CHAPTER 1
INTRODUCTION

1.1 OVERVIEW OF MOBILE AD HOC NETWORK

Mobile Ad hoc Network (MANET) is a self-configuring system of mobile routers linked by wireless links, which consequently combine to form arbitrary topologies. Thus, the network's wireless topology may alter rapidly and unpredictably. However, due to the lack of fixed infrastructure in the links, it becomes complicated to exploit the present routing techniques for network services and provides some high challenges in providing the security of the communication, moreover this is not made effortlessly as the number of demands of network security conflict with the demands of mobile networks, generally due to the nature of the mobile devices such as low power consumption, poor processing load.

Ad-hoc wireless networks are a comparatively new paradigm in multi-hop wireless networking that is increasingly becoming popular and will become an essential part of the computing circumstances, consisting of infra-structured and infrastructure-less mobile networks [Forman G, et.al, 1994]. MANET is an infrastructure-less multi-hop network where each node communicates with other nodes directly or indirectly through intermediate nodes. The credit for growth of ad-hoc network goes to its self-configuring and self-organizing habitant. All nodes in a MANET basically function as mobile routers participating in some routing protocol required for deciding and maintaining the routes. While MANETs [Perkins C.E.et.al,2001] are infrastructure-less, self-organizing, promptly deployable wireless networks, they are extremely suitable for applications involving special outdoor events such as communications in regions with no wireless infrastructure, emergencies, natural calamities, military operations, mine site operations, urgent business meetings and robot data acquisition [Jubin J, et.al, 1987 and Royer E. M, et.al,2001]. In general, routes between nodes in an ad hoc network may include multiple hops and, therefore it is suitable to call such networks “multi-hop wireless ad hoc networks”.

An ad-hoc network is a collection of wireless mobile nodes (or routers) dynamically forming a temporary network without the use of any existing network infrastructure or centralized management. The routers are free to move randomly and organize themselves arbitrarily according to wireless topology, which may change rapidly and impulsively, such a network may operate in a stand-alone manner, or may be connected to the Internet.

Multi-hop, mobility, large network size combined with heterogeneity device, bandwidth, and battery power [Harshit et.al,2010] constraints makes the design of adequate routing protocols a major challenge.

In general some form of routing protocol is necessary in such environment when two host Mobile users will want to communicate in a situation where no fixed wired infrastructure is available. For example, a group of researchers would be in need of connecting to the wide area network, students may need to interact during a lecture, or fire fighters want to connect to an ambulance en route to an emergency prospect. In such situations, a set of mobile hosts with wireless network interfaces may form a temporary network without the aid of any established infrastructure or centralized administration.

A Mobile Ad hoc Network (MANET) group has been created within IETF. The primary focus of this working group is to develop and evolve MANET specifications and it introduces them to the Internet standard track. The goal is to support mobile ad hoc networks with hundreds of routers and solve challenges in this kind of network. A few common challenges that are faced by ad hoc networking are restricted wireless transmission range, concealed terminal problems, and packet losses due to transmitting errors, mobility-induced route changes, and battery constraints. Mobile ad hoc network can reinforce the service area of network access and could provide wireless connectivity into areas with poor or no coverage earlier (e.g., cell edges). Connectivity to wired infrastructure will be provided through multiple gateways with possibly different efficiency and usage. To improve the performance level, the mobile host should have the potentiality to adapt the performance variation, coverage and to switch gateways when it is beneficial.
To embellish the prediction of the best overall performance, a network-layer metric has a better overview of a network. Ad hoc networks convey features like easy connection to access networks, dynamic multi-hop network construction, and direct peer-to-peer communication. The multi-hop property of an ad hoc network needs to be bridged by a gateway to the wired backbone, the gateway should also have a network interface for both type of networks along with be a part of the global routing and the local ad hoc routing.

Users could be benefited from pervasive networks in several ways. User mobility enables users to switch between devices, migrate from one session to more, and still get the same personalized services. The host mobility enables the users devices to move around the networks/maintain connectivity as well as the ability to reach.

Mobile ad-hoc networks can turn out the dream of getting connected "anywhere and at any time" into realism. The typical application example includes a disaster recovery and a military operation. Besides these explicit situations, these networks may equally show better performance in other places.

Figure 1.1: Mobile Ad-hoc Network
Figure 1.1 shows MANETs which does not depend on pre-existing infrastructure or base-station. Network nodes in the MANETs are free to travel randomly. Therefore, the network topology of a MANET may change rapidly and randomly. The nodes themselves should execute all network activities, for instance discovering the topology and delivering data packets, either individually or as a group. Depending upon its application, the structure of a MANET may differ from a small static network with a highly power-constrained to a large-scale, mobile, highly dynamic network. It is a group of autonomous mobile nodes or devices connected through wireless links without the support of a communication infrastructure. The topology of the network transform dynamically as nodes move and the nodes reorganize themselves to enable communications with nodes beyond their immediate wireless communications range by relaying messages from one another, i.e. multi-hop.

Power of the MANET relies on the cooperation of the participating nodes. A MANET will be more powerful when more nodes are encompassed to transfer traffic. Still supporting a MANET is a cost-intensive activity for a mobile node, forwarding packets consumes network-bandwidth, detecting routes, local CPU time, memory and energy. Hence there should be a strong motivation for a node to deny packet forwarding to others, meanwhile using their services to deliver own data.

MANET has various potential applications, which are usually set up in the situations of temporary operations for emergency or if there are no resources to set up elaborate networks. Some conventional examples are emergency rescue-search operations, conventional events, conferences and battlefield communication between moving vehicles and soldiers. With the growing possibilities to meet the new demands of mobile computation, the MANET has a very bright future.

As of their flexibility, MANETS are seen as important components in 4G architecture and ad hoc networking capabilities are believed to form a significant part of overall functionalities of next generation. The application of MANET has become widespread and varied from email to ftp to web services. Some general MANET applications are:
Figure 1.2: Application of MANET

- **Military Environments**: Since it is not possible to install base station in the enemy territories, no hospital zone/terrain, MANET provides communication services, where soldier’s acts like nodes. The required coordination among the soldiers and the military objects can be seen as another application of MANET in military services.

- **Civilian Environments**: MANET finds its use in many civilian activities like gathering-hall / conference-room, boats, taxi/cab network, small aircraft and sports stadium etc.,

- **Emergency Operations**: the uses of MANET, in situations like crowd control, search / rescue operations, disaster recovery and commando operations, the use of mobile ad hoc networks are very much suited. MANET can also be established when conventional infrastructure based communication is damaged due to any calamities.

- **Local Level**: Ad hoc networks can autonomously link an instant and temporary multimedia network using notebook computers or palmtop computers to spread and share information among participants. Another appropriate local level application might be in home networks where devices can communicate directly to exchange information.
1.2 DESIGN ISSUES OF MANET ROUTING PROTOCOLS

1.2.1 Routing Architecture

The routing architecture of a self-organized network can be either hierarchical or flat. In most self-organized networks, the hosts will be acting as an independent router, which implies that routing architecture should practically be flat, that is, each address serves only as an identifier and does not convey any information about one host that is topologically located with respect to another node. In flat self-organized network, the mobility management is not essential because all the nodes are visible to each other via routing protocols.

In, flat routing algorithms such as Destination-Sequenced Distance Vector (DSDV) and Wireless Routing Protocol (WRP), the routing tables have entries to all the hosts in the self-organized network. Nevertheless, a flat routing algorithm does not have a good scalability. The routing overhead increases rapidly, when the network becomes larger. Hence, to control channel reuse spatially (in terms of frequency, time, or spreading code) and reduce routing information overhead, some form of hierarchical scheme should be employed. Clustering is the most common technique employed in hierarchical routing architectures.

The idea behind hierarchical routing is to divide the hosts of a self-organized network into a number of overlapping or disjoint cluster. Here, one node is elected as cluster head for each cluster. This cluster head maintains the membership information for the cluster.

Nodes that are not cluster heads will from this time forth, be referred to as “ordinary nodes.” When an ordinary node wants to send a packet, the node can send the packet to the cluster head that routes the packet toward the destination. Cluster head Gateway Switch Routing (CGSR) and Cluster-Based Routing Protocol (CBRP) [Jiang M, et.al, 1999] belong to this type of routing scheme. Hierarchical routing involves cluster, address, and mobility management.

1.2.2 Unidirectional Links Support

Every existing routing protocol tends to assume that all links are bidirectional. On the other hand, there are a number of factors that can make wireless links unidirectional.
They are as follows:

**Different radio capabilities:** Radios within a network can have different transmit powers or receive sensitivities. This is quite likely in a tactical environment where man-pack and vehicular radios exist.

Vehicular radios, being less constrained by size and weight, typically have 12 decibels (dB) greater transmit power than their man-pack counterparts. Unidirectional links are exceedingly likely in tactical networks.

**Interference:** This is due to either hostile jammers or friendly interference, which may reduce a nearby receiver’s sensitivity. For instance, host A can receive packets from host B as there is very little interference in A’s vicinity. Conversely, B may be in the vicinity of an interference node, and consequently cannot receive packets from A., hence the link between A and B is directed from B to A.

**Message broadcast requirement:** There is an increasing emphasis on the wide area broadcast of messages. Satellite-based transmitters are being used for the downward links, whereas the upward links use alternative paths.

**Mute mode:** An extreme instance, applicable only in tactical mobile networks, when the host cannot transmit due to an impending threat. In such cases, it still receives information; however, it does not participate in bidirectional communications.

**The state of link direction is time varying:** The directional state of the wireless link may be either a persistent or a transient phenomenon. The frequency of transitions and the duration of stay in each state will be a function of existing traffic, terrain, mobility pattern, and energy availability.

### 1.2.3 Usage of Super hosts

All existing routing protocols assume that all mobile hosts have the same properties based on the spirit of a self-organized network as a collection of “equal” peers opportunistically using each other’s services to communicate. Although this is true in some circumstances, there are also situations where the network will include hosts with preponderant bandwidth, guaranteed power supply, and high-speed wireless links. Such hosts are referred to as ‘Super Hosts’. For example, accompanying a military environment
consists of a number of walking soldiers equipped with low-capacity man-pack radios and a few tanks with high-capacity vehicular radios.

Usually, the self-organized networks in this situation have two-level network architecture: Backbone area and Subarea. Backbone area is composed of Super Hosts. In addition, Super Hosts are often assumed to have lower mobility than normal hosts so as to maintain the stability of the backbone. Normal hosts need not make routing decisions. For example, a satellite host (a Super Host) can easily collect the routing information from the normal hosts’ geographical locations, build the routing table, and propagate these routes. The example is just analogous that a person on a stage is likely to have a much better view of the wireless network throughout an auditorium.

1.2.4 Quality of Service (QoS) Routing

Till now most of the routing protocols that have been proposed for ad hoc wireless networks optimized the solution for only one metric: hop distance. So the shortest path is generally preferable. For datagram traffic, shortest path routing may perhaps be sufficient. However, the wireless links in self-organized networks, are naturally scarce and dynamic, make it difficult to perform efficient resource utilization or to execute critical real-time applications in such environments. Based on this consideration, it is necessary to provide QoS routing support to effectively control the total traffic that can flow into the network. QoS routing is a routing mechanism under which paths for flows are determined according to resource availability in the network as well as the QoS requirement of flows. QoS routing means the selection of routes with sufficient resources for the requested QoS parameters. There are two goal for QoS routing. The first goal is to meet the QoS requirements for each admitted connection, and the second is to achieve global efficiency in resource utilization.

Thus, QoS routing will consider multiple constraints, and provide better load balance by allocating traffic on different paths, subject to the QoS requirement of different traffic. On the contrary, current routing protocols seem to favour the Routing traffic based on the shortest path, thereby causing a bottleneck. In a self-organized network, many metrics have been considered: (1) the most reliable path, (2) the most stable path, (3) the maximum total power remained path, (4) the maximum available bandwidth path, and so forth. It is desirable to select the routes with a minimum cost.
based on the above metrics and it’s not unlikely to provide just the shortest path based on the hop distance.

1.2.5 Multicast Support

Multicast routing is a network-layer function that constructs long path, which distributes data packets from one source to arrive at many, though not all destinations in a communication network. Then, multicast routing sends a single copy of a data packet simultaneously to multiple receivers over a communication link that is shared by the paths to the receivers. The sharing of links in the collection of the paths to receivers implicitly defines a tree used for the distribution of multicast packets. In contrast to unicast routing, multicast routing is very useful and efficient way to support group communication. This is especially the case in self-organized networks where the bandwidth is limited and energy is constrained. In addition, a self-organized network often consists of several cooperative work groups. The deployment of multicast routing in a self-organized network will provide collaborative visualization and multimedia conferencing as well as information dissemination in critical situations such as disaster or military scenarios. Multicast routing in self-organized networks became an active research topic only very recently, and much research has focused on designing the unicast Routing Protocols. However, a self-organized network is better suited for multicast than unicast because of its broadcast characteristics.

1.3 AD HOC ROUTING PROTOCOLS OVERVIEW

Routing protocols in ad hoc networks vary depending on the type of network. In general, ad hoc network routing protocols are classified into three major categories based on the routing information updated mechanism. They are proactive (table driven routing Protocols), reactive (on-demand routing protocols) and hybrid routing protocols. In addition, protocols may also be classified according to the utilization of specific resources, such as power aware routing protocol, load aware routing protocols and so on.

A. Proactive Routing Protocols

Routes to all destinations are maintained by sending periodical control messages. There may be unnecessary bandwidth wastage for sending control packets. Proactive routing protocols are inappropriate for larger networks, as it requires to maintain route
information every node’s routing table. This causes more overhead leads for consumption of more bandwidth. Example: DSDV [Perkins C.E, et.al, 1994].

B. Reactive Routing Protocols

Routes are found when there is a need (on demand), for the reason being it reduces the routing overhead. It does not demand to search and maintain the routes were there is no route request. Reactive routing protocols are very pleasant in the resource-limited environment. However the source node should wait until a route to the destination is discovered. This approach is well suited when the network is static and traffic is very light. Example: DSR, AODV. [Johnson D B, et.al, 1999, Perkins C.E, et.al, 1998].

C. Hybrid Routing

The Ad Hoc network can use the hybrid routing protocols that have the advantage of both proactive and reactive routing protocols to balance the delay and control overhead (in terms of control packages). The complication of all hybrid routing protocols is the complexity in organizing the network essential to the network parameters. The frequent drawback of hybrid routing protocols is that the nodes have high-level topological information, which maintains more routing information, that results to more memory and power consumption.

Protocols are be assumed to operate at unicast, multicast, geocast or broadcast situations. In unicast protocols, one source transmits messages or data packets to a single destination, which is the most accepted operation at any network. The unicast protocols are also the most frequent in ad hoc environment to be developed and they are the basis on which it is possible to construct other type of protocols. Unicast protocol lacks when there is a need to send same message or stream of data to multiple destinations. So there arises an evitable need for multicast protocols.

Multicast routing protocols try to construct a desirable routing tree or a mesh from one source to several destinations. These protocols should have also to coupe-up with information of joining and leaving to a multicast group. The purpose of geocast protocols is to deliver data packets for a group of nodes, which are positioned, at specified geographical area. This kind of protocol can also help to improve the routing procedure by providing location information for route acquisition.
A routing protocol is a protocol that specifies how routers communicate with one another, disseminating information that enables them to select routes between any two nodes on a computer network. Each router has a prior knowledge of networks attached to it directly. The routing protocol shares this information first among immediate neighbors, and then throughout the network. By way of this process, routers gain knowledge of the topology of the network.

There are unicast, single channel protocols, which are uniform or non-uniform. Uniform protocols are divided to topology-based protocols, in where nodes are aware of the topology information of all other nodes in the network or to destination-based protocols, in where nodes only know the preferred next hop to a destination.

One protocol to belong to that topology-based class is GSR (Global State Routing) and the other is DSR (Destination Source Routing). One main difference between these protocols is the scheduling method. GSR is a proactive protocol, which will all the time have the information needed for routing. DSR is on its behalf a reactive protocol, which will obtain needed information only on demand.

To destination-based protocols belong such protocols as DSDV, AODV, TORA, ABR and WRP. The well known difference between e.g. DSDV and AODV is the scheduling method. The DSDV is proactive as is WRP, but AODV, TORA and ABR all are reactive protocols.

To be classified to single channel, non-uniform protocols there are such protocols as ZRP, FSR, OLSR, CEDAR and CBRP. Form these protocols ZRP, FSR, and OLSR belong to neighbor selection protocols, which have a common feature to select network subsets by individual nodes themselves. In partitioning protocols there are some kind of clustering and cluster head selection mechanism. To partitioning protocols belongs e.g. CEDAR and CBRP.

To unicast multi-channel protocols include such protocols as CGSR and Epidemic. CGSR is a nonuniform protocol and Epidemic is a uniform protocol. The Types of routing protocols is shown in Figure 1.3.
1.3.1 AD-HOC ON-DEMAND DISTANCE VECTOR (AODV)

An ad hoc network is the cooperative engagement of a collection of mobile nodes without the required intervention of any centralized access point or existing infrastructure. AODV is a narrative algorithm for the operation of such ad hoc networks, each mobile host operates as a specialize router, and routes are obtained when needed (i.e., on demand with little or no reliance on periodic advertisements). The AODV routing algorithm is relatively suitable for a dynamic self-configuring network as required by users wishing to utilize ad hoc networks. AODV [Perkins C.E, et.al, 1998] provides loop-free routes, even while repairing broken links. For the reason that the protocol does not require global periodic routing advertisements and the demand on the overall bandwidth available to the mobile nodes is significantly to that of those protocols that necessitate such advertisements.
AODV uses symmetric links among neighbouring nodes. It does not attempt to follow path between nodes, when one of the nodes cannot hear the other. Nodes do not lie on active path, they neither maintain any routing information nor participate in any periodic routing table exchange. Furthermore, a node does not have to ascertain and maintain a route to another node until the two needs to communicate unless the former node is offering its services as an intermediate forwarding station to maintain connectivity between two other nodes.

When the local connectivity of the mobile node is of interest, each mobile node can be aware of the other nodes in its neighbourhood by the use of several techniques, including local (not system wide) broadcasts known as Hello messages. The routing table of the nodes within the neighbourhood are organized to optimize response time to local movements and provide quick response time for requests for establishment of new routes.

The primary objectives of the algorithm are as follows:

Broadcast the discovery packets, only when necessary. AODV uses a broadcast route discovery mechanism, which is also required for modifications in the DSR [Johnson D B, et.al, 1999] algorithm. Instead of source routing, AODV relies on dynamically establishing route table entries at intermediate nodes. Thus, difference pays off in networks with many nodes where a larger overhead is incurred by carrying source routes in each data packet.

To maintain the most recent routing information between nodes, the idea of destination sequence numbers from DSDV is used. Unlike in DSDV, however, each ad hoc node maintains a monotonically increasing sequence number counter, which is used for superseding stale cached routes. The combination of these techniques yields an algorithm that uses bandwidth efficiently by minimizing the network load to control, data traffic is liable for changes in topology and ensures loop-free routing.

1.3.1.1 Routing request

<table>
<thead>
<tr>
<th>Source Address</th>
<th>Request ID</th>
<th>Source Sequence No.</th>
<th>Destination Address</th>
<th>Destination Sequence No.</th>
<th>Hop Count</th>
</tr>
</thead>
</table>

**Figure 1.4: Routing request of AODV**
Figure 1.4 illustrates the routing request of AODV. When a route is not available for the destination, a Route Request packet (RREQ) is flooded right through the network. The RREQ contains the following fields. The request ID is incremented each time as the source node sends a new RREQ, so the pair (source address, request ID) identifies a RREQ distinctively. On receiving a RREQ message each node checks the source address and the request ID. Provided if the node has already received a RREQ with the same pair of parameters the new RREQ packet will be discarded or else the RREQ will be either forwarded (broadcast) or replied (unicast) with a RREP message.

The number of RREQ messages that a node can send per second is limited. There is an optimization of AODV used when flooding RREQ messages. Every RREQ carries a Time to Live (TTL) value that specifies the number of times this message should be re-broadcasted.

This value is set to a predefined value at the first transmission and is increased at retransmissions. Retransmissions occur only if no replies are received. Historically such flooding is used by TTL vast enough, better than the diameter of the network to reach all nodes in the network, so as to guarantee successful route discovery in only one round of flooding. Besides this low delay time approach causes high overhead and unnecessary broadcast messages. The minimal cost flooding search problem can be solved via a sequence of flooding with an optimally chosen set of TTLs.

1.3.1.2 Routing reply

Figure 1.5 illustrates the routing reply of AODV. If a node is the destination, or has a valid route to the destination, it unicasts a route reply message (RREP) back to the source. Thus the message has the following format.

<table>
<thead>
<tr>
<th>Source Address</th>
<th>Destination Address</th>
<th>Destination Sequence No.</th>
<th>Hop Count</th>
<th>Life time</th>
</tr>
</thead>
</table>

Figure 1.5: Routing reply of AODV

The reason one can unicast RREP back, is that every node forwarding a RREQ message caches a route back to the source node.
1.3.1.3 Path Discovery

The path discovery process is initiated whenever a source node needs to communicate with another node for which it has no routing information in its table. Each node maintains two separate counters, a node sequence number and a broadcast ID. The source node initiate path discovery by broadcasting a Route Request (RREQ) packet to its neighbours.

When an intermediate node receives an RREQ, it has already received an RREQ with the same broadcast ID and source address, it leaves the redundant RREQ and does not rebroadcast it. If, a node cannot satisfy the RREQ, it maintains track of the following information to implement the reverse-path setup as well as the forward path setup that accompany the transmission of the eventual RREP. Destination IP address, Source IP address, Broadcast ID, Expiration time for reverse-path route entry, Source node’s sequence number, and Time To Live.

1.3.1.4 Reverse-Path Setup

Two sequence numbers are (in addition to the broadcast_id) included in an RREQ: the source sequence number and the last destination sequence number known to the source. The source sequence number is used for maintaining the freshness of information about the reverse route to the source and the destination sequence number specifies how fresh a route to the destination must be before it can be accepted by the source. As the RREQ travels from a source to various destinations, it automatically sets up the reverse path from all nodes back to the source. To set up a reverse path, a node records the address of the neighbour from which it received.

1.3.1.5 Merits of AODV

- The AODV routing protocol does not need any central administrative system to control the routing process. Reactive protocols like AODV tend to reduce the control traffic messages overhead at the cost of increased latency in finding new routes.
- AODV reacts relatively fast to the topological changes in the network and updates only the nodes affected by these changes.
• The HELLO messages supporting the routes maintenance are range-limited, so they do not cause unnecessary overhead in the network.

1.3.1.6 Drawbacks of AODV

• It is possible that a valid route is expired. Determining a reasonable expiry time is difficult, because the nodes are mobile, and sources sending rates may differ widely and can change dynamically from node to node.

• Moreover, AODV can gather only a very limited amount of routing information; route learning is limited only to the source of any routing packets being forwarded.

• This causes AODV to rely on a route discovery flood more often, which may carry significant network overhead. Uncontrolled flooding generates many redundant transmissions, which may cause so-called broadcast storm problem.

1.3.1.7 Properties

The advantage with AODV compared to classical routing protocols like distance vector and link-state is that AODV has greatly reduced the number of routing messages in the network. AODV achieves this by using a reactive approach. This is probably necessary in an ad-hoc network to get reasonable performance when the topology is changing often. AODV is also routing in the more traditional sense compared to instance source routing based proposals like DSR. The advantage with a more traditional routing protocol in an ad-hoc network is that connections from the ad-hoc network to a wired network like the internet is most likely easier.

The sequence numbers that AODV uses represents the freshness of a route and is increased when something happens in the surrounding area. The sequence prevents loops from being formed, but can however also be the cause for new problems. This can happen when the network becomes partitioned, or the sequence numbers wrap around.

AODV only support one route for each destination. It should however be fairly easy to modify AODV, so that it supports several routes per destination. Instead of requesting a new route when an old route becomes invalid, the next stored route to that destination could be tried.
1.3.2 **802.11 PROTOCOL**

The IEEE 802 standards committee defines two separate layers, the Logical Link Control (LLC) and media access control, for the Data-Link layer of the OSI model. The IEEE 802.11 wireless standard defines the specifications for the physical layer and the media access control (MAC) layer communicates up to the LLC layer, as shown in the following Figure 1.6

![Figure 1.6: 802.11 and OSI Model](image)

### 1.3.2.1 The MAC Layer

The MAC Layer defines two different access methods, the Distributed Coordination Function and the Point Coordination Function:

### 1.3.2.2 CSMA/CA method

The basic access mechanism, called the Distributed Coordination Function, is basically a Carrier Sense Multiple Access with Collision Avoidance mechanism (usually known as CSMA/CA). CSMA protocols are well-known in the industry, the most popular being the Ethernet, which is a CSMA/CD protocol (CD standing for Collision Detection). A CSMA protocol works as follows: A station desiring to transmit senses the medium. If the medium is busy (i.e. some other station is transmitting) then the station defers its transmission to alter time. If the medium is sensed free then the station is allowed to transmit.
These kinds of protocols are very effective when the medium is not heavily loaded since it allows stations to transmit with minimum delay. But there is always a chance of stations simultaneously sensing the medium as being free and all the components in the 802.11 architecture fall into either the Media Access Control (MAC) sub layer of the data-link layer or the physical layer. If Transmission occurs at the same time, it causes a collision. These collision situations must be identified so the MAC layer can retransmit the packet by itself and not by upper layers, which would cause significant delay.

In the Ethernet case, this collision is recognized by the transmitting stations which go into a retransmission phase based on an exponential random back off algorithm. While these Collision Detection[Seema,2012] mechanisms are a good idea on a wired LAN, they cannot be used on a Wireless LAN environment for two main reasons:

1. Implementing a Collision Detection Mechanism would require the implementation of a Full Duplex radio capable of transmitting and receiving at once, an approach that would increase the price significantly.
2. In Wireless environment it cannot be assumed that all stations hear each other (which is the basic assumption of the Collision Detection scheme), and the fact that a station wants to transmit and senses the medium as free doesn’t necessarily mean that the medium is free around the receiver area.

In order to overcome these problems, the 802.11 uses a Collision Avoidance (CA) mechanism together with a Positive Acknowledge scheme, as follows:

- A station wanting to transmit senses the medium. If the medium is busy then it defers. If the medium is free for a specified time (called Distributed Inter Frame Space (DIFS) in the standard), then the station is allowed to transmit.
- The receiving station checks the CRC of the received packet and sends an acknowledgment packet (ACK). The receipt of acknowledgment indicates to the transmitter that no collision occurred. If the sender does not receive the acknowledgment then it retransmits the fragment until it receives acknowledgment or is thrown away after a given number of retransmissions.

1.3.2.3 802.11 MAC Frame

There are two types of proposals for an MAC algorithm: Distributed access protocols, which, like Ethernet, distribute the decision to transmit over all the nodes using
a carrier-sense mechanism; and centralized access protocols. Figure 1.7 illustrates the frame format of MAC. It contains the following fields.

**FC** - Frame Control

**SC** - Sequence Control

**Oct.** - Octets

**D/I** - Duration/Connection control

**FCS** - Frame Checks Sequence

---

**Figure 1.7: MAC Frame Format**

- Frame control indicates the type of frame and provides control information.
- Duration/connection ID indicates the allocation of the time and the channel for the successful transmission of an MAC frame.
- Address field indicates the transmitter and receiver address, SSID and source & destination address.
- Sequence control is used for fragmentation and reassembly.
  - Protocol version: so far only 1 version (number 0).
  - Type and subtype: due to different classes of frames.
  - To DS and From DS bits: indicate if frame destined/from to/a distribution system. For infrastructure NTs equals to “1”.
  - More fragments bit: it equals to “1” in any non-final fragments of a fragmented frame. Otherwise it set to “0”.
  - Retry bit: any retransmitted frame sets it to “1”.
  - Power management bit: (after completion of current atomic exchange) “1” means the devices are in power-saving mode, “0” marks that it is active.

---
(APs not allowed to power-save mode because perform management functions).

- More data bit: APs buffer frames from distribution system towards power-saving mode stations.

- WEP bit: set to “1” when the frame is encrypted.

- Order bit: set to “1” when fragments and frames transmitted in order.

- When bit 15 equals to “0”: duration field used to set Network Allocation Vector (NAV). All the stations monitor the header of received frames and update NAV.

- During contention-free periods: bit14 =0, bit15 =1, all other bits =0 field value = 32768 interpreted as Network Allocation Vector (NAV). Any station that did not hear Beacon (announce contention free period) updates NAV avoiding collision.

- In PS-Poll frames bit14 and bit15 =0. For battery saving stations turn antennas off. When they wake up periodically they transmit a PS-Poll and check if any buffered data is available from AP. An association ID (AID, identifies to which BSS they belong to) is included.

- Four address fields: 48 bits long (as in Ethernet); not always used; not for all the MAC frames.

- Different types of addresses (48-bit MAC identifier):
  - Destination address (i.e, final destination).
  - Source address (i.e, originated frame).
  - Receiver address.
  - Transmitter address.
  - Basic Service Set ID (BSSID).
  - Rule of thumb: address1 used for the receiver; address2 used for the transmitter; address 3 used for filtering by the receiver (frames discarded from a BSS other than the associated one).
• When the first bit towards medium is “0” the address represents a single station (unicast).
• When the first bit is “1” the address represents a group of physical stations (multicast).
• If all bits are “1” the frame is a broadcast.

- Used for de-fragmentation and discarding duplicate frames. It includes:
  - 4-bit fragment number field: associated to fragments. It has value “0” for the first fragment. The successive ones increase it by 1.
  - 12-bit sequence number field: associated to higher-level frames. It operates as modulo-4096 counter of the frames transmitted. It begins as “0”, increments by “1” for each higher-level packet handled by the MAC. Retransmitted frames or fragments keep the same sequence number and moves the higher-layer payload from station to station.
  - 802.11 can transmit frames with a maximum payload of 2312 bytes of higher-level data (8 bytes header of 802.2 LLC + 2296 bytes of NT payload + WEP overhead).
  - Empty field for control and management frames.
  - Allows checking the integrity of received frames (all the fields included into FCS).
  - Recalculated by APs because the MAC address in 802.11 differs from 802.3.

The Sequence Control field contains two subfields, the Fragment Number field and the Sequence Number field, as shown in the following figure 1.8.

<table>
<thead>
<tr>
<th>12 Bits</th>
<th>4 Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence Number</td>
<td>Fragment Number</td>
</tr>
</tbody>
</table>

**Figure 1.8: Sequence Control Field**

A description of each Sequence Control field subfield is as follows:
- **Sequence Number** indicates the sequence number of each frame. The sequence number is the same for each frame sent for a fragmented frame; otherwise, the number is incremented by one until reaching 4095, where it begins at zero again.

- **Fragment Number** indicates the number of each frame sent by fragmented frame. The initial value is set to 0 and then incremented by one for each subsequent frame sent for a fragmented frame.

At the physical (PHY) sub layer, IEEE 802.11 defines a series of encoding and transmission schemes for wireless communications, the most common of which are the Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), and Orthogonal Frequency Division Multiplexing (OFDM) transmission schemes.

### 1.3.3 DYNAMIC SOURCE ROUTING (DSR)

The DSR Protocol [Johnson D B, et al., 1999] is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes. Using DSR, the network becomes completely self-organizing and self-configuring, requiring no existing network infrastructure or administration. Network nodes (computers) cooperate to forward packets for each other to allow communication over multiple “hops” between nodes not directly within wireless transmission range of one another. As nodes in the network move about or join or leave the network, and as wireless transmission conditions such as sources of interference change, all routing is automatically determined and maintained by the DSR Routing Protocol. Because the number or sequence of intermediate hops needed to reach any destination may change at any time, the resulting network topology may be quite rich and rapidly changing. The DSR Protocol allows nodes to dynamically discover a source route across multiple network hops to any destination in the ad hoc network.

Each data packet sent then carries in its header the complete, ordered list of nodes through which the packet must pass, allowing packet routing to be trivially loop-free and avoiding the need for up-to-date routing information in the intermediate nodes through which the packet is forwarded. By including this source route in the header of each data packet, other nodes forwarding or overhearing any of these packets may also easily cache this routing information for future use.
While designing DSR, it is needed to create a routing protocol that has very low overhead yet is capable enough react quickly to changes in the network, providing highly reactive service to help ensure successful delivery of data packets in spite of node movement or other changes in network conditions.

An overview of the Dynamic Source Routing (DSR) protocol is provided as an example on-demand routing protocol, on which is based the framework for ensuring cache freshness is based upon. In general, the proposed framework can be applied to any routing protocol employing some form of route cache. DSR is a source routing protocol, in which source of a packet specifies complete route in the packet header before forwarding packet to the next hop. The DSR protocol consists of two components, route discovery and route maintenance, both of which operate entirely on-demand. In route discovery a node searches for a route to the destination by propagating RREQ packet in the network. A node receiving RREQ packet performs route cache lookup and sends RREP back to the sender if its route cache has route to the destination. However, if an intermediate node, which has received RREQ, does not have route in its route cache, cannot broadcast RREQ further in the network. When a node gets Route Reply (RREP) either by an intermediate node or by the destination itself, it stores route information in its route cache to avoid route discovery for the same route in near future. Route discovery of DSR is shown in figure 1.9.

![Figure 1.9: Route Discovery of DSR](image)

When some node S originates from a new packet destined to some other node D, it places in the header of the packet a “source route” giving the sequence of hops that the packet should follow on its way to D. Normally, S will obtain a suitable source route by searching its “route cache” of routes previously learned, but if no route is found in its cache, it will initiate the route discovery protocol to dynamically find a new route to D.
In this case, it is called S the “initiator” and D the “target” of the route discovery. For example, Figure 1.9 illustrates an example of route discovery, in which a node A is attempting to discover a route to node E. To initiate the route discovery, A transmits a route request message as a single local broadcast packet, which is received by (approximately) all the nodes currently within the wireless transmission range of A. Each route request message identifies the initiator and target of the route discovery, and also contains a unique “request ID,” determined by the initiator of the request.

Each route request also contains a record listing the address of each intermediate node through which this particular copy of the route request message has been forwarded. This route record is initialized to an empty list by the initiator of the route discovery.

When another node receives a route request, if it is the target of route discovery, it returns a route reply message to the initiator of the route discovery, giving a copy of the accumulated route record from the route request; when the initiator receives this route reply, it caches this route in its route cache for sending the subsequent packets to this destination. Otherwise, if this node receiving the route request has recently seen another route request message from this initiator bearing the same request ID or if it finds that its own address is already listed in the route record in the route request message, it discards the request.

Otherwise, this node appends its own address to the route record in the route request message and propagates it by transmitting it as a local broadcast packet (with the same request ID).

In returning the route reply to the initiator of the route discovery, such as node E replying back to A, node E will typically examine its own route cache for a route back to A and, if found, will use it for the source route for delivery of the packet containing the route reply. Otherwise, E may perform its own Route maintenance and this is invoked when link failure is detected. When a node detects link failure, it sends RERR packet back to the source. Source node after receiving RERR packet, either uses alternate route present in its route cache, or starts route discovery, if there is no alternate route in its cache. A number of optimizations are proposed for the basic DSR protocol. Two such optimizations that are prone to affectation by this framework are packet salvaging and packet overhearing to learn routing information. Nodes learn routing information by
forwarding packets to other nodes, or by overhearing packet transmissions of other nodes. DSR makes aggressive use of route cache to avoid route discovery.

When originating or forwarding a packet using a source route, each node transmitting the packet is responsible for confirming that the packet has been received by the next hop along the source route; the packet is retransmitted (up to a maximum number of attempts) until this confirmation receipt is received. For example, in the situation illustrated in figure 1.10, node A has originated a packet for E using a source route through intermediate nodes B, C, and D. In this case, node A is responsible for receipt of the packet at B, node B is responsible for receipt at C, node C is responsible for receipt at D, and, finally, node D is responsible for receipt at the destination E.

![Figure 1.10: Route maintenance of DSR](image)

This confirmation receipt in many cases may be provided at no cost to DSR, either as an existing standard part of the MAC protocol in use (such as the link-level acknowledgment frame defined by IEEE 802.11) or by a “passive acknowledgment” (in which B confirms receipt at C by overhearing C transmit the packet to forward it on to D). If neither of these confirmation mechanisms is available, the node transmitting the packet may set a bit in the packet’s header to request that a DSR-specific software acknowledgment be returned by the next hop; this software acknowledgment will normally be transmitted directly to the sending node, but if the link between these two nodes is unidirectional, this software acknowledgment may travel over a different multi-hop path.

If the packet is retransmitted by some hop the maximum number of times and no receipt confirmation is received, this node returns a route error message to the original sender of the packet, identifying the link over which the packet could not be forwarded. If C is unable to deliver the packet to the next hop D, then C returns a route error to A, stating that the link from C to D is currently “broken.” Node A then removes this broken link from its cache; any retransmission of the original packet is a function for upper-layer protocols such as TCP.
For sending such a retransmission or other packets to this same destination E, if A has in its route cache packets), it can send the packet using the new route immediately. Otherwise, it may perform a new route discovery for this target another route to E (for example, from additional route replies from its earlier route discovery, or from having overheard sufficient routing information.

1.3.3.1 Overview and Important Properties of the DSR Protocol

The DSR Protocol is composed of two mechanisms that work together to allow the discovery and maintenance of source routes in the ad hoc network: Route discovery is the mechanism by which a node S wishing to send a packet to a destination node D obtains a source route to D. Route discovery is used only when S attempts to send a packet to D and does not already know a route to D. Route maintenance is the mechanism by which node S is able to detect, while using a source route to D, if the network topology has changed such that it can no longer use its route to D because a link along the route works no longer. When route maintenance indicates that a source route is broken, S can attempt to use any other route it happens to know to reach D, or can invoke route discovery again to find a new route.

Route maintenance is used only when S is actually sending packets to D. Route discovery and route maintenance each operate entirely on demand. In particular, unlike other protocols, DSR requires no periodic packets of any kind at any level within the network. For example, DSR does not use any periodic routing advertisement, link status sensing, or neighbour detection packets, and does not rely on these functions from any underlying protocols in the network. This entirely on-demand behaviour and lack of periodic activity allow the number of overhead packets caused by DSR to scale all the way down to zero, when all nodes are approximately stationary with respect to each other and all routes needed for current communication have already been discovered. As nodes begin to move more or as communication patterns change, the routing packet overhead of DSR automatically scales to only that is needed to track the routes currently in use. In response to a single route discovery (as well as through routing information from other packets overheard), a node may learn and cache multiple routes to any destination.
This allows the reaction to routing changes to be much more rapid, because a node with multiple routes to a destination can try another cached route if the one it uses should fail. This caching of multiple routes also avoids the overhead of needing to perform a new route discovery each time a route in use breaks.

The operations of route discovery and route maintenance in DSR are designed to allow unidirectional links and asymmetric routes to be easily supported. In wireless networks, it is possible that a link between two nodes may not work equally well in both directions, due to differing antenna or propagation patterns or sources of interference. DSR allows such unidirectional links to be used when necessary, improving the overall performance and network connectivity in the system.

DSR also supports internetworking between different types of wireless networks, allowing a source route to be composed of hops over a combination of any types of network available. For example, some nodes in the ad hoc network may have only short-range radios, while other nodes have both short-range and long-range radios; the combination of these nodes can be considered by DSR as a single ad hoc network. In addition, the routing of DSR is integrated into standard Internet routing, where a “gateway” node connected to the Internet also participates in the ad hoc network routing protocols, and is integrated into Mobile IP routing, where such a gateway node also serves the role of a Mobile IP foreign agent.

1.3.3.2 Properties

DSR uses the key advantage of source routing. Intermediate nodes do not need to maintain up-to-date routing information in order to route the packets they forward. There is also no need for periodic routing advertisement messages, which leads to reduce network bandwidth overhead, particularly during periods when little or no significant host movement is taking place. Battery power is also conserved on the mobile hosts, both by not sending the advertisements and by not needing to receive them; a host could go down to sleep instead.

This protocol has the advantage of learning routes by scanning for information in packets that are received.
A route from A to C through B means that A learns the route to C, but also that it learns the route to B. The source route will also mean that B learns the route to A and C and that C learns the route to A and B. This form of active learning is very good and thereby it reduces overhead in the network.

However, each packet carries a slight overhead containing the source route of the packet. This overhead grows when the packet has to go through more hops to reach the destination. So the packets sent will be slightly bigger, because of the overhead. Running the interfaces in promiscuous mode is a serious security issue. Since the address filtering of the interface is turned off and all packets are scanned for information. A potential intruder could listen to all packets and scan them for useful information such as pass words and credit card numbers. Applications have to provide the security by encrypting their data packets before transmission. The routing protocols are prime targets for impersonation attacks and must therefore also be encrypted. One way to achieve this is to use IP-sec [Stephen Kent, et.al, 1998]. DSR also has support for unidirectional links by the use of piggybacking the source route a new request.

1.3.4 DESTINATION SEQUENCED DISTANCE VECTOR (DSDV)

DSDV [Perkins C.E, et.al, 1994] is a proactive unicast mobile ad hoc network routing protocol. Like WRP, DSDV is also based on the traditional Bellman-Ford algorithm. However, its mechanisms to improve routing performance in mobile ad hoc networks are quite different. In routing tables of DSDV, an entry stores the next hop toward a destination, the cost metric for the routing path to the destination, and a destination sequence number that is created by the destination. Sequence numbers are used in DSDV to distinguish stale routes from fresh ones and avoid the formation of route loops. The route updates of DSDV can be either time driven or event driven.

Every node periodically transmits updates, including its routing information, to its immediate neighbours. While a significant change occurs from the last update, a node can transmit its changed routing table in an event-triggered style. Moreover, the DSDV has two ways when sending routing table updates. One is the “full-dump” update type in which the full routing table is included inside the update. An incremental update, in contrast, contains only those entries with metrics that have been changed since the last update was sent. Additionally, the incremental update fits in one packet.
1.3.4.1 Distance vector

In distance-vector algorithms, every node $i$ maintains, for each destination $z$, a set of distances $\{d_{xij}\}$ where $j$ ranges over the neighbours of $i$. Node $i$ treats neighbour $k$ as a next hop for a packet destined for $x$ if $d_{xik}$ equals $\min_j \{d_{xij}\}$. The succession of next hops chosen in this manner leads to $x$ along the shortest path. To keep the distance estimates up-to-date, each node monitors the cost of its outgoing links and periodically broadcasts, to each one of its neighbours, its current estimate of the shortest distance to every other node in the network.

The distance-vector algorithm is the classical Distributed Bellman-Ford (DBF) algorithm. Compared to link state method, it is computationally more efficient, is easier to implement, and requires much less storage space; however, it is well-known that this algorithm can cause the formation of both short-lived and long-lived loops. The primary cause for formation of routing loops is that the nodes choose their next hops in a completely distributed fashion based on information which can possibly be stale and, therefore, incorrect.

Almost all proposed modifications to the DBF algorithm eliminate the looping problem by forcing all nodes in the network to participate in some form of intermodal coordination protocol. Such inter nodal coordination mechanisms might be effective when topological changes are rare.

However, within an ad hoc mobile environment, enforcing any such inter nodal coordination mechanism will be difficult due to the rapidly changing topology of the underlying routing network.

Simplicity is one of the primary attributes which makes any routing protocol preferred over others for implementation within operational networks. Routing Information Protocol (RIP) is a classic example. Despite the counting-to-infinity problem, it has proven to be very successful within small-size internetworks. The usefulness of RIP within an ad hoc environment, however, is limited as it was not designed to handle rapid topological changes. Furthermore, the techniques of split horizon and poisoned-reverse are not useful within the wireless environment due to the broadcast nature of the transmission medium.
Packets are transmitted between the stations of the network by using routing tables that are stored at each station of the network. Each routing table, at each of the stations, lists all available destinations and the number of hops to each. Each route table entry is tagged with a sequence number which is originated by the destination station. To maintain the consistency of routing tables in a dynamically varying topology, each station periodically transmits updates immediately when significant new information is available.

Because it is not assumed that the mobile hosts are maintaining any sort of time synchronization, it is also made no assumption about the phase relationship of the update periods between the mobile hosts. These packets indicate which stations are accessible from each station and the number of hops necessary to reach these accessible stations, as is often done in distant-vector routing algorithms. Any new metrics for route selection other than the freshness of the sequence numbers associated with the route; cost or other metrics might easily replace the number of hops in other implementations. The packets may be transmitted containing either layer-2 (MAC) addresses or layer-3 (network) addresses.

1.3.4.2 Properties

DSDV is dependent on periodic broadcasts. It needs some time to converge before a route can be used. This converging time can probably be considered negligible in a static wired network, where the topology does not change so frequently.

In an ad-hoc network on the other hand, where the topology is expected to be very dynamic, this converging time will probably mean a lot of dropped packets before the detection of a valid route. The periodic broadcasts also add a large amount of overhead into the network.

1.3.5 ZONE ROUTING PROTOCOL (ZRP)

The ZRP Protocol combines the advantages of the proactive and reactive approaches by maintaining an up-to-date topological map of a zone centred on each node. Within the zone, routes are immediately available. For destinations outside the zone, ZRP employs a route discovery procedure, which can benefit from the local routing information of the zones.
Proactive routing uses excess bandwidth to maintain routing information, while reactive routing involves long route request delays. Reactive routing also inefficiently floods the entire network for route determination. ZRP [Haas, et al., 1998] aims to address the problems by combining the best properties of both approaches. ZRP can be classed as a hybrid reactive and proactive routing protocol. In an ad hoc network, it can be assumed that the largest part of the traffic is directed to the neighbouring nodes. Therefore, ZRP reduces the proactive scope to a zone centered on each node. In a limited zone, the maintenance of routing information is easier.

Further, the amount of routing information that is never used is minimized. Still, nodes that are farther can be reached with reactive routing. Because all nodes proactively store local routing information, route requests can be more efficiently performed without querying all the network nodes. Despite the use of zones, ZRP has a flat view over the network. In this way, the organizational overhead related to hierarchical protocols can be avoided.

Hierarchical routing protocols depend on the strategic assignment of gateways or landmarks, so that every node can access all levels, especially the top level. Nodes belonging to different subnets must send their communication to a subnet that is common to both the nodes.

This may congest parts of the network. ZRP can be categorized as a flat protocol because the zones overlap. Hence, optimal routes can be detected and network congestion can be reduced. The behaviour of ZRP is adaptive and it depends on the current configuration of the network and the behaviour of the users.

### 1.3.5.1 Properties

ZRP is self adaptive protocol, however this is not done dynamically but instead it is suggested that this zone radius should be set by the administration of the network or with a default value by the manufacturer. The performance of this protocol depends quite a lot on this decision. Since this is a hybrid between proactive and reactive schemes, this protocol use advantages from both. Routes can be found very fast within the routing zone, while routes outside the zone can be found by efficiently querying selected nodes in the network. One problem is however that the proactive intrazone routing protocol is not specified. The use of different intrazone routing protocols would mean that the nodes
would have to support several different routing protocols. This is not a good idea when dealing with thin clients. It is better to use the same intrazone routing protocol in the entire network. ZRP also limits propagation of information about topological changes to the neighborhood of the change only. However, a change in topology can affect several routing zones.

1.3.6 TEMPORALLY ORDERED ROUTING ALGORITHM (TORA)

TORA is a distributed routing protocol for mobile, multihop wireless networks. Its intended use is for the routing of IP datagram’s within an autonomous system. The basic, underlying algorithm is neither a distance vector nor a link state; it is one of a family of algorithms referred to as “link-reversal” algorithms. The protocol’s reaction is structured as a temporally ordered sequence of diffusing computations, each computation consisting of a sequence of directed link reversals. The protocol is highly adaptive, efficient, and scalable, and is well suited for use in large, dense, mobile networks.

In these networks, the protocol’s reaction to link failures typically involves only a localized “single pass” of the distributed algorithm. This desirable behaviour is achieved through the use of a physical or logical clock to establish the “temporal order” of topological change events. The established temporal ordering is subsequently used to structure (or order) the algorithm’s reaction to topological changes.

TORA [Park V.D, et.al, 1997] design is predicated on the notion that a routing algorithm that is well suited for operation in this environment should possess the following properties:

- Executes in a distributed manner.
- Provides loop-free routes.
- Provides multiple routes (i.e., to reduce the frequency of reactions to topological changes, and potentially to alleviate congestion).
- Establishes routes quickly (i.e., so they may be used before the topology changes).

Minimizes communication overhead by localizing algorithmic reaction to topological changes when possible (i.e., to conserve available bandwidth and increase scalability). Routing optimality (i.e., determination of the shortest path) is of less importance. It is also not necessary (or desirable) to maintain routes between every
source–destination pair at all times. The overhead expended to establish a route between a given source–destination pair will be wasted if the source does not require the route prior to its invalidation due to topological changes. TORA is designed to minimize reaction to topological changes. A key concept in its design is that it decouples the generation of potentially far-reaching control message propagation from the rate of topological changes. Control messaging is typically localized to a very small set of nodes near the change without having to resort to a dynamic, hierarchical routing solution with its attendant complexity.

TORA includes a secondary mechanism, which allows far-reaching control message propagation as a means of infrequent route optimization and soft-state route verification. This propagation occurs periodically at a very low rate and is independent of the network topology dynamics. TORA is distributed in that nodes need only to maintain information about adjacent nodes (i.e., one-hop knowledge). It guarantees all routes are loop free, and typically provides multiple routes for any source–destination pair that requires a route.

TORA is “source initiated” and quickly creates a set of routes to a given destination only when desired. Because multiple routes are typically established and having a single route is sufficient, many topological changes require no reaction at all. Following topological changes that do require reaction, the protocol quickly re-establishes valid routes. This ability to initiate and react infrequently serves to minimize communication overhead. Finally, in the event of a network partition, the protocol detects the partition and erases all invalid routes. A logically separate version of TORA is run for each destination to which routing is required. The following discussion focuses on a single version running for a given destination, j. TORA can be separated into three basic functions: creating routes, maintaining routes, and erasing routes. Creating a route from a given node to the destination requires establishment of a sequence of directed links leading from the node to the destination. This function is only initiated when a node with no directed links requires a route to the destination. Thus creating routes essentially corresponds to assigning directions to links in an undirected network or portion of the network.

The method used to accomplish this is an adaptation of the query–reply process, which builds a Directed Acyclic Graph (DAG) rooted at the destination (i.e., the destination is the only node with no downstream links). Such a DAG will be referred to as a “destination-oriented DAG.”
“Maintaining routes” refers to reacting to topological changes in the network in a manner such that routes to the destination are re-established within a finite time meaning that its directed portions return to a destination-oriented DAG within a finite time.

TORA incorporates a new algorithm, in the same general class, that is more efficient in reacting to topological changes and capable of detecting a network partition. This leads to the third function, erasing routes. Upon detection of a network partition, all links (in the portion of the network that has become partitioned from the destination) must be marked as undirected to erase invalid routes.

TORA accomplishes these three functions through the use of three distinct control packets: query (QRY), update (UPD), and clear (CLR). QRY packets are used for creating routes, UPD packets are used for both creating and maintaining routes, and CLR packets are used for erasing routes.

1.3.6.1 Properties

The protocols underlying link reversal algorithm will react to link changes through a simple localized single pass of the distributed algorithm. This prevents CLR packets to propagate too far in the network. The overhead in TORA is quite large. The graph is rooted at the destination, which has the lowest height. However, the source originating the query does not necessarily have the highest height. This can lead to the situation, where multiple routes are possible from the source to the destination, but only one route will be discovered. The reason for this is that the height is initially based on the distance in number of hops from the destination.

1.3.7 Optimized Link State Routing (OLSR)

Optimized Link State Routing [Clausen, et.al,2003] is a proactive protocol that is based on the link state algorithm. OLSR has been modified and optimized to efficiently operate MANET routing. The main concept of the protocol is to adapt the changes of the network without creating control messages overhead due to the protocol flooding nature. Thus, the designers of OSLR decided to have only a subset of nodes, named Multipoint Relays (MPRs), in the network responsible for broadcasting control messages and generating link state information. A second optimization is that every MPR may choose to broadcast link state information only between itself and the nodes that have selected it as an MPR.
1.3.8 Cluster Based Routing Protocol

Cluster Based Routing Protocol (CBRP) is a routing protocol that has a hierarchical-based design [Jiang M, et.al, 1999]. This protocol divides the network area into several smaller areas called cluster. The clustering algorithm of CBRP is Least Cluster Change or LCC [Chiang, et.al, 1997] means the node with the lowest ID among its neighbours elects as cluster head. Other nodes lie into radio range of this cluster head will be the ordinary nodes of that cluster. Because of mobility of nodes in ad hoc network this is probable that elected cluster head to be too mobile. In addition, because nodes with cluster head role consume more power than ordinary nodes, mobile node with lower ID discharge soon. Through these reasons cluster head election procedure used in CBRP is not suitable. The idea of CBRP is to divide the nodes of an Ad hoc network into a number of overlapping or disjoint clusters. Each cluster elects a node as cluster head. A cluster head exerts gateway nodes to communicate with other cluster head through them. In other word a gateway node has at least one cluster head or more. Other nodes in the cluster are ordinary nodes. Cluster head record the membership information for the clusters in two neighbouring tables. CBRP’s clustering algorithm creates clusters that their diameters are 2 hops. Intra-cluster routes (routes within a cluster) are discovered dynamically using the membership information. CBRP is based on source routing, similar to DSR. This means that inter-cluster routes (routes between clusters) are found by flooding. Cluster structure in CBRP network with Route Requests (RREQ). The difference is that the cluster structure generally means that the number of nodes disturbed is much less. Flat routing protocols, i.e. only one level of hierarchy, might suffer from excessive overhead when scaled up.

In CBRP, each node transmits some packets named “Hello message” to announce its presence to its neighbour nodes. Upon receiving a hello message, each node updates its neighbor tables. Each node enters the network in the “undecided” state. Every node upon receiving hello message from its neighbors compares its own ID with its neighbor’s. If a node distinguishes that its own ID is the lowest ID between its neighbors, this node declares itself as cluster head. Every node that has a bi-directional link to this cluster head will be a member of this cluster [Jiang M, et.al, 1999]. Clusters are identified by their respective cluster heads, which means that the cluster head must change as infrequently as possible. The algorithm is therefore not a strict “lowest-ID” clustering algorithm. A non-cluster head never challenges the status of an existing cluster head. Only when two cluster-heads move next to each other, one of them loses its role as cluster head (LCC).
1.4 Properties of Ad hoc Routing Protocols

1.4.1 Distributed operation

The protocol should of course be distributed. It should not be dependent on a centralized controlling node. This is the case even for stationary networks. The difference is that nodes in an ad-hoc network can enter/leave the network very easily and because of mobility the network can be partitioned.

1.4.2 Loop free

To improve the overall performance, the routing protocol must guarantee that the routes supplied are loop-free. This avoids any waste of bandwidth or CPU consumption.

1.4.3 Demand based operation

To minimize the control overhead in the network and thus not wasting network resources more than necessary, the protocol should be reactive. This means that the protocol should only react when needed and that the protocol should not periodically broadcast control information.

1.4.4 Unidirectional link support

The radio environment can cause the formation of unidirectional links. Utilization of these links and also the bi-directional links improves the routing protocol performance.

1.4.5 Security

The radio environment is especially vulnerable to impersonation attacks, to ensure the relevant behavior from the routing protocol; some sort of preventive security measures are needed. Authentication and encryption is probably the way to go and the problem here lies within distributing keys among the nodes in the ad-hoc network.

1.4.6 Power conservation

The nodes in an ad-hoc network can be laptops and thin clients such as PDAs that are very limited in battery power and therefore uses some sort of standby mode to save power. It is therefore important that the routing protocol supports these sleep-modes.
1.4.7 Multiple routes

To reduce the number of reactions to topological changes the congestion multiple routes could be used. If one route has become invalid, it is possible that another stored route could still be valid. The routing protocol is saved from initiating another route discovery procedure.

1.4.8 Quality of service support

Some sort of Quality of Service support is probably necessary to incorporate into the routing protocol. This has a lot to do with what these networks will be used for. It could for instance be real-time traffic support.

None of the proposed protocols from MANET have all these properties, but it is necessary to remember that the protocols are still under development and are probably extended with more functionality.

1.4.9 Load Balancing

The protocol should not overload one node and should be designed to keep the load even on all the nodes. This will also help in avoiding the occurrence of congestion near certain nodes.

1.4.10 Scalability

The performance of the protocol should not be affected by increasing or decreasing the number of nodes in the network.

1.4.11 Sleep function

As nodes in an ad hoc wireless network are energy constrained some nodes may decide to go to sleep (inactive mode) for a random period of time. The protocol should be able to handle such nodes without causing any effect on the rest of the network.

1.4.12 Benchmarking Matrices

The Benchmarking for the performance analysis of routing protocols for wireless ad hoc networks can be obtained by varying the parameters listed below.
1.4.13 Network Size

The Network size is the size of area on which the network is spread off. It may be in meters or kilometres.

1.4.14 Network connectivity

Network Connectivity is defined as the average degree of a node (i.e. the average number of neighbours of a node)

1.4.15 Link capacity

Link Capacity is defined as the effective link speed measured in bits/second, after accounting for losses due to multiple access, coding, framing, etc.

1.4.16 Type of link

Type of link is defined as the pattern of link, i.e. whether it is an Unidirectional or a bi directional link.

1.4.17 Offered load

The network offered load is defined as the ratio between the total network traffic load and the network capacity.

1.4.18 Mobility

Mobility is defined as the circumstances under which and when, is temporal and spatial topological correlation is relevant to the performance of routing protocols.

1.4.19 Fraction of sleep nodes

As a result of energy conservation, or some other need to be inactive, nodes of a MANET may stop transmitting and or receiving (even receiving requires power) for arbitrary time periods. A routing protocol should be able to accommodate such sleep periods without overtly adverse.

1.4.20 Fading models

Fading is deviation of the attenuation that a carrier-modulated telecommunication signal experiences over certain propagation media. The fading may vary with time, geographical position and/or radio frequency, and is often modelled as a random process.
1.4.21 Traffic Pattern

Traffic Pattern is defined as how much effective is a protocol in adapting to non-uniform or burst traffic patterns.

1.4.22 Topology

Network topology is the layout pattern of interconnections of the various elements (links, nodes, etc.) of a computer network. Network topologies may be physical or logical. Physical topology means the physical design of a network including the devices, location and cable installation. Logical topology refers to how data is actually transferred in a network as opposed to its physical design.

1.5 Salient Features of Mobile Ad Hoc Networks

1.5.1 Dynamic Topologies

All nodes in the network are free to move arbitrarily making the network topology to change randomly and rapidly at unpredictable times. The topology may consist of both bidirectional and unidirectional links.

1.5.2 Bandwidth constrained

Variable capacity links wireless links continue to have significantly lower capacity when compared to traditional hardwired links. The realized throughput of wireless communications after accounting for the effects of multiple access, fading, noise and interface conditions, is often much less than a radio’s maximum transmission rate.

1.5.3 Energy constrained operation

Some or all of the nodes in a MANET may rely on batteries or other exhaustible means for their energy. For these nodes, the most important system design criteria for optimization may be energy conservation.

1.5.4 Limited Physical Security

Mobile wireless networks are generally more prone to physical security attacks. Existing link security techniques are often applied to reduce such security threats.
1.5.5 **Autonomous**

No centralized administration entity is required to manage the operation of different mobile hosts.

1.6 **APPLICATIONS OF MANETS**

As MANETs do not require a fixed infrastructure they have a number of benefits and versatility for certain environments and applications scenarios/services.

1.6.1 **Tactical networks**

- Military communication and operations
- Automated battlefields

1.6.2 **Emergency services**

- Search and rescue operations
- Disaster recovery
- Replacement of fixed infrastructure in case of environmental disasters
- Policing and fire fighting
- Supporting doctors and nurses in hospitals

1.6.3 **Commercial and civilian environments**

- E-commerce: electronic payments anytime and anywhere
- Business: dynamic database access, mobile offices
- Vehicular services: road or accident guidance, transmission of road and weather conditions, taxi cab network, inter-vehicle networks
- Sports stadiums, trade fairs, shopping malls
- Networks of visitors at airports

1.6.4 **Home and enterprise**

- Home/office wireless networking
- Conferences, meeting rooms
- Personal area networks (PAN), Personal networks (PN)
- Networks at construction sites
- Universities and campus settings
- Virtual classrooms
- Ad hoc communications during meetings or lectures
- Multi-user games
• Wireless P2P networking
• Outdoor Internet access
• Robotic pets
• Theme parks

1.6.5 Sensor networks
• Home applications: smart sensors and actuators embedded in consumer electronics
• Body area networks (BAN)
• Data tracking of environmental conditions, animal movements, chemical/biological detection.

1.6.6 Context aware services
• Follow-on services: call-forwarding, mobile workspace
• Information services: location specific services, time dependent services
• Infotainment: touristic information

1.6.7 Coverage extension
• Extending cellular network access
• Linking up with the Internet, intranets, etc.

1.7 ROUTING SECURITY IN MANETS

Security always implies the identification of potential attacks, threats and vulnerabilities of a certain system.

**Black hole**: In this attack, a malicious node uses the routing protocol to advertise itself as having the shortest path to the node whose packets it wants to intercept.

**Denial of service**: The DoS attack results when the network bandwidth is hijacked by a malicious node. It has many forms: the classic way is to flood any centralized resource so that the network instance, a route request is generated whenever a node has to send data to a particular destination. A malicious node might generate frequent unnecessary route requests to make the network resources unavailable to other nodes.

**Routing table overflow**: The attacker attempts to create routes to nonexistent nodes. The goal is to have enough routes so that creation of new routes is prevented or the implementation of routing protocol is overwhelmed.

**Impersonation**: A malicious node may impersonate another node while sending the control packets to create an anomaly update in the routing table.
**Energy consummation**: Energy is a critical parameter in the MANET. Battery-powered devices try to conserve energy by transmitting only when absolutely necessary. An attacker can attempt to consume batteries by requesting routes or forwarding unnecessary packets to a node.

**Information disclosure**: The malicious node may leak confidential information to unauthorized users in the network, such as routing or location information. In the end, the attacker knows which nodes are situated on the target route.

### 1.8 CROSS-LAYER DESIGN IN WIRELESS NETWORKS

#### 1.8.1 DEFINITIONS OF CROSS-LAYER DESIGN

Cross Layer design is said to be the violation of the layered architecture in order to get some improvements in the network parameters. The defined the cross-layer design [Sandeep Sharma, et.al, 2012] as follows:

**Definition**: Protocol design by the violation of layered communication architecture is cross-layer design with respect to the original architecture.

**Comment 1**: Violation of a layered architecture involves giving up the luxury of designing protocols at the different layers independently. Protocols so designed impose some conditions on the processing at the other layer(s).

For example, consider a model in the Figure 1.11 which consists of three layers viz. layer-1, layer-2 and layer-3 and follows the traditional layered architecture. Layer-1 is the lowest layer which provides its services to the layer-2 and layer-2 provide service to its layer just above it i.e. layer-3 via well defined interfaces which exists between layers. Define a interface which can communicate directly between the layer-1 and layer-2 bypassing the layer-2 then it is the violation of the layered protocol and hence it is a cross-layer design. While doing this the designer must take care of the headers which are combined at the layer-2 (as layer-2 is responsible of various operations and convert the layer-1 frame as required by the layer-3 by adding its own header).

![Cross-layer design between Layer1-3](image)

**Figure 1.11: Cross-layer design between Layer1-3**
1.8.2 THE MOTIVATING FACTORS FOR CROSS-LAYER DESIGN

In the following discussion will examine the motivating factors for ad-hoc networks and sensor networks. The motivating factors for cross-layer design for ad hoc networks include:

A. Cross-Layer Aspects: Nodes in wireless ad hoc networks [Marco Conti, et.al, 2004.] have to manage several performance aspects like system management, power management, and security management that cut across traditional layers. For example, both medium access and routing decisions have significant impact on power consumption, and the joint consideration of both can yield more efficient power consumption thereby increasing the battery life. The strict boundary separation of layers in the layered architecture and standard interlayer interfaces in traditional approaches do not permit adequate communication among layers to make joint decisions to optimize these cross-layer aspects. This has led to the proposal of new interaction models to support cross-layering, ranging from a more relaxed information flow and sharing between layers to full-fledged merging of layer functionalities.

B. Distributed State: In the traditional infrastructure models the base stations has a global view of the network state, where as in contrast with the traditional view, the network state in ad hoc networks is generally distributed across the nodes. Each node forms its own local view of state, representing a partial view of the overall network state. In most of the cases, it is not feasible to collect network state at any one of the node, which prevents the use of any centralized optimization algorithms. As such, each node can run distributed algorithms locally using its partial view of network state. Distributed algorithms can exploit a cross-layer design to enable each node to perform fine-grained optimizations locally whenever it detects changes in network state.

C. Mobility: Mobility introduces an additional challenge for ad hoc network design. Routing protocols would have to cope with this mobility of the mobile terminals by constantly adapting routing state to the changing user positions. Let us now consider mobility in the context of ad hoc networks, where no node has global view of network state. Mobility management poses an added challenge to the battery-powered nodes in ad hoc networks, which have to adjust their behaviour to the changing node locations. Mobility causes changes for the physical layer (for e.g. interference levels), the data link layer (for e.g. link schedules), the routing layer (for e.g. new neighbouring nodes), and the
transport layer (for e.g. connection timeouts). As such, a cross-layer based design enhances the capability of the node to manage its resources in the mobile environments.

D. Wireless Link Properties: Wireless links are more susceptible as compared to the wired links to interference variations and channel errors. For instance, in the example of the TCP congestion control problem [Floyd, 1994], over wireless links, in which TCP misinterprets a packet loss due to channel error as a sign of congestion. Wireless links are also more vulnerable to security attacks because of easy access to the wireless channel as the wireless channel is open. If the wireless link status information is provided at the higher layers the nodes can adapt their configuration in a better way at the physical layer. For example, a routing protocol detects degradation in the signal strength of a particular wireless link then it can divert the traffic to another wireless link which has an adequate quality on the link.

E. New Communication Modalities: Ad hoc network design can exploit the broadcast nature of the channel to enhance performance. For example, nodes can sneak on the neighbouring transmissions in order to estimate and evaluate the quality of links with neighbours. Antenna arrays can also enable the reception of multiple packets simultaneously on the wireless channel and the data packets corresponding to several connections could also arrive simultaneously at a node. The cooperation of various layers such as routing, data link, and physical layer can ensure the forwarding of data for all the connections within time.

D. Inherent Layer Dependencies: in a layered protocol stack there exist a number of interlayer dependencies which motivate cross-layer design for ad hoc and sensor networks. The data link and routing layers in ad hoc networks exhibit both variable interaction as well as algorithmic interaction, telling the need for design through coupling of these layers. The data link layer is also closely related with the physical layer. The physical layer deals with the channel state and the data link layer with the error control and flow control. If the change in the channel state at the physical layer is provided to the data link layer then it can adapt error control mechanisms in a adaptive manner, thereby improving the throughput.

E. Security: due to the fact that a wireless channel is open and could be access easily by an attacker security has become more and more important to secure our communication.
Security is an important concern in wireless networks due to their increased vulnerability and exposure to varying types of attacks. Unreliable wireless links, frequently changing network topology and lack of a centralized system to handle the security needs of the network contribute to insecure standalone systems in wireless networks. Intrusion detection systems located on concentrated points such as network gateways and wireless access points are not guaranteed to achieve the desired security level in the network. There exists a need of an efficient and reliable intrusion detection system to manage the access control and provide a monitoring unit to detect any anomalous behaviour in the network [Trappe W, et.al, 2009]. In a wireless network protocol stack, every layer is vulnerable to attacks (internal and external) by adverse nodes in the network. Independent security solutions at different layers might lead to conflicting actions and result in performance degradation. Hence, ensuring security and network reliability, has to be jointly addressed in all of the protocol layers. Proper interaction and coordination among different protocol layers helps in developing a robust intrusion detection system suitable for wireless networks. Such interactions are the key elements to building cross layer architectures. Apart from the need to make a collaborative decision, adopting a cross-layer approach to intrusion detection facilitates effective fault diagnosis and reduced false alarms. Physical layer authentication for the detection of the intruder when integrated with the cross-layer design can improve the security of the wireless networks.

F. Resource-Constrained Nodes: The mobile nodes for ad hoc networks are decreasing in size, which results in the use of smaller batteries for these nodes. Cross-layer design approaches can expose power related variables at several layers, enabling nodes to efficiently utilize their energy resources and to maximize the battery life of the node.

1.9 PROBLEM STATEMENT

Packet loss and bandwidth dilapidation are caused due to congestion, and so, time and energy is wasted during its recovery. To avoid this, there is a need of congestion control mechanism. In the layered protocol stack each layer communicates only with the adjacent layers using well defined interfaces and hence there is no performance optimization, so there is demand for Cross-layer design which emphasizes on the network performance optimization by enabling different layers of the communication stack to share state information or to coordinate their actions in order to jointly optimize network performance.
The vulnerability of attacks may introduce damage to the whole network connectivity; therefore physical layer authentication for the detection of the intruder is integrated with the cross-layer design which can improve the security of the Ad hoc Networks. In addition to this efficient energy consumption algorithm is required to enabling nodes to efficiently utilize their energy resources, to maximize the battery life of the node and high throughput.

1.10 RESEARCH CONTRIBUTION

The research work mainly concentrates on developing the cross layer based secure congestion control scheme which attains more energy efficiency, efficient authentication, data integrity and improved network lifetime which depends on cross layer congestion control design, multipath routing, energy consumption model and security schemes. This objective intrinsically comprises several tasks that should be done to achieve the final goal.

There are four main challenges of the Mobile Ad hoc Networks. They are:

- Congestion Control
- Efficient Authentication and Data Integrity
- More energy efficiency
- Improved Network Lifetime and Connectivity

These challenges are very important to analyze the performance of the ad hoc network. Packet failure in MANETs is primarily caused due to obstruction. To avoid this, there is a need for efficient congestion avoidance protocol which attempt to improve packet delivery ratio and high network throughput.

Mobile nodes which are organized in random manner and communicate with each other without any centralized infrastructure where information sharing is very vital. So there is demand for cross-layer design which emphasizes on the network performance optimization by enabling different layers of the communication stack to share state information or to coordinate their actions in order to jointly optimize network performance.

The vulnerability of attacks may introduce damage to the whole network connectivity and lack to attain correct balance between data integrity and congestion status. To overcome this, efficient authentication scheme is needed to make the network
more secure and have high data integrity. In addition to this, an Efficient Energy consumption model is required to improve energy efficiency of the mobile nodes and achieve high throughput. If these schemes are focused on the MANET, the above said challenges can be easily achieved to introduce the popularity of the network.

The aim of the work is to achieve the challenges of the Ad hoc Networks. The existing works in Ad hoc Networks have some limitations on their reliability in terms of the congestion control, security and energy consumption.

1.11 THESIS LAYOUT

The thesis is organized into nine chapters and one Appendix. Some of these chapters are based on developments from former chapters, but in general, all of them can be read independently as self-contained entities.

- Chapter 1 presents a deep revision of challenges, threats and methodologies used in MANET.
- The literature review is carried out in Chapter 2.
- Chapter 3 describes Effective Congestion Avoidance Scheme (ECAS) which consists of congestion monitoring, effective routing establishment and congestion less routing. The overall congestion status is measured in congestion monitoring. In routing establishment, it is proposed the contention metric in the particular channel in terms of queue length of packet, overall congestion standard, packet loss rate and packet dropping ratio to monitor the congestion status. Based on the congestion standard, the congestion free routing is established to reduce the packet loss, high overhead, long delay in the network when compared with the existing work.
- Chapter 4 discusses Cross Layer Based Congestion Control Scheme (CLCCS) for reducing the packet losses in the network. The proposed scheme contains three parts. In the first part, the cross layer design is proposed to ensure that the information sharing can be done between the different layers in protocol stack. In the second part, the congestion detection scheme is explored which attains packet loss rate and congestion scale factor. In the third part, congestion control is achieved using cross layer approach. Hence the congestion route is determined
based on the path gain and buffer tenancy fraction. Each node maintains the congestion scale value and buffer tenancy fractional value.

- Chapter 5 describes a Cross layer based Secure Multipath Routing (CLSMRSCA) developed for congestion avoidance. The cross layer model is introduced to give more security among the nodes. The multipath routing is focused to provide the load balancing and network lifetime. Each node chooses multipath to avoid the network congestion based on optimized encryption and decryption algorithm. Here the asymmetric key is chosen. Both encryption and decryption phases are used to validate the cipher text and manipulate the plaintext according the given message. Integrity and authentication of data is improved in this scheme.

- Chapter 6 explains an Efficient Energy based Congestion Control Scheme (EECCS) proposed for congestion avoidance and to improve energy efficiency of the mobile nodes. Cross layer design is deployed to improve the network performance and multipath routing is used to avoid congestion and to increase network lifetime. Probability of the retransmission of packets is reduced with the help of calculating the energy level of data and acknowledgment packets.

- The concluding remarks with the findings of the research are presented in Chapter 7 along with future research directions.

- The works of several researchers are quoted and used as evidence to maintain the concepts explained in this dissertation. All such evidences used are listed in the Reference section of the dissertation.

1.12 SUMMARY

This chapter summarizes the major definitions, issues, importance and tasks used in Mobile Ad-hoc Networks. In mobile ad hoc networks (MANETs) congestion occurs with limited resources. Large amount of real-time traffic involves high bandwidth and it is liable to congestion.