. Following figure shows specification based IDS.

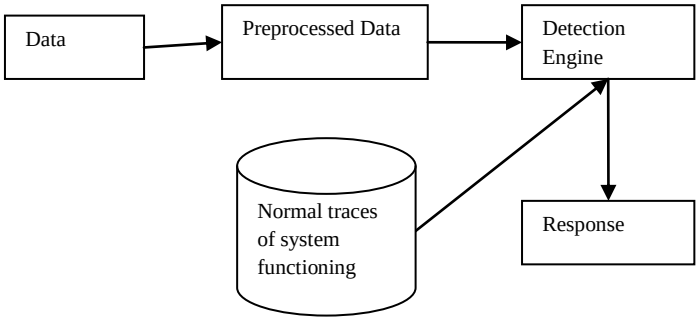


Fig 2.13 Specification based IDS

Table 2.2 Pros and cons of anomaly, signature and specification based IDS

	Anomaly based IDS	**Signature based IDS**	**Specification based IDS**
Pros			
	➢ Detection of new attacks ➢ Efficient against inside attack ➢ Dynamic attack dictionary	➢ Detection of known attacks ➢ Low false positive rate ➢ Efficient against outside attack	➢ Low false negative rate ➢ No training phase ➢ Effective against inside attack
Cons			
	➢ Weak profile	➢ Inefficient for	➢ Complex in

	building ➢ Time critical	new attack ➢ Needs to update signature periodically ➢ Time consuming	generation of specification ➢ No method to train data

CHAPTER-3

LITERATURE REVIEW

This section gives the details about various research works that has been done in securing VANETs. The solution to the attacks can be broadly classified as cryptographic solutions and Intrusion detection system based solutions. Zhou T et al. [33, 34] proposed Public Key Infrastructure wherein each node has public-private keys. While sending the information the sender signs the message with it's private key and add Certificate Authority (CA) certificate. The receiver verifies the signature. Another alternative proposed by Hao Y. et al. [36] was group signature but this is complex in nature as every time a vehicle enters the group its public key and vehicle session key has to be changed and transmitted to the group. Another solution is to provide authentication is where the vehicles will sign the message with its private key and add the certificate along with it. This solution was proposed by Daeinabi et. al. [32]. Grover et al. [35] proposes solution to combat Sybil attack by the use of session keys and digital signature with sequence numbers.

Tajbakhsh et al. [21] proposed Fuzzy Association Rules IDS (FARIDS) and Association Based Classification (ABC). The detection model comprises of learning and the detection phase. In the learning phase features were transformed into items based on fuzzification, the rules were induced and formed after selecting appropriate items and later filtered. The detection phase used classifier to match and label the rule for identifying attack. The model is used for both known and unknown attacks.

Shanmugavadivu et al. [16] proposed fuzzy logic based intrusion detection system. It used KDD Cup 99 dataset and selected 34 continuous attributes. These attributes were mined to form the most critical attribute that affects the whole attack. After this fuzzy rules were created and passed to inference engine to test on testing data. It provided good accuracy for all types of attack-DOS, U2R (user to root), probe and R2l (remote to local).

Hoque et al. [22] proposed an intrusion detection system using genetic algorithm to detect different network intrusions. The famous KDD Cup99 dataset was used. The paper presented two phases- the first phase gave rise to new chromosomes by proving the network data. The second phase outputs the type of data whether attack or normal

by taking previous phase as input. The detection rate gave remarkable results but can be improved with more hybrid techniques used in detection phase.

Benaicha et al. [23] proposed IDS using GA. The features extracted from five different attack types and rules are formed for each of them. The paper builds 80 rules for each attack type which are then fed to GA model. After these 400 rules the evolution process takes place and fitness function is calculated and rule set are formed.

Alheeti et al. [26] proposed a multi stage intrusion detection system. The model is composed of 3 stages. At the first level the data is aggregated and processed. The second stage consists of training dataset. The data after processing from the previous stage is passed to Artificial Neural Network for training of data. The last stage is detection engine where the data is classified as normal or abnormal. The paper is designed for detection of DOS attack.

Sen et al. [5] proposed back propogation Neural Network (BPNN). The BPNN is composed of multiple hidden layers. This paper performs 2 experiments one with 70-30 split and another with 80-20 split of dataset and uses different number of nodes in each hidden layer. The training is done on 1000 epochs and the number of hidden layers is fixed to 4. The number of features selected is 40 and then all of them are assigned a numeric value for normalization to take place and the confusion matrix is formed.

Saied et al. [24] proposed an intrusion detection system based on ANN to combat known and unknown DDOS attack. It used SNORT IDS to monitor the network. The IP identifier identifies the IP address and is then passed to ANN engine which compares it with existing pattern to detect attack. After the detection phase it is passed to defense phase whose role is to stop the attack and allow only legal packets to pass through. It also took the output from other snort ids as well to determine if it has an out dated algorithm in which case retraining is required. The last phase is knowledge share where each detector sends message to other ids. These messages were encrypted. The dataset uses 80% training and 20% testing set.

Barati et al. [15] proposed a network based anomaly IDS to detect DDOS attack using GA and MLP of ANN. The ANN is composed of 3 layers. The data received from GA

is passed to internal layer. The middle layer processes the data and the external layer gives the output.

Panja et al. [25] proposed a hybrid intrusion detection system that used Adaptive Neuro Fuzzy Inference System (ANFIS) at the first place and later uses Genetic Algorithm (GA) and application level filtering. The ANFIS is combination of fuzzy logic and neural networks. This module consist of 4 stages namely data collection, processing, classify and response. The ANFIS has 5 input layers for classifying the attacks as normal, probe, DOS, R2l, and U2R. After this it is passed to fuzzy inference module where fuzzificztion and defuzzificztion takes place. The GA is applied to improve the efficiency of detection. The application level filtering is used for making the detection more accurate by filtering the result from GA

Selim et al. [27] proposed IDS using ANN and Decision trees. The decision tree is if-else statements used as effective classifier. The system consists of 3 layers. The first layer classifies the data. The second layer is responsible for determining the type of attack. The last layer consists of modules for each type of attack.

Balkanli et al. [28] proposed the paper which uses two classifiers Navie Bayes and CART decision tree. The CART and Naïve Bayes classifiers were employed to the model and compared. The experiments were performed on different data sets and features to get realistic results.

Ying et al. [29] presented a host based intrusion detection system using two detection technologies- log file and Back Propogation Neural Network (BPNN). Both these technologies felicitate each other. It combines both anomaly and signature based IDS into its model. The CPU utilization is sent to BPNN as input and its output depicts whether the data is normal or not.

Sandhya G et al. [30] presented a model by combining k means clustering and genetic algorithm to detect unknown attacks without the presence of any actual signatures. The population in current iteration depends on the fitness function for survival in next iteration. The crossover phase was determined by k means operations which include calculation of center and reassignment of each data to closest cluster.

Jongsuebsuk et al. [31] proposed Fuzzy Genetic IDS to detect known and unknown attacks by conducting experiment for each. Later in case of unknown attack the Fuzzy

Genetic is compared to Decision Tree as well. The arguments of fuzzy rules are passed to GA in this model.

Table 3.1 Comparison of literature review

Technique	**IDS**	**Attack covered**	**Merits**	**Demerits**	**Result obtained**
FAR IDS [21]	Anomaly, signature	DOS, probe, U2R, R2L	Less execution time, Efficient classification on large dataset	Decrease in DR in known attacks in anomaly	91% DR and 2.95% false rate
Fuzzy logic based IDS [16]	Anomaly	DOS, probe, U2R, R2L	Appropriate features are used from sample available to reduce cost and memory	Overhead of 1 mined algorithm	Accuracy > 90% for all attacks covered
GA based IDS [22]	Traffic	DOS, R2L, U2R, probe	Filtering of data makes system less complex	Lacks in experiments with distinct GA parameters Huge FP rate	95% DR 30% FP
GA [23]	Traffic	Neptune, pod, smurf, teardrop, back	Remarkable DR with 4 iterations and FP with 10 iterations	Outdated data set used FP is not brilliant as compared to other research	99.74% DR 3.7% FP
BPNN [5]	Anomaly	DOS, probe, U2R, R2L	Different data used among experiments,	High computation cost	Execution time=156ms, DR=98.97%
ANN based IDS [24]	Cooperative, signature	DDOS	Real physical environment was used instead of simulation	Could not detect encrypted DDOS	100% precision and specificity, 98% accuracy
GA and MLP based IDS [15]	Anomaly	DDOS	Only most appropriate features are used	The model lacks robustness	99.9% correctly classified instances
GA and ANFIS based IDS [25]	Cooperative	DOS, R2L, U2R, probe	Application level filtering adds to good accuracy of the model	GA applied is obsolete and can be improved	DR close to 95%

ANN based IDS [26]	Anomaly, signature	DOS	Implemented both anomaly as well as signature based IDS to detect attacks	High FP in anomaly detection	Accuracy=99.12% FN=0.17% (signature), 1.93% (anomaly)
ANN and Decision tree based IDS [27]	Traffic	DOS, R2L, U2R, probe	Detects the abnormal behavior even without knowing the class of attack	Difficulty in detecting U2R. High false rate of C5 decision tree	95.6% DR for both known and unknown attacks
CART and Naïve Bayes based IDS [28]	Anomaly, signature	Backscatter traffic(caused by DOS or DDOS)	Classification without usage of IP and port no. No training required for Bro and Corsaro	High processing time in Bro No default script for DDOS detection Corsaro misclassify ACK and ACK+RST	99% DR with 20% data set used for training
Log file and BPNN based IDS [29]	Anomaly, signature, behavior	-	The system is efficient and improves the accuracy of IDS	Does not identify the type of attack , only notifies abnormal data	The abnormal behavior detected by log file analysis and BPNN uses CPU utilization to detect abnormality
GKA [30]	Anomaly	Blackhole	Can detect attacks without any signature Suitable for dynamic environment	High computation cost	DR better than k means and Enhanced IDS
Fuzzy Genetic IDS [31]	Traffic	DOS, Probe	Low computation cost Detect known and unknown attacks effectively	Preprocessing time is more than detecting time	DR=97.9% for known attacks and can even detect 100% for some unknown attacks

CHAPTER 4

PROBLEM STATEMENT

4.1 Gaps in Study

Ad hoc networks are vulnerable to many security threats especially vehicular networks. The ill effect is so severe that it may cause even loss of life. The different solutions have been proposed by various researchers as discussed in previous section. Some of them include cryptography while some uses intrusion detection. Cryptography however does not provide optimum results as seen earlier. Cryptographic solutions being resource intensive consume more energy of vehicles. Though a lot of research has been conducted in IDS based solutions, there are still a lot of demerits with these works like requirement of hardware, high false positive rate, lacks in experiments and robust dataset. So there is still need for research to be conducted in securing VANET against RREQ flooding attack so that an efficient system could be developed.

4.2 Problem Statement

Vehicular networks are vulnerable to many security threats and the ill effect could even risk the life of person. So there is a need to have secure routing of data packets in VANETs as routing protocols do not incorporate security by default. There is a need to have an intrusion detection system to detect the RREQ flooding attack in reactive protocols such as AODV. Intrusion detection techniques range from machine learning, statistical to game theory etc. Out of these techniques, the machine learning suits the detection problem as it solves classification problem by allocating the objects to the class it belongs to. In this thesis we have implemented RREQ Flooding attack, there needs to be a detection technique which classifies the features into attack class or normal class.

4.3 Objectives

The following research objectives are formulated

- To launch RREQ flooding attack and calculate packet delivery ratio and throughput trend on increasing the malicious nodes.

- To create our own dataset having substantial amount of normal and attack records.
- To implement the machine learning technique for intrusion detection and classifying the features into attack or normal class.
- To reduce the number of features available using GA.
- To calculate the performance metrics such as accuracy, false positive rate etc of our proposed system with other existing works.

CHAPTER 5

PROPOSED ALGORITHM AND IMPLEMENTATION

5.1Proposed System Architecture

The proposed system is used to detect RREQ Flooding attack using ANN. It is optimized in terms of feature subset selection using GA. The network simulator ns-2 is used for launching the RREQ Flooding attack. This algorithm works well for multiple numbers of malicious nodes and gives remarkable results on evaluating the performance metrics like accuracy and false positive rate. The simulation of VANET environment is done through SUMO. The tcl file which is generated by MOVE simulator, is used as input to NS-2. The implementation generally involves three stages. The first stage is creation of dataset by launching the attack. The second step is Data Preprocessing and the last step is the classification and optimization engine. The classification technique used is Artificial Neural Networks and the optimization techniques make use of Genetic algorithm. Figure 5.1 explains the overall architecture of proposed system.

The proposed architecture consists of three stages which are explained in detail in this chapter.

- Creation of data set
- Data Pre-processing
- Classification and Optimization engine

5.1.1 Creation of data set

For creation of data set, a VANET environment was set up by integration of SUMO, MOVE and NS-2. The integration of SUMO and MOVE is shown in figure 5.2(a), 5.2(b) and 5.3. The output of MOVE file is a tcl script used by NS-2. The output of trace files were collected as output for two different scenarios- normal AODV and AODV under RREQ Flooding attack.

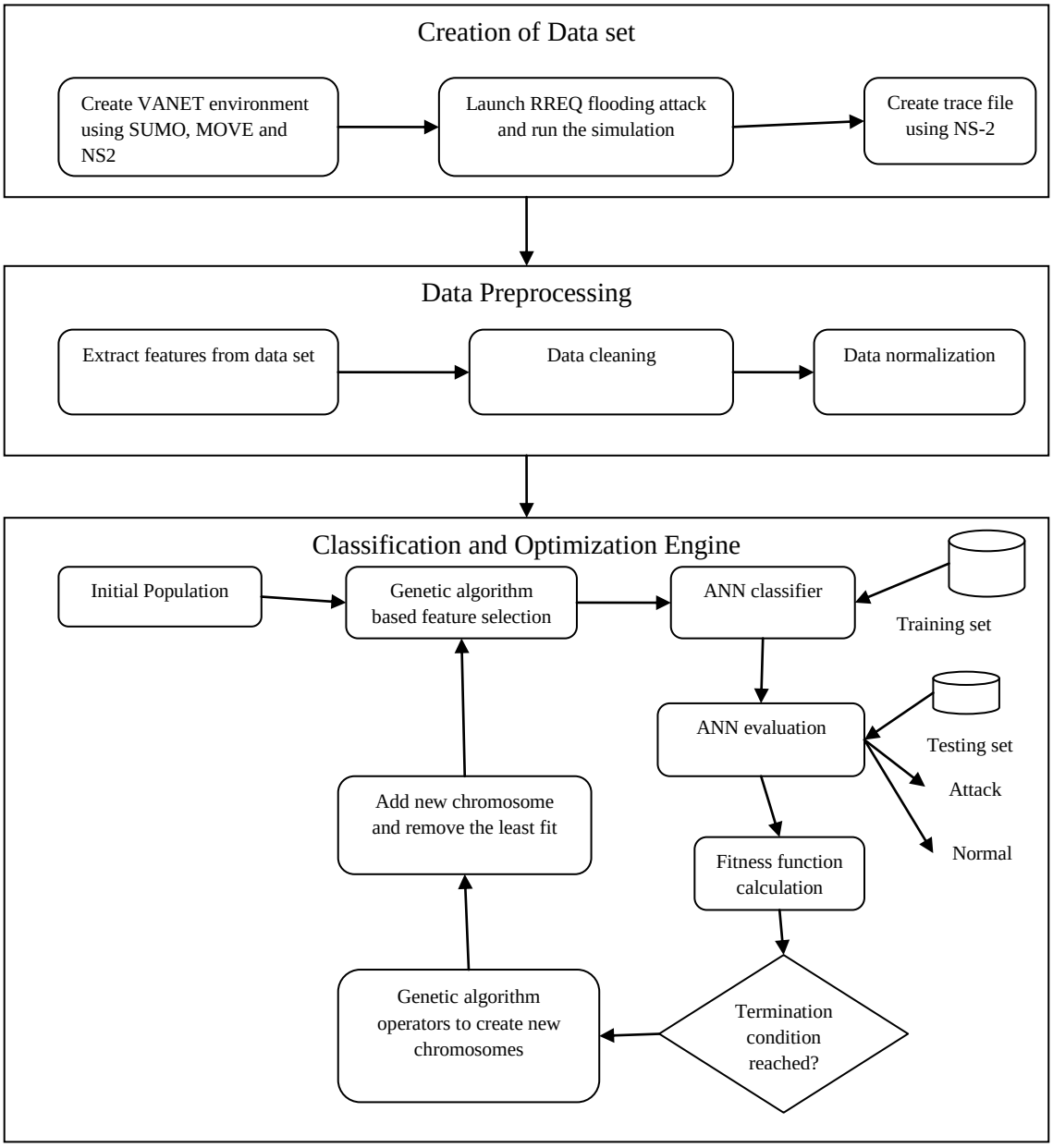
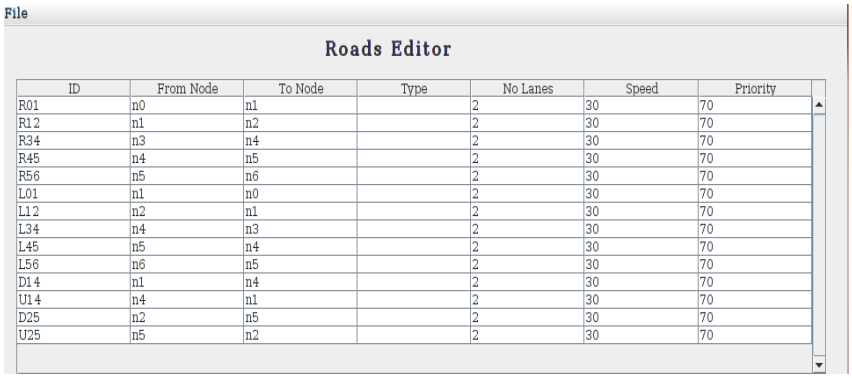
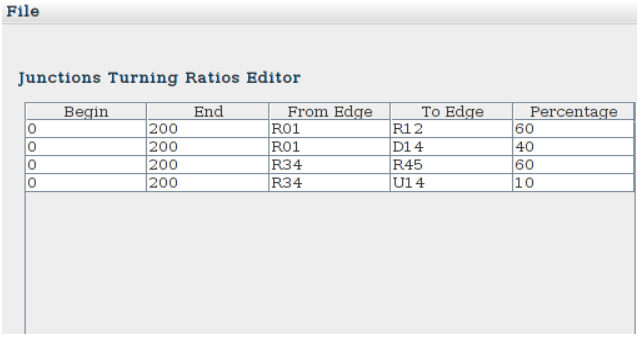


Fig 5.1 Proposed System Architecture



ID	From Node	To Node	Type	No Lanes	Speed	Priority
R01	n0	n1		2	30	70
R12	n1	n2		2	30	70
R34	n3	n4		2	30	70
R45	n4	n5		2	30	70
R56	n5	n6		2	30	70
L01	n1	n0		2	30	70
L12	n2	n1		2	30	70
L34	n4	n3		2	30	70
L45	n5	n4		2	30	70
L56	n6	n5		2	30	70
D14	n1	n4		2	30	70
U14	n4	n1		2	30	70
D25	n2	n5		2	30	70
U25	n5	n2		2	30	70

Fig 5.2 (a) Creating VANET environment



Begin	End	From Edge	To Edge	Percentage
0	200	R01	R12	60
0	200	R01	D14	40
0	200	R34	R45	60
0	200	R34	U14	10

Fig 5.2 (b) Creating VANET environment

Fig 5.3 Visualization using SUMO

Fig 5.4 gives the pseudocode for creation of data set. The normal.tr and malicious.tr are the two trace file generated by running the simulation in ns-2 for normal data and malicious entries of data respectively.

```
Input: MOVE.jar
Output: Two trace files normal.tr and attack.tr
1.  no_of_vehicles ←20
2.  simulation_time ← 200
3.  for each node n
4.          set node's cordinates
5.  end for
6.  for each edge e
7.          lanes ← 20
8.          speed ←30
9.          priority ←70
10.         initialize edge id
11. end for
12. for each flow f
13.         initialize flow id
14.         no_of _vehicle_per_flow ←no_of_vehicles/no_of_ flow
15. end for
16.  CreateVehicle()
17. ConfigVehicle()
18. call Visualize()
19. add_Connection()
20. normal.tr ←NS2()
21. RREQ_Flooding()
22. Attack.tr ←NS2()
```

Fig 5.4 Pseudocode for creation of data set

The normal.tr and malicious.tr combined forms data set in which the normal entries belong to one particular class normal (normal.tr) and the malicious entries belong to another class abnormal (malicious.tr).The malicious node will create multiple number of RREQ to a destination node, which doesn't exist in the network topology. The purpose of this attack is to consume the network bandwidth and to exhaust the network resources all the time. Following figure 5.5 depicts pseudocode for launching RREQ Flooding in AODV.

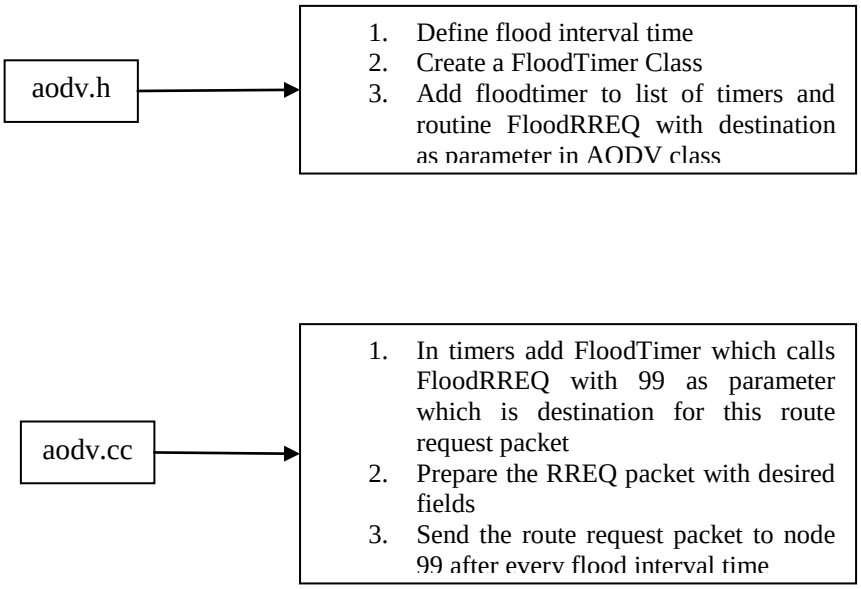


Fig 5.5 Pseudocode for launching RREQ Flooding

5.1.2 Data Preprocessing

The real data consists of erroneous data and might also be not very useful as it is raw data. There is a need to convert this raw data into meaningful information. To achieve this objective, data preprocessing is required. The data preprocessing phase as described above in our proposed architecture involves three main steps- Extraction of features from data set, data cleaning and data normalization.

- **Extracting features from the data set**

The normal.tr and malicious.tr files obtained while launching the attack collectively forms a data set. These trace files contains many fields and are separated by space delimiter. The following figure gives snapshot of the data set.

```
r 0.000963000 _3_ MAC  --- 0 AODV 48 [0 ffffffff 1 800] ------- [1:255 -1:255 30 0] [0x2 1 1 [6 0] [1 4]] (REQUEST)
r 0.000963000 _5_ MAC  --- 0 AODV 48 [0 ffffffff 1 800] ------- [1:255 -1:255 30 0] [0x2 1 1 [6 0] [1 4]] (REQUEST)
r 0.000963000 _7_ MAC  --- 0 AODV 48 [0 ffffffff 1 800] ------- [1:255 -1:255 30 0] [0x2 1 1 [6 0] [1 4]] (REQUEST)
```

Fig 5.6 Snapshot of dataset-1

The trace file is space separated and contains many fields. The trace file is divided into three traces namely Basic Trace, IP trace and AODV trace. Table 5.1, 5.2 and 5.3 gives brief overview of trace files:

- **Basic Trace**

Table 5.1 Basic Trace

Column Number	Description

1	Represents event. s ,r, f and d represents that packet is sent, received, dropped or forwarded.
2	Represents the time in seconds from start of simulation.
3	Represents the node id.
4	Trace level-AGT, RTR, MAC.
5	Reason. This is generally blank but if the event is dropped it list down the reason for dropping packet like collision.
6	Represents packet id.
7	Represents packet type-AODV, ACK, TCP, RTS, CTS.
8	Represents packet size corresponding to the type of packet.
9	Represents the duration in MAC layer.
10	Destination MAC addresses.
11	Source MAC Address.
12	Ethernet packet type 0x800 represents IP (ETHERTYPE_IP) and 0x860 represents ARP (ETHERTYPE_ARP).

- **IP trace**

Table 5.2 IP Trace

Column Number	Description
13	Reserved for flags
14	Source ip address and port number separated by colon as IP address: Port number
15	Destination ip address and port number separated by colon as in column 14.
16	Represents TTL (time to live)
17	Represents the next hop

- **AODV Trace**

Table 5.3 AODV Trace

Column Number	Description
18	Hexadecimal number representing Route Request and Route

	Reply. 0x02 represents the request and 0x04 represents the reply
19	Represents the hop count
20	Represents the broadcast id
21	Destination IP address
22	Destination Sequence number
23	Source IP address
24	Source Sequence number
25	Represents the label like REQUEST, REPLY etc

After going through the trace file, the features were extracted. Here normal.tr file is extracted column wise. Similarly we run it again for malicious.tr and get the columns of malicious records as well. We get twenty five columns as output. These twenty five columns act as feature attributes of our data set. The figure 5.7 shows the data set

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
r	0.001	_28_	RTR	0	AODV	48	[0	ffffffff	1	800]	[1,255	-1,255	30	0]	[0x2	1	1	[6	0]	[1	4]]	(REQUEST)
r	0.001	_30_	RTR	0	AODV	48	[0	ffffffff	1	800]	[1,255	-1,255	30	0]	[0x2	1	1	[6	0]	[1	4]]	(REQUEST)
r	0.001	_32_	RTR	0	AODV	48	[0	ffffffff	1	800]	[1,255	-1,255	30	0]	[0x2	1	1	[6	0]	[1	4]]	(REQUEST)
r	0.001	_34_	RTR	0	AODV	48	[0	ffffffff	1	800]	[1,255	-1,255	30	0]	[0x2	1	1	[6	0]	[1	4]]	(REQUEST)
r	0.001	_36_	RTR	0	AODV	48	[0	ffffffff	1	800]	[1,255	-1,255	30	0]	[0x2	1	1	[6	0]	[1	4]]	(REQUEST)
r	0.001	_38_	RTR	0	AODV	48	[0	ffffffff	1	800]	[1,255	-1,255	30	0]	[0x2	1	1	[6	0]	[1	4]]	(REQUEST)
s	0.001	_6_	RTR	0	AODV	44	[0	0	0	0]	[6,255	1,255	30	1]	[0x4	1	[6	4]	10.000	(REPLY)		
s	0.0014	_23_	RTR	0	AODV	48	[0	ffffffff	1	800]	[23,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0019	_3_	RTR	0	AODV	48	[0	ffffffff	1	800]	[3,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0019	_35_	RTR	0	AODV	48	[0	ffffffff	1	800]	[35,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.002	_3_	MAC	0	AODV	106	[0	ffffffff	3	800]	[3,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0021	_30_	RTR	0	AODV	48	[0	ffffffff	1	800]	[30,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0021	_27_	RTR	0	AODV	48	[0	ffffffff	1	800]	[27,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0026	_21_	RTR	0	AODV	48	[0	ffffffff	1	800]	[21,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0027	_33_	RTR	0	AODV	48	[0	ffffffff	1	800]	[33,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
s	0.0029	_34_	RTR	0	AODV	48	[0	ffffffff	1	800]	[34,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
r	0.0029	_5_	MAC	0	AODV	48	[0	ffffffff	3	800]	[3,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)
r	0.0029	7	MAC	0	AODV	48	[0	ffffffff	3	800]	[3,255	-1,255	29	0]	[0x2	2	1	[6	0]	[1	4]]	(REQUEST)

Fig 5.7 Snapshot of dataset-2

- **Data Cleaning**

It refers to correcting or removing of inconsistent records from the data set. In this phase we try to make the data set robust and unbiased. The redundant data or the data that assumes value zero all the time are removed. Handling of missing values is also done in this phase. As described above, the data set consists of 25 fields but we remove some of the fields in this phase. The column 5 is removed as it is mostly empty and contains reason COL (collision) only for the event type drop. The column 13 is also removed as it is empty. The column 18 is removed which represents the

hexadecimal code for route request or route reply. The columns 5, 13 and 18 are removed and we are left with 22 columns. After removing these fields, there arises the problem of handling the missing data. Some of the records had insufficient fields. The solution to this problem could be removing all instances where we have insufficient data. But doing so would result in very less data set size. Another solution to this problem could be filling the insufficient field with some value. The main problem was with what value the missing data could be filled in. One solution is filling it with most common feature value but this could make dataset biased. Another alternative would be filling it with either special value or mean of the feature value. We filled the missing data with mean of that feature value.

- **Data normalization**

This phase is further subdivided into two stages- Converting the values into numeric and the second stage is normalizing the data set so that the whole data set lies in one particular range.

1. Converting non numeric into numeric fields

In this phase the non integer values like TCP, AODV, MAC, RTR, AGT, ACK, s, r, d, f etc are converted into integer. For eg in column1 s, r, d, f is represented by 1, 2, 3 and 4 respectively. Likewise it is done for all the columns. The columns which consist of hexadecimal values are also converted into integers. This phase is important to be performed for dataset to be normalized. All the square brackets etc are also removed in this phase. Removal of brackets can also be considered in Data cleaning as well.

2. Converting data into standardized form

The dataset in the previous phase contains only integer values. But those integer values have a wide range. It would be selfish act to let that wide range of inputs. As when we apply the detection techniques the comparison between the fields would make no sense if they have huge deviation. The data with huge values will affect the data with lesser value, hence converting the data into standard form becomes indispensable. The figure 5.10 gives snapshot of normalized dataset.

1. Calculate mean of all the columns as m1, m2, and so on up to m22.
2. Calculate variance of all the columns as v_1, v_2, and so on up to v22
3. Normalize each value X_i as

$$X_i = \frac{(Xi - \bar{X})}{v} \text{ where I ranges from 1 to 22.}$$

$$\text{Mean } (\bar{x}) = \frac{Sum\ of\ all\ values}{Total\ Number\ of\ values}$$

$$\text{Variance (v)} = \frac{\sum(Xi - \bar{X})^2}{n}$$

Fig 5.8 Steps for Data Normalization

A	B	C	D	E	F	G	H
-1.4627	-0.7954	-0.9619	-1.4942	-0.3865	-0.7640	-0.4991	-0.4335
-1.4627	-0.7953	-0.9619	0.3822	-0.3865	-0.7640	-0.3477	-0.4335
0.1067	-0.7946	-0.7959	0.3822	-0.3865	-0.7640	-0.4991	-0.4335
0.1067	-0.7946	-0.6298	0.3822	-0.3865	-0.7640	-0.4991	-0.4335
0.1067	-0.7946	-0.4638	0.3822	-0.3865	-0.7640	-0.4991	-0.4335
0.1067	-0.7946	-0.2977	0.3822	-0.3865	-0.7640	-0.4991	-0.4335
0.1067	-0.7946	-0.1317	0.3822	-0.3865	-0.7640	-0.4991	-0.4335

Fig 5.9 Snapshot of normalized dataset

5.1.3 Classification and Optimization Engine

After normalizing the data set, we come to third phase of our proposed architecture which is classification and optimization engine. In the proposed algorithm, Artificial Neural Networks has been used as classification model and Genetic Algorithm has been used for feature sub-selection. These two techniques are described in detail and later gives an overview of the third phase of proposed architecture

5.1.3.1 Artificial Neural Networks (ANN)

ANN is a branch of machine learning techniques which are influenced by the central nervous system of animals specifically the brain. ANN is the system of connected neurons where in each neuron is connected to all other neurons. The neurons are the generally the most important processing unit. The brain processes the information by interchanging the pulses among neurons. The neurons are connected to each and every input with weight associated with particular input and results in the outcome. The same idea is applied by ANN in computer science for classification or prediction based issues. ANN is flexible system by virtue of which the organization of network

changes due to change in the inputs, weights associated with inputs or any other parameter. In this neural network, there is another layer known as hidden layer. The Multi Layer Perceptron consists of input layer, one or more hidden layer and finally the output layer. The number of hidden layer depends on the application. Also the number of neuron in each hidden layer is application specific. The hidden layer is also referred to as the processing layer.

The proposed architecture is based on this model of ANN. The neural network consists of minimum of three layers as explained in previous section. The input layer consists of as many neurons as the number of features used in dataset for solving classification problem. In our case after the normalization phase there are 22 features, so we have twenty two neurons in the input layer. The next layer is the hidden layer which is the processing layer of this model. It can have any number of neurons. In our case, we fixed its size to 10 neurons. We have also used only one hidden layer. After the hidden layer comes the output layer. The output consists of the class of the input. We used numeric value for class, hence used one neuron for the output layer. Apart from the three layers we have weights associated with the inputs. We also have initial biases. The neurons are fully connected between the layers of ANN model. This Multi Layer model is also known as feed forward model because of lack of any backward connectivity in the model. However the error is calculated using back propogation concept. We will discuss about this later in this section. As shown in figure 5.11 there are inputs at first level with associated weights and biases values. The sum of product of input with its weights is calculated and is passed to transfer function. The result of this transfer function is passed to training function to yield the outcome of the model. The hidden layer consist of 10 neurons namely N_1, N_2 and so on up to $N_{10.}$

The ANN works in 2 phases – training phase and testing phase. During the training phase the inputs are passed at input layer. The sum of product of input and weights along with the bias values are passed to processing layer. At the processing layer lies the transfer function. We can choose from many transfer functions available. Purelin transfer function is used in our proposed algorithm as it gives best results while working in feed forward networks. The figure 5.10 shows different transfer functions. The output of this transfer function is passed to training function. Again we have multiple training functions to choose from but we stick to Levenberg-Marquardt algorithm as training function.

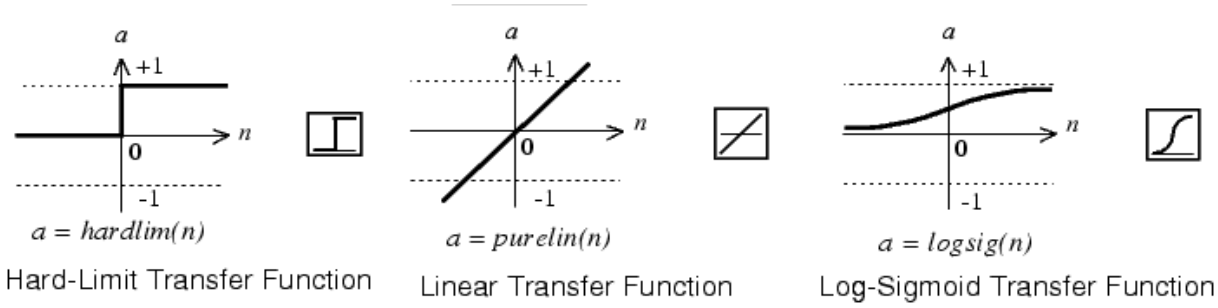


Fig 5.10 Transfer functions

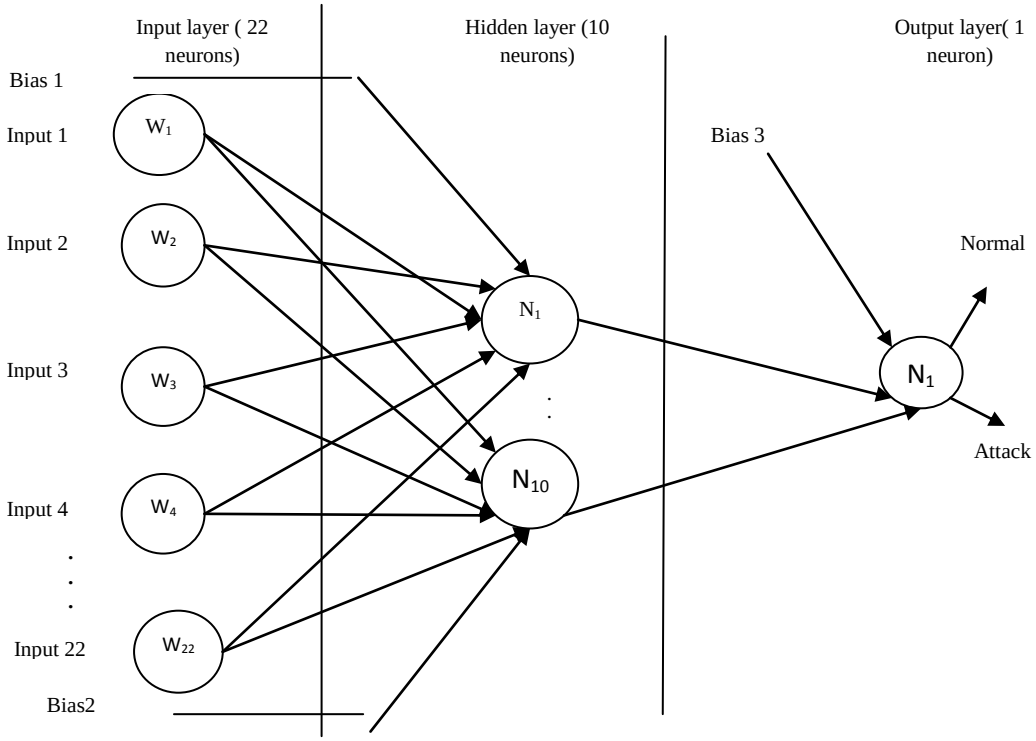


Fig 5.11 Artificial Neural Network working with 5 inputs

Back Propagation

As mentioned above the ANN is feed forward network. During the training phase the weights and biases are calculated and performance metrics are noted. However there needs to be change in weights and biases values if the performance metric thresholds are not obtained. Due to change in weights and biases values we need to move back at the first layer and again calculate the values using the training function [5]. Since we move from last layer to first layer back, it is known as back propagation. This back propagation works by calculating gradient at each layer known as local gradient.

The weights and biases values are updated according to Levenberg-Marquardt optimization. If the performance function contains sum of squares then the Hessian matrix [37] is defined as

$$\mathbf{H} = \mathbf{J}^T\mathbf{J}$$

and the gradient can be computed as

$$\mathbf{g} = \mathbf{J}^T\mathbf{e}$$

where **J** is the Jacobian matrix . This matrix contains first order derivatives of the errors with respect to the weights and biases, and **e** is the network error vector. The Jacobian matrix can be deduced through a standard backpropagation technique which is less complex to Hesian approach. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix

$$\mathbf{x}_{k+1}=\mathbf{x}_k-[\mathbf{J}^T\mathbf{J}+\mu\mathbf{I}]^{-1}\mathbf{J}^T\mathbf{e}.$$

Figure 5.12 shows the workflow of Artificial Neural Network along with back propagation technique.

5.1.3.2 Genetic Algorithm (GA)

Genetic algorithm is an artificial intelligence heuristic approach which imitates the methodology of natural evolution. Evolution is the process by virtue of which the organisms improve successively over generations through the GA operators described later in this section. The Genetic Algorithm is used for optimization problems. It follows the principle of survival of the best which means the best feature individual will be selected over successive generations and hence improving and making the system more efficient. It is more efficient than normal Artificial Intelligence as GA are more stable and do not get affected by minute change in inputs.

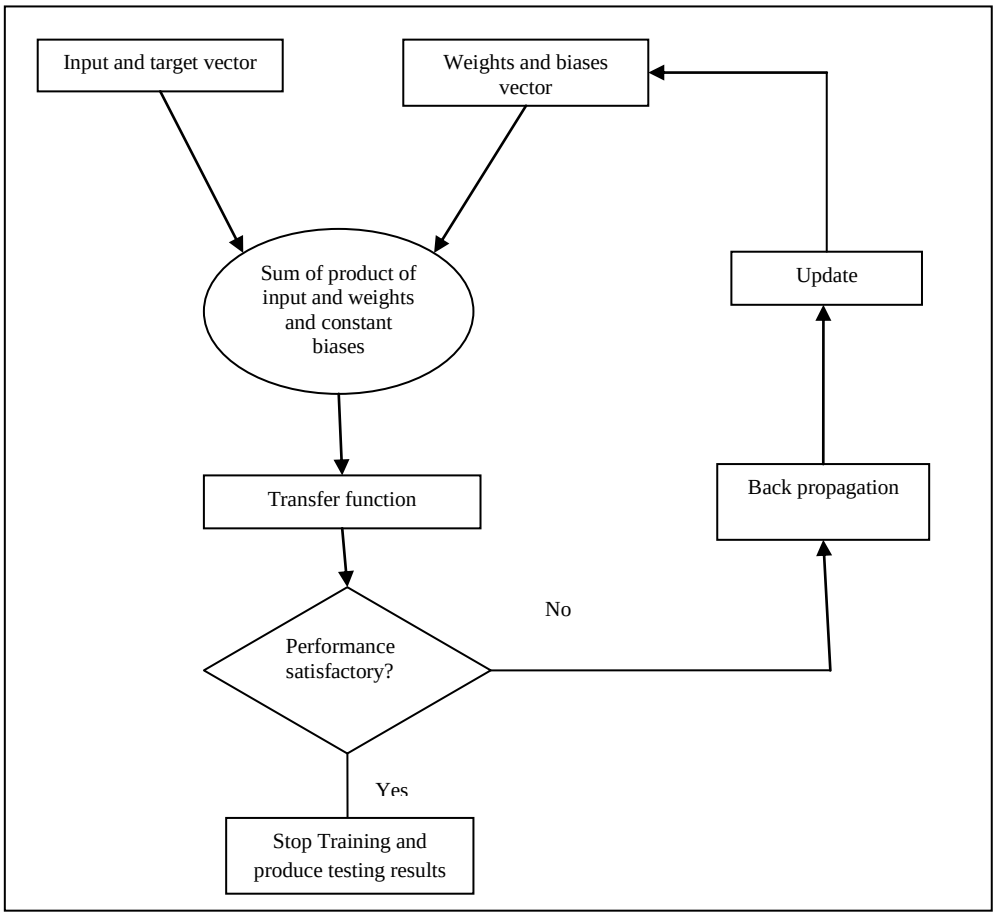


Fig 5.12 Workflow of ANN along with back propagation

In GA the population is represented as a string known as chromosomes [17]. Every chromosome represents part of space to be evaluated. The space referred here is known as search space which is the possible values of all the features of the system. The chromosomes are represented are generally represented as bit strings. In some case we use double vector as well. The representation is required as it allows the parameters to be mapped into strings which make the algorithm easy to work on. The proposed algorithm uses double vector as mean of representing the chromosomes.

- **Fitness Function**

This is utmost important component in Genetic Algorithm evaluation. It is also known as driver function of GA. It indicates the problem to be solved. The fitness function is evaluated at each generation in order to make a decision for a particular feature. The fitness function is kept as the classification accuracy in our thesis. This is calculated at every generation and only survival of best is kept for next round.

- **Feature Subset Selection**

This is vital topic in machine learning and is especially indispensable wherein the data set is huge having many features. Generally only few features are relevant and contain the valid information and the others are irrelevant. These irrelevant features which do not provide any useful information when passed as a part of input to the algorithm tend to decrease the performance and robustness of the algorithm. So there is a need to remove the unwanted and undesirable features which does not contain valid information and does not contribute much in increasing the system performance and stability.

Generally two approaches are used for this purpose. These are the filter approach and the wrapper approach [15]. The filter approach does not depend on classification algorithm. Instead the feature subset selection is done prior to classification algorithm is used. In this case each feature is selected independently using statistic test. This method might select the redundant features as it does not consider the relationship between the features. The wrapper method makes use of classification algorithm to select the subset of features. The best set of features are retained and the least one is removed from the list hence modifies the feature set. This method is more accurate than the previous one but takes more computation cost as well. Figure 5.13 and 5.14 shows the two feature subset selection using the two approaches discussed above.

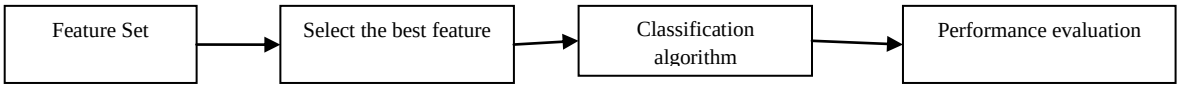


Fig 5.13 Filter approach

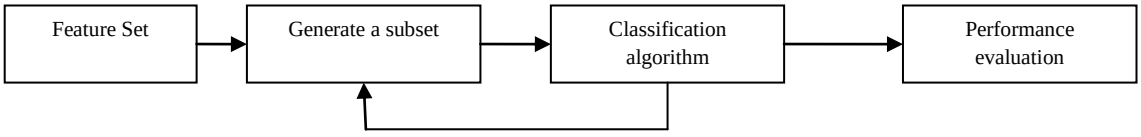


Fig 5.14 Wrapper Approach

The performance evaluation mentioned in figure 5.14 is generally either the classification accuracy, runtime or the number of selected feature. Our proposed algorithm have used wrapper method to select feature subset.

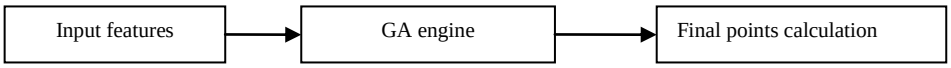


Fig: 5.15 Feature subset selection

- **Genetic algorithm operators**

The Genetic algorithm starts with an initial population. This initial population is randomly chosen from list of population. After initializing the initial population, the fitness function is evaluated. Based on the result of evaluation the genetic algorithm operators like selection, crossover and mutation are applied. The whole procedure is repeated until stopping criteria is matched [14]. The following block diagram gives an overview of Genetic algorithm operators.

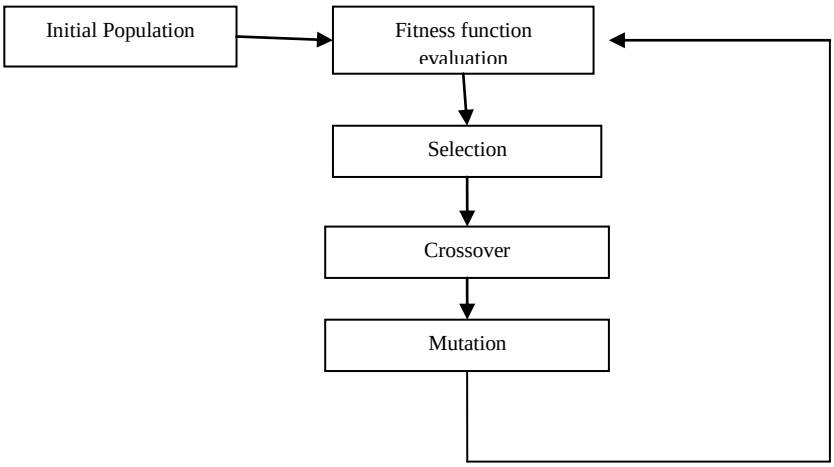


Fig 5.16: Genetic Algorithm Operators

- Selection

This is the process of selecting the chromosome based on fitness values and survival of the best. The GA determines which chromosome to be included into feature set. There are various selection algorithms available to choose from. Some of them are roulette, tournament, uniform etc. In our proposed algorithm, stochastic uniform selection algorithm is used for selection. In this algorithm 'n' equal spaces are marked. A random number is chosen from [0,1/n]. This random number marks the position of first point.

- Crossover

This is vital operator of GA. The performance of the system depends on crossover and mutation. It is the process of swapping genes of the chromosomes between the individuals. There are various crossover functions like single point, dual point etc. Let us consider single point crossover. In our proposed algorithm conventional crossover is used. Suppose P_1 and P_2 are the two parents, C_1 and C_2 are two child and M_1 and M_2 are two masks, then according to scattered crossover.

$C_1=M_1\&P_1 + M_2\&P_2$

$C_2=M_2\&P_1 + M_1\&P_2$

M_1 and M_2 are mask values where M_2 is NOT(M_1)

The following figure demonstrates the scattered crossover

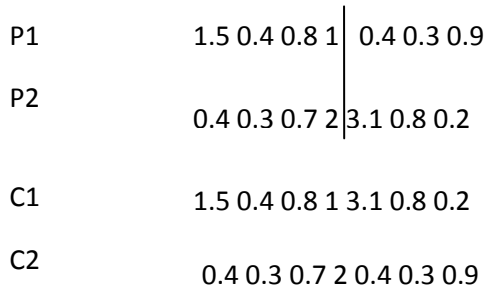


Fig 5.17 Demonstration of conventional crossover

- Mutation

This GA operator is similar to biological mutation. It changes the value of genes. The mutation is responsible for maintenance of diversity of genes of the chromosomes. There are various mutation functions available. In uniform mutation function, the mutation operator replaces the value of gene. A random number is selected and it lies between the the bounds constraints specified by user. This function is used by integer and float. Apart from uniform there are various other mutation functions available to choose from. The mutation scenario demonstration is given in Fig 5.18

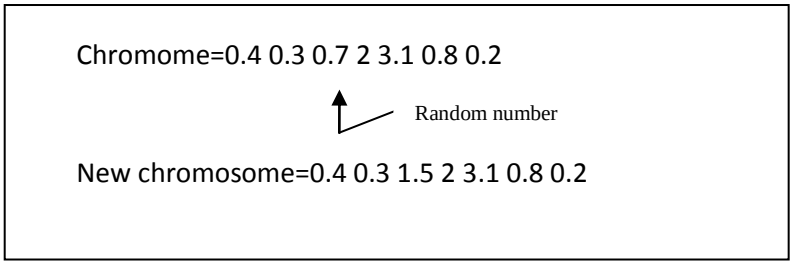


Fig 5.18 Mutation Process to add new chromosome

Fig 5.19 shows an algorithm using genetic algorithm approach to select subset of feature with best fit. According to this algorithm a new chromosome is added to the population if the fitness value is improved or if the number of iteration reaches 100. Every generation, the values of parameters are noted and the least fit is removed from the population. This process is repeated until there is improvement in fitness value.

Input: Intial population

Output: Feature sub selection

1. do{ For each chromosome c
2. Calculate fitness function
3. Create vector using mutation
4. Create new individual using crossover
5. end for
6. if fitness value is greater than previous individual
7. add the individual to the list
8. else
9. remove the least fit individual
10. }while(stopping criteria is not met)

Fig 5.19 Algorithm for feature sub-selection

CHAPTER 6

SIMULATION RESULTS AND DISCUSSIONS

This section elaborates about the simulation and environment setup followed by the results and later does the comparative analysis of the proposed approach with other related work. The simulation results are discussed. The awk programming is used to deduce the throughput and packet delivery ratio from trace file to make sure that the attack has been launched. After this, the Artificial Neural Network and genetic algorithm techniques are used to detect the RREQ Flooding attack.

6.1 Simulation and Parameter Setup

The simulation environment parameters are shown in table 6.1. The Artificial Neural Network and Genetic algorithm parameters are shown in Table 6.2 and Table 6.3 respectively. The RREQ flooding attack scenario is simulated in NAM which gives graphical view of network. The RREQ flooding attack with two nodes malicious is shown in figure 6.1.

Table 6.1 Simulation Environment Parameters

Parameter	Value
No of nodes	20
No of malicious nodes	2
Channel Type	Wireless
Routing Protocol	AODV
MAC_TYPE	802.11
Packet Size	1000
Interface Queue Type	Queue/Drop Tail/Priority
Simulation Time	200

Table 6.2 ANN Parameters

Parameter	Value
Network Type	Feed forward backdrop
Training Function	Levenberg-Maquardt
Performance Function	Mean Square Error
No of layers	2
No of neurons per layer	10
Transfer Function	Transig
Maximum Epochs	1000

Validation check	6
Data Division	Random

Table 6.3 GA Parameters

Parameter	Value
Genetic Operations	Scattered crossover, single point mutation
Selection Method	Stochastic
Crossover Rate	0.8
Mutation Rate	0.02
Population Size	200
Max Generations	100
Stall Generations	50

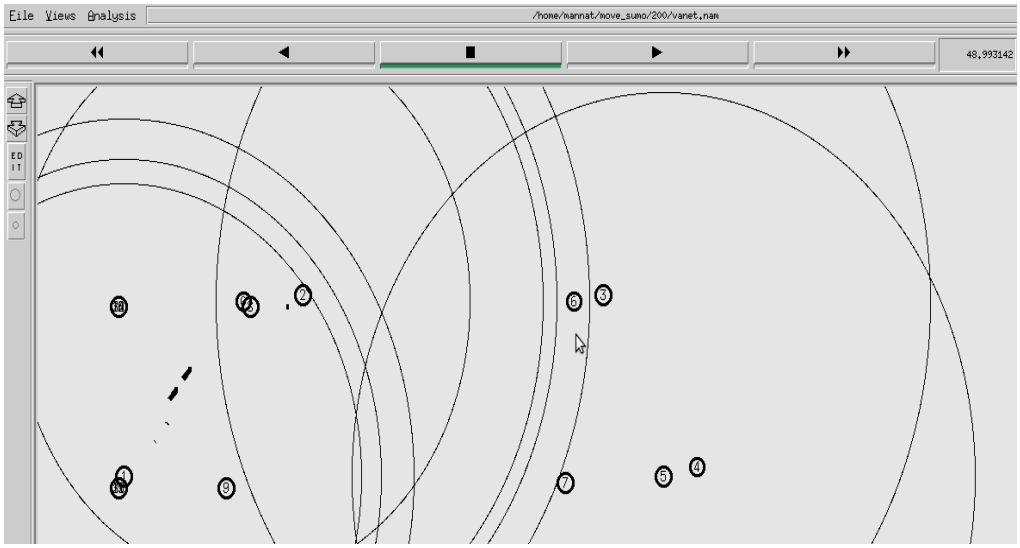


Fig 6.1 Flooding attack graphic view in NAM

6.2 Performance Metrics

After launching the flooding attack in ns-2, some of the parameters of the network like Packet Delivery Ratio and Throughput are noted .The decrease in these parameter values with increase in number of malicious nodes confirmed that the attack has been launched. Packet Delivery Ratio or PDR is the ratio of number of packets sent to the number of packets received. Throughput is defined as rate of successful message sent. The graph shows the network parameters in fig 6.2, 6.3.

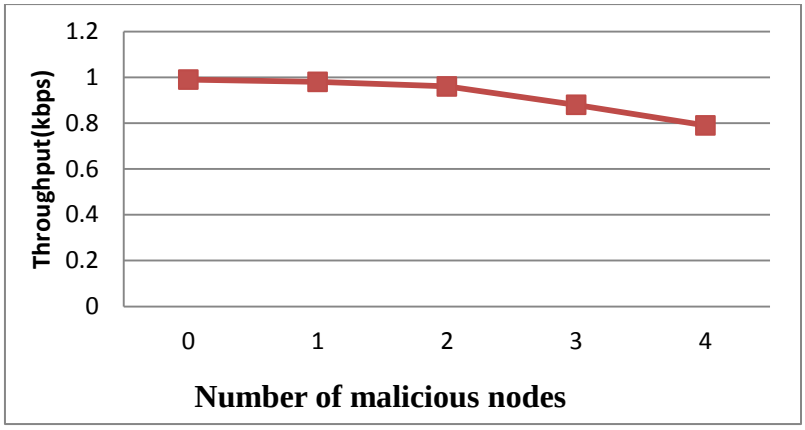


Fig 6.2 Effect of PDR with increase in malicious nodes

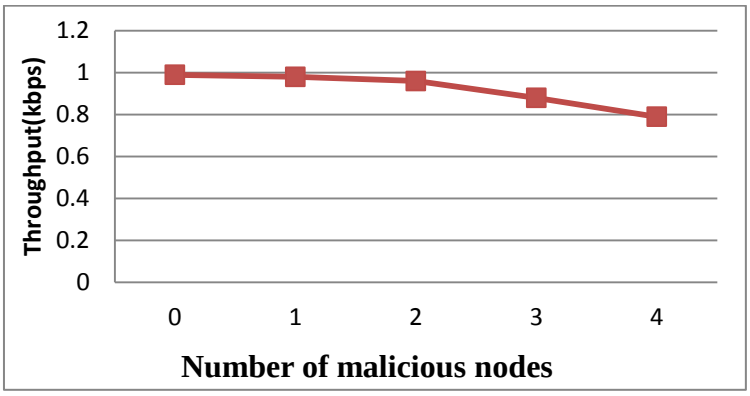


Fig 6.3 Effect of throughput with increase in malicious nodes

The performance metrics that we have used to evaluate the intrusion detection system are Accuracy (fraction of correctly marked records), Precision (ratio of positive predicted values), Specificity (rate of true negative), Sensitivity (rate of true positive), and False Positive and False Negative rate. These performance metrics can be calculated with the help of following four parameters as mentioned in equation 1 to 6 [16].

- True Positive (TP) is defined as the number of normal records which are marked as normal
- False Positive (FP) is defined as number of normal records which are marked as abnormal
- True Negative (TN) is defined as number of abnormal records which are marked abnormal
- False Negative (FN) is defined as number of abnormal records which are marked as normal

$$\text{Accuracy}=\frac{Number\ of\ correctly\ marked\ records}{Total\ number\ of\ records} \quad (1)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (2)$$

$$\text{Specificity}=\frac{TN}{TN+FP} \quad (3)$$

$$\text{Sensitivity}=\frac{TP}{TP+FN} \quad (4)$$

$$\text{False Positive Rate}=\frac{FP}{FP+TN} \quad (5)$$

$$\text{False Negative Rate}=\frac{FN}{FN+TP} \quad (6)$$

6.3 Results and Discussion

The implementation of ANN and GA has been performed in two scenarios- misuse and anomaly and the results for the same have been noted. The whole dataset is split into ten equal subsets with 10% records set for testing in each dataset. The average of all these results has been taken into consideration to avoid any biased results. The genetic algorithm has been implemented for feature sub-selection. The final number of features left in the system is 18.

6.3.1 Results for misuse detection

During the training and testing phase the same dataset is used for misuse detection. As it can be seen, the performance metrics in misuse detection gives remarkable results with all values greater than 0.99 as shown in Table 6.4. Also the false positive rate comes to 0 which is remarkable as false positive affects the performance of intrusion detection system and is one of the factors for evaluating intrusion detection system.

Table 6.4 Misuse Detection Results

Precision	1
Specificity	1
Sensitivity	0.99
Accuracy	0.99
F_measure	1
False Positive Rate	0

6.3.2 Results for anomaly detection

During anomaly detection, there is usage of different dataset in training and testing phase. As evident in the results, the false positive rate is very less and thus can be considered as remarkable outcome. Also the other performance metric shows promising results as shown below in Table 6.5.

Table 6.5 Anomaly Detection Results

Precision	0.97
Specificity	0.97
Sensitivity	0.99
Accuracy	0.95
F_measure	0.98
False Positive Rate	0.03

During the genetic algorithm implementation for selecting subset of feature the top five performing features are listed in Table 6.6

Table 6.6 Top five performing feature

Feature Number	Description
F2	Simulation time
F9	MAC
F12	Source IP and port number
F13	Destination IP address and port number
F18	Time to live

6.3.3 Comparative Analysis

The following table compares our approach with various other researchers' work

Table 6.7 Comparison of results with other proposed protocols

Algorithms	Precision	Specificity	Sensitivity	Accuracy	F_measure	False Positive
Proposed approach	1	1	0.99	0.99	1	0.01

ANN based driverless car [26]	NA	.87	.98	NA	NA	0.12
ANN and Decision Tree based IDS [27]	NA	NA	NA	0.95	NA	0.10
ANN based IDS [24]	1	1	0.96	0.98	NA	NA
SNORT	0.96	0.97	0.9	0.93	NA	NA
GA and MLP based IDS [15]	1	NA	0.99	NA	0.99	0.03

As evident in Table 6.7, our proposed algorithm shows better results than the previous existing algorithms in terms of all performance metric calculated. This improvement in results is due to GA which reduced the features to 18. Moreover our approach made use of self created data set. Most of the previous algorithms make use of 20 to 21 features and takes readily available datasets which has the redundant data. The graph based comparison of accuracy of our approach with other existing algorithms is shown in Fig 6.4.

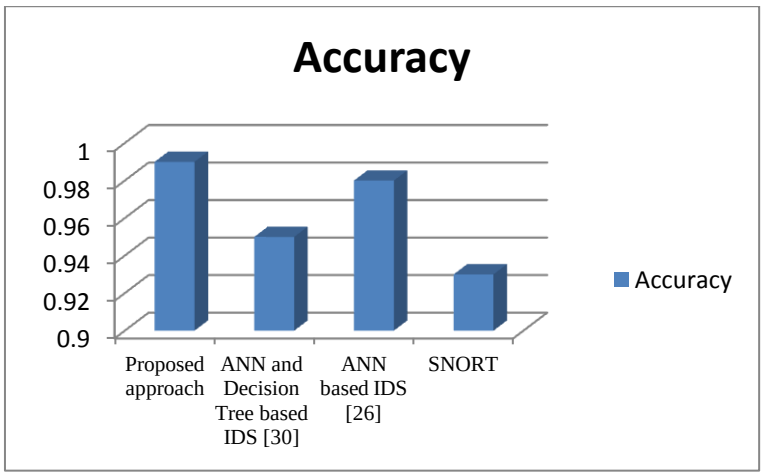


Fig 6.4 Accuracy graph of different protocol

CHAPTER 7

CONCLUSIONS AND FUTURE SCOPE

7.1Conclusion

As security is indispensable part of wireless networks especially vehicular networks, there was a need to tackle the threats which could arise in vehicular networks. One of the most serious security threats is flooding in which the legitimate user is denied to get the service due to unavailability of resources. In our thesis we have launched the flooding attack in ns-2 and evaluated the packet delivery and throughput of the network. There was sharp fall in packet delivery and throughput. We proposed our algorithm based on ANN to detect the attack and further applied GA for feature sub-selection to obtain better results under two different scenarios misuse and anomaly. Our proposed algorithm can detect multiple malicious nodes with higher accuracy as compared with existing approaches. The accuracy of our system came to 99%. Moreover the number of features was reduced to 18. There is no need for any hardware, hence simple and cost effective.

7.2 Future Scope

Research is never ending process. This thesis showed remarkable results but the future scope lies in detecting the attacks with encrypted malicious entries. Also the data set used is specifically for flooding attack. We would like to extend our proposed algorithm to make it more generic by adding more records of other attacks as well. There could be use of other techniques as well to reduce the number of features to even lesser than 18.

REFERENCES

[1] R. S. Raw, M. Kumar, and N. Singh, "Security challenges, issues and their solutions for VANET". *International Journal of Network Security & Its Applications*, vol 5, No. 5, pp. 95, 2013.

[2] B. Patel and K. Shah, "A survey on vehicular ad hoc networks". *IOSR Journal of Computer Engineering (IOSR-JCE) vol* 15,No.4, pp.34–42, 2013.

[3] S. Dhankhar and S. Agrawal," Vanets: A survey on routing protocols and issues."*International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, vol3, No.6, pp.13427–13435, 2014.

[4] M. Erritali and B. El. Ouahidi, "A review and classification of various VANET intrusion detection systems". In *Security Days (JNS3), 2013 National*, Rabat pp. 1–6. IEEE, 2013.

[5] N.Sen, R.Sen, and M. Chattopadhya, "An effective back propagation neural network architecture for the development of an efficient anomaly based intrusion detection system." In *Computational Intelligence and Communication Networks (CICN), 2014 International Conference Bhopal*, pp. 1052–1056. IEEE, 2014.

[6] S. Selim, M. Hashem, and T. M. Nazmy, "Intrusion detection using multi-stage neural network". *International Journal of Computer Science and Information Security*, vol8, No.4, pp.14-20, 2010.

[7] F. Barani,. "A hybrid approach for dynamic intrusion detection in ad hoc networks using genetic algorithm and artificial immune system". In *Intelligent Systems (ICIS), 2014 Iranian Conference Bam*, pp. 1–6, IEEE, 2014.

[8] S. BENQDARA, et al. "Ensemble of clustering algorithms for anomaly intrusion detection system". *Journal of Theoretical and Applied Information Technology*, vol 70, No.3, pp.425-431, 2014.

[9] A.S. Subaira and P. Anitha. "Efficient classification mechanism for network intrusion detection system based on data mining techniques: a survey". In

Intelligent Systems and Control (ISCO), 2014 IEEE 8th International Conference on, pp. 274–280. Coimbatore IEEE, 2014.

[10] N. El. Moussaid and A. Toumanari, "Overview of intrusion detection using data-mining and the features selection". In *Multimedia Computing and Systems (ICMCS), 2014 International Conference* Marrakech, pp 1269–1273. IEEE, 2014.

[11] P Sivaranjanadevi, et.al. "An effective intrusion system for mobile ad hoc networks using rough set theory and support vector machine. *IJCA Proceedings on EGovernance and Cloud Computing Services-2012*, (2) pp.1–7, 2012.

[12] B. Paul and M. J. Islam. "Survey over VANET routing protocols for vehicle to vehicle communication". *IOSR Journal of Computer Engineering (IOSRJCE), ISSN*, pp: 2278–0661, 2012.

[13] R. Mitchell and R. Chen, "A survey of intrusion detection in wireless network applications". *Computer Communications*, vol 1, No.42, pp:1–23, 2014.

[14] D. Pal and A. Parashar, "Improved genetic algorithm for intrusion detection system". In *Computational Intelligence and Communication Networks (CICN), 2014 International Conference Bhopal*, pp: 835–839. IEEE, 2014.

[15] M. Barati, et.al, "Distributed denial of service detection using hybrid machine learning technique". In *Biometrics and Security Technologies (ISBAST), 2014 International Symposium Kuala Lumpur*, pp 268–273. IEEE, 2014.

[16] R. Shanmugavadivu and N. Nagarajan., "Network intrusion detection system using fuzzy logic." *Indian Journal of Computer Science and Engineering (IJCSE)*, vol 2, No.1, pp:101– 111, 2011.

[17] A. Goyal and C. Kumar,. "GA_NIDS: a genetic algorithm based network intrusion detection system". *Northwestern university Evaston*, 2008.

[18] G. Samara, W. AH Al-Salihy, and R. Sures, "Security issues and challenges of vehicular ad hoc networks (VANET)". In *New Trends in Information Science and Service Science (NISS), 2010 4th International Conference* Gyeongju, pp: 393–398. IEEE, 2010.

[19] P. Tyagi and D. Dembla, “Investigating the security threats in vehicular ad hoc networks (VANETs): Towards security engineering for safer on-road transportation.” In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference New Delhi*, pp: 2084–2090. IEEE, 2014.

[20] C. Intanagonwiwat, R. Govindan, and D. Estrin. “Directed diffusion: a scalable and robust communication paradigm for sensor networks”. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pp: 56–67. ACM, 2000.

[21] A. Tajbakhsh, M. Rahmati, and A. Mirzaei., “Intrusion detection using fuzzy association rules”. *Applied Soft Computing*, vol 9,No.2, pp:462–469, 2009.

[22] M.S. Hoque, M. Mukit, Md Bikas. “An implementation of intrusion detection system using genetic algorithm”. *International Journal of Network Security & Its Applications (IJNSA) vol 4, No.2, pp:109-120, 2012*

[23] S. E. Benaicha et.al, “Intrusion detection system using genetic algorithm”. In *Science and Information Conference (SAI), 2014*, London pp: 564–568. IEEE, 2014.

[24] A. Saied, R. E. Overill, and T. Radzik,. “Detection of known and unknown ddos attacks using artificial neural networks”. *Neurocomputing*, vol 172, pp:385–393, 2016.

[25] B. Panja, O. Ogunyanwo, and P. Meharia, “Training of intelligent intrusion detection system using neuro fuzzy”. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2014 15th IEEE/ACIS International Conference Las Vegas*, pp: 1–6. IEEE, 2014.

[26] K. M. A. Alheeti, A. Gruebler, and K. D. McDonald-Maier, “An intrusion detection system against malicious attacks on the communication network of driverless cars”. In *Consumer Communications and Networking Conference (CCNC), 2015 12th Annual IEEE*, Las Vegas, pp: 916–921. IEEE, 2015.

[27] S. Selim, M. Hashem, and T. M. Nazmy,. “Hybrid multi-level intrusion detection system”. *International Journal of Computer Science and Information Security*, vol 9, No.5, pp::23, 2011.

[28] E. Balkanli, J. Alves, and A. N. Zincir-Heywood. “Supervised learning to detect ddos attacks. In *Computational Intelligence in Cyber Security (CICS), 2014 IEEE Symposium*, Orlando, pp 1–8. IEEE, 2014.

[29] L. Ying, Z. Yan, and O. Yang-jia,. “The design and implementation of host-based intrusion detection system”. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium*, Jinggangshan pp: 595–598. IEEE, 2010.

[30] G Sandhya and A. Julian,. “Intrusion detection in wireless sensor network using genetic k-means algorithm”. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on*, Ramanathapuram ,pp: 1791–1794. IEEE, 2014.

[31] P. Jongsuebsuk, N. Wattanapongsakorn, and C. Charnsripinyo,. “Network intrusion detection with fuzzy genetic algorithm for unknown attacks”. In *Information Networking (ICOIN), 2013 International Conference on*, Bangok, pp: 1–5. IEEE, 2013.

[32] A. Daeinabi and A. G. Rahbar. “Detection of malicious vehicles (DMV) through monitoring in vehicular ad-hoc networks”. *Multimedia tools and applications*, vol 66, No.2, pp:325–338, 2013.

[33] T. Zhou,et.al, “P2dapsybil attacks detection in vehicular ad hoc networks”. *Selected Areas in Communications, IEEE Journal*, vol 29,No.3, pp:582–594, 2011.

[34] T. Zhou et.al, “Privacy-preserving detection of sybil attacks in vehicular ad hoc networks”. In *Mobile and Ubiquitous Systems: Networking & Services, 2007. MobiQuitous 2007. Fourth Annual International Conference on*, Philadelphia , pp: 1–8. IEEE, 2007.

[35] J. Grover, M. S. Gaur, and V. Laxmi, “A novel defense mechanism against sybil attacks in VANET”. In *Proceedings of the 3rd international conference on Security of information and networks*, pp: 249–255. ACM, 2010.

[36] Y. Hao, et.al, "A distributed key management framework with cooperative message authentication in VANETS". *Selected Areas in Communications, IEEE Journal*, vol 29,No.3, pp::616–629, 2011.

[37] http://mathworld.wolfram.com/Levenberg-MarquardtMethod.html

SUPPLEMENT INFORMATION

- **Video Link**

https://www.youtube.com/channel/UC9gaToSHoIo8-vtXvUOMW3g

- **List of Publications**
  - M. Aneja, T. Bhatia, "Artificial Intelligence based IDS in VANETs : A Review" in Proceedings of 2nd IEEE International Conference on Engineering & Technology (ICETECH) 17th & 18th March 2016,Coimbatore,TN.India,2016
  - M. Aneja, T.Bhatia, "Artificial Intelligence based Intrusion Detection System to detect Flooding Attack in VANETs" (Communicated)