2. Ad-hoc Networks

Wireless communication and mobility in ad-hoc networks, unlike traditional wired networks require different types of routing protocols that should be based on new and different principles. Routing protocols for traditional wired networks are designed under the assumption that the nodes’ relative positions generally remain unchanged and the links between them exist at all times. In a mobile ad-hoc network, however, the topology changes so frequently as the mobile nodes move. Furthermore, links existence between neighbour nodes is not always assured due to the nature of the wireless medium, such as the presence of obstacles and unidirectional links.

In this chapter new routing protocols in ad-hoc networks are proposed. These routing protocols belong to several classifications based on when and how the routes are discovered. In the following sections, the properties and limitation of existing conventional routing protocols in wired networks that make them inappropriate in mobile ad hoc networks are discussed. Then different ad-hoc routing protocols are being proposed discussing their advantages and limitations. Finally the impacts of mobility models on ad-hoc routing are illustrated.

2.1. The need for ad-hoc routing protocols

Routing is not a new issue in computer networking. Link-state routing e.g. Open shortest Path First (OSPF) (Moy 1998) and distance vector routing e.g. Routing Information Protocol (RIP) (Hedrick, 1998 and Malikin, 1998) have existed for a long time and are widely used in conventional networks. However, they are not very suitable for use in mobile ad-hoc networks, for a number of reasons:
• They were designed with a quasi-static topology in mind. In an ad-hoc network, with a frequently changing network topology, these routing protocols may fail to converge, i.e., to reach a steady state.

• They were designed with a wired network in mind. In such a network, links are assumed to be bidirectional. In mobile ad-hoc networks, this is not always the case; differences in wireless networking hardware of nodes or radio signal fluctuations may cause uni-directional links, which can only be traversed in one direction.

• They try to maintain routes to all reachable destinations. In mobile ad-hoc networks with a very large number of nodes, such as large wireless sensor networks, this may become infeasible because of the large number of routing entries.

In mobile ad-hoc networks, periodic flooding of routing information is relatively expensive, since all nodes compete for access to the wireless medium which usually offers a limited amount of bandwidth. Therefore, special routing protocols are needed for ad-hoc networks.

2.2. Challenges in Routing

The nodes in an ad-hoc network function as routers that discover and maintain routes to other nodes in the network. The primary goal of such an ad-hoc network routing protocol is to establish correct and efficient route between a pair of nodes so
that messages may be delivered in a timely manner. Route construction should be
done with minimum overhead and bandwidth consumption. One of the major
challenges in designing a routing protocol for the ad-hoc networks stems from the fact
that, on one hand, determining a packet route requires a node needs to know at least
the reachability information to its neighbours. On the other hand, in an ad-hoc
network, the network topology can change quite often. Host mobility results in
frequent link breaks and all routes that use the link need to be repaired. Furthermore,
as the number of network nodes can be large, the potential number of destinations is
also large, requiring large and frequent exchange of data (e.g. route requests, route
tables, route errors) among the network nodes. Thus, the amount of update traffic can
be quite high. This is in contradiction with the fact that all updates in a wireless
interconnected ad-hoc network travel over the air and thus, are expensive in terms of
resources. Since the nodes are battery powered, the routing should also be energy
efficient. Due to all these constraints, no single protocol works well in different
environments.

2.3. Wireless Medium

There have been industrial, scientific, medical (ISM) wireless local area
networks operating at 900 MHz, 2, 4 and 5 GHz since 1990s, until IEEE approved an
international interoperability standard IEEE 802.11 (LANMAN Standards
Committee, 1998). The standard specifies Medium Access Control (MAC) protocol
procedures and three types of Physical Layers (PHY). There are two radio-based
PHYs and the third PHY uses infrared light. All PHYs support 1 Mb/s or 2 Mb/s data
rates. The frequency used is 2.4GHz. Later in July 1998, an 11 Mb/s PHY running
on standard MAC was developed. Smart or adaptive antennas for mobile communications have been of great interest, due to their provision of opportunity for increased bandwidth and range. Major drawbacks of such systems are transceiver complexity and more complex radio resource management. Lehne and Pettersen (1999) have done a detailed survey on such antenna technologies.

2.3.1. Infrared

Infrared PHY is simple and inexpensive. They use the same signal frequencies on fibre optic links. Infrared systems detect only the amplitude of the signal so the interference is reduced greatly. Bandwidth is not limited or restricted, so transmission speeds can be larger than other systems. Since it operates in the light spectrum, it does not require permission from Federal Communications Commission (FCC). This makes it very attractive. The range can be increased to few kilometres in outdoors by aiming. In indoor usage, Omni directional transmitters can be used and a multi-hop network can be built. The only requirement is the true line of sight. It is cheap, however its spectrum is shared with the sun and fluorescent lights and they can cause heavy interference.

2.3.2. Microwave

Microwave operates at less than 500 mW (FCC 2000). It uses narrow-band transmission with single frequency modulation and is usually set up at 5.8 GHz band. Microwave achieves higher throughput, because it does not involve the overhead of spread spectrum systems. Radio LAN is an example to microwave technology.
2.3.3. Radio

2.3.3.1. Direct Sequence Spread Spectrum (DSSS):

Direct sequence spread spectrum spreads the transmission signal over an allowed band (for example 25 MHz). A random binary string which is the spreading code is used to modulate the transmitted signal. The data bits are mapped to a pattern of chips and mapped back into a bit at the destination. The number of chips that represent a bit is equal to the spreading ratio. Higher the spreading ratio, the more the signal is resistant to the interference. IEEE 802.11 standard requires spreading ratio of 11 (Nee, et.al 1999) the transmitter and receiver must be synchronized with the same spreading code.

If orthogonal spreading codes are used, then more than one LAN can share the same band. There is a study on distributed assignment of codes for multi-hop packet-radio networks, which has an implementation of code assignment algorithm as a part of the Medium Access Controller (MAC) (Garcia and Raju, 1997). However, since DSSS systems use wide sub channels, the number of adjacent LANs is limited by the size of sub channels. Recovery from interference is fast in DSSS, because of the ability to spread the signal over a wider band. Wave-LAN is an example of DSSS product by Lucent Technologies.
2.3.3.2. Frequency Hopping Spread Spectrum (FHSS):

When a band is split into small sub channels like 1 MHz, the signal may hop from one sub channel to another transmitting short burst of data on each for a set period of time, which is called dwell time. The hopping sequence must be synchronized at the sender and receiver nodes or the data will be lost. Frequency hopping is less vulnerable to the interference because frequency is always shifting. It is also difficult to intercept a frequency hopping communication. This provides higher security point. The whole band needs to be jammed to interfere with the transmission. Since sub channels are smaller than DSSS, more LANs can run simultaneously in the same band. An example to this technology is Wave Access and Jaguar.

2.3.3.3. Multipath Interference:

When a signal bounces off walls or other barriers, interferences are generated and they reach the destination at different times. This is called multipath interference. This kind of interference affects IR, RF and MW systems. FHSS intuitively solves this problem by hopping to other frequencies whereas other systems apply algorithms to avoid this. A subset of multipath is Rayleigh fading (Goodman 1997). This happens when signals are arriving from different directions and the difference in path length is a multiple of half the wavelength. Rayleigh fading has the effect of completely cancelling out the signal. IR systems are not affected by Rayleigh fading because the wavelengths used in IR are so small.
2.3.4. Medium Access

2.3.4.1. Medium Access Control Protocol

Most wired LANs use Carrier Sense Medium Access with Collision Detection (CSMA/CD) (Tanenbaum 1996) as the MAC protocol. In a carrier sense system, the station will listen for a carrier before it starts transmitting. If another station is already transmitting the sender waits and tries again later. When two stations send at the same time, transmissions collide and information will be lost. Collision detection handles this situation by listening to the signal it is transmitting to ensure that everything is going right. Whenever a collision occurs, nodes stop transmitting and will try again after a delay. The delay is determined by the back off algorithm. This technique cannot be used in wireless medium. Figure 2.1 illustrates the hidden terminal problem. Node C cannot hear node A, so if A is transmitting, node C will not know that A is in transmission and it will decide to transmit as well. This will result in collision at B.

Figure 2.1 Hidden terminal problems that occurs when node A unwittingly interferes with the transmission of node C
Carrier Sense Medium Access with Collision Avoidance (CSMA/CA or MACA) (Gerla, et.al 1999) solves this problem. In this technique the sender listens before it sends. If a node is already transmitting the sender waits. If collision occurs, both senders execute back off algorithm like in CSMA/CD. If no node is transmitting, the sender sends a Request to Send (RTS) message, which indicates that the node is ready to transmit data. RTS contains the destination address and the duration of the transmission. Other stations will now wait till the end of transmission. Destination node sends back Clear to Send (CTS) message. Each packet is acknowledged. When an acknowledgement is not received, MAC layer retransmits the data. This is a 4 way handshake, accepted as a standard and used (Gerla et al 1999) by 802.11 as seen in Figure 2.2

![Figure 2.2 The 4-way handshake in 802.11](image)
There is a QoS modified CSMA proposal for supporting real-time traffic. It is a distributed MAC scheme. Nodes requiring real-time transmissions make back-bursts (BB) which are special pulses of energy to the channel. By this method real-time data is transmitted before normal data packets (Sobrinho, and Krishnakumar, 1999). In addition, there is a study on MACs to benefit from the use of directional and adaptive antennas (Ko, Shankarnumar, and Vaidya, 2000). The performance of another MAC scheme, Floor Acquisition Multiple Access with Non persistent carrier (FAMA-NCS) that guarantees the sender to send data packets free of collisions to a given receiver at any time, is analyzed by (Garcia, and Fullmer, 1998). Talucci and Gerla, 1997, have introduced new wireless MAC protocol Multiple Access Collision Avoidance by Invitation (MACA-BI). MACA-BI removes RTS part of the handshake, and gains considerable performance for high-speed, steady traffic environments, which is typically for indoor usage only. Another MAC protocol called Hop-Reservation Multiple Access (HRMA) for wireless ad-hoc networks is introduced by Tang and Aceves, which is based on simple half-duplex and very slow FHSS radios. HRMA takes advantage of the time synchronization necessary for frequency hopping, thus providing a reservation and handshake mechanism that guarantees collision-free data transmission in the presence of hidden terminals (Tang, Garcia, 1999). In addition, Tang et al 1999 did a comprehensive survey on ad-hoc MAC mechanisms.

With growing demand on mobile computing, IEEE has formed a working group to develop a medium access control and physical layer standard for wireless connectivity.
2.3.4.2. 802.11 DSSS Architecture

This section gives basic information about 802.11 DSSS 1 Mb/s and 2 Mb/s signal structure. The packet signature is shown in Figure 2.3.

![Diagram of IEEE 802.11 DSSS packet structure](After Nee et al 1999)

Presentation Protocol Data Unit (PPDU) consists of three segments: The preamble, which is used for signal detection and synchronization. Second is the header, including data rate and packet length information. Third part Message Protocol Data Unit (MPDU) contains data. The preamble and header is transmitted at 1 Mb/s, whereas data is transmitted at one of four possible rates viz., 1 Mb/s, 2 Mb/s, 5.5 Mb/s or 11 Mb/s. 1 and 2 Mb/s DSSS signal is created using fixed spreading sequence formed by using the 11 chips Barker code (Nee et al 1999). For 1 Mb/s, the fixed spreading sequence is used to spread a 1 Mb/s binary phase shift keying (BPSK)
signal and for 2 Mb/s the same spreading sequence is quaternary PSK (QPSK) modulated. For all data rates of 1-11 Mb/s, the chip rate is 11 Mchips/s.

2.4. Routing protocols

As shown in Figure 2.4 routing protocols may be generally categorized as

1. Table driven
2. Source initiated (demand driven)
Figure 2.4 Categorization of ad hoc routing protocols
2.4.1. Table Driven Protocols:

Table driven protocols maintain consistent and up to date routing information about each node in the network. These protocols require each node to store their routing information and when there is a change in network topology the entire network has to be updated. Some of the existing table driven protocols are given below.

- Destination sequenced Distance vector routing
- Wireless routing protocol
- Global state routing protocol
- Fish Eye state routing protocol
- Hierarchical state routing protocol
- Zone based hierarchical link state routing protocol
- Cluster Gateway switch routing protocol

2.4.1.1. Destination sequenced distance vector routing (DSDV)

DSDV is a table driven algorithm based on Bellman-Ford routing mechanism (Ford Jr, and Fulkerson, 1962). In DSDV, packets are routed between nodes within a network, using routing tables stored at each node. Every node in the network has a routing table, which specifies all possible destinations in the network and the number of hops to each destination is recorded. Each entry is marked with the sequence numbers, which helps to identify the stale routes from new ones. To maintain consistency updated table are transmitted periodically.
Figure 2.5 illustrates the routing procedure of DSDV. Consider a packet which needs to be sent from node 1 to node 3. From node 1, the next hop for the packet is node 4. When node 4 receives the packet, it looks up the routing table with the destination address (node 3). As per the routing table entry in node 4, it transmits the packet to the next hop as specified by node 5. This procedure repeats until the packet finally reaches the destination.

To reduce the network traffic overhead, updation can employ two possible types of packets. The first is known as full dump. This type of packet carries all routing information and can require multiple network protocol data units. During the periods of occasional movement, these packets are transmitted infrequently. Incremental packets are used to relay the information, which has changed since the last full dump. These broadcasts should fit into a standard size Network Protocol Data Unit (NPDU), thereby decreasing the amount of traffic generated. The mobile nodes maintain an additional table where they store the data sent in the incremental routing information packets (Perkins and Bhaqat 1994).

New routes, which are broadcasted, contain the address of the destination, the number of hops to reach the destination, the sequence number of the information received during the destination, as well as new sequence number unique to broadcasts (Royer and Toh 1999. The route labelled with the most recent sequence number is used. If two updates have the same sequence number, the one with the smaller metric is used to shorten the path.
Mobile nodes cause broken links as they move from place to place. A broken link can be detected by hardware or if no information has been received from its neighbour. A broken link is described as having a metric with infinity. The detecting node will broadcast an update message for the lost destination. The update message will have a new sequence number and a metric of infinity which will cause the routing table entries for the lost node to be flushed from the network. Routes to the lost node will be established when the lost node sends its next broadcast.
Figure 2.5 – Destination sequenced distance vector routing
To avoid nodes and their neighbours generating conflict sequence numbers when topology changes, nodes generate even sequence numbers for themselves and neighbours responding to link changes generate odd sequence numbers. DSDV has more procedures for handling network layer addresses, dealing with non-mobile stations, damping fluctuations caused by topology changes.

Advantages: By delaying the broadcast of updation and length of settling time, mobiles can reduce network traffic and optimize routes by eliminating those broadcasts that would occur if a better route was discovered in the near future.

Disadvantages: DSDV requires broadcasting periodic updates, regardless of network traffic. As the number of nodes in the network grows, the size of the routing table and the bandwidth required to update also grows. This remains as overhead, which has been found in simulation with the network consisting of 50 nodes (Broch et al 1998). DSDV does not perform well when there is a high rate of node mobility.

2.4.1.2. Wireless routing protocol (WRP)

Wireless routing protocol is a table-based protocol where it maintains the routing information at all nodes in the network (Murthy, and Garcia, 1996). Each node in the network maintains four tables.

- Distance table
- Routing table
- Link cost table
• Message retransmission list table (MRL)

The distance table of a node x contains the distance of each destination y via each neighbour z of x. It also contains the distance information of downstream neighbour of z through which the path is realized.

The routing table of node x contains the distance of each destination node y from node x, the predecessor and successor of node x on this path. It also contains a tag entity to determine whether the entry is a simple path, a loop or invalid. Storing predecessor and successor in the table helps to detect loops and avoid count to infinity problems.

The link cost table contains cost of link to each neighbour of the node and the number of time-outs since an error free message was received from that neighbour. Each entry of MRL contains the sequence number of update message, a retransmission counter, an acknowledgement required flag vector with one entry per neighbour, lists of updates sent in the update message. MRL keeps a record of which updates in an update message has to be retransmitted. The neighbours should acknowledge the retransmission.

Mobiles inform each other about the link changes with an update message. Update message are sent between the neighbouring nodes, it contains a list of updates, list of responses specifying which of the mobiles should acknowledge the update. Update messages are sent when there is a change in link and when there is a loss of link. The neighbours then update their distance table entries and check for new
possible paths through other nodes. If any new paths are relayed back to the original nodes then they can update their tables accordingly.

The existence of neighbours is known through the receipt of acknowledgement and other messages. To ensure connectivity a node must send a hello message, within a particular period of time. Lack of message from the node indicates a failure of link, which causes a false alarm. When a hello message is received from a new node, the new node is added to the mobile routing table and the mobile sends the new node a copy of its routing information.

Advantages: Wireless routing protocol belongs to path finding algorithm. It eliminates looping situation and provides faster route convergence when a link failure event occurs.

It enhances distance vector routing in four ways:

- When there is no link change, wireless routing protocol exchanges hello messages rather than exchanging the whole route table.
- If there are topology changes, only the path vector tuples that reflect the updates are sent. The path vector tuple contains destination, distance, and the predecessor node id.
- To improve reliability while delivering update message, every neighbour must send an acknowledgement for the update packets received. Retransmissions are made if positive acknowledgements are not received within the particular period.
• The predecessor node id information allows the protocol to recursively calculate the entire path from source to destination. With this information, wireless protocol reduces looping situation, reduces the time for convergence and is less prone to count to infinity problem.

2.4.1.3. Global state routing (GSR)

The nodes exchange vectors of link states among their neighbours during routing information exchange (Tsu-Wei Chen and Gerla, 1998). Based on these link states, nodes maintain a global knowledge of network topology and optimize their routing decision locally. Global state routing avoids flooding of message.

An ad-hoc network is an undirected graph \( G = (V, E) \) where \( V \) is a set of nodes and \( E \) is the set of undirected links connecting nodes \( V \). Each node has a unique identifier which represents a host with a wireless communication device with transmission range \( R \) and an infinite storage space. Nodes can move around and change their speed independently. An undirected link can be formed between nodes when the distance between two nodes is less than or equal to \( R \). A link is removed when the nodes move out of transmission range \( R \).

Each node maintains a neighbour list, a topology table, a next hop table and a distance table. Neighbour list specifies the list of all neighbours, the topology table contains the link state information as reported by the destination and the time stamp of the information. The next hop table specifies the next hop to which the packets should be forwarded. The distance table specifies the shortest distance to each destination node.
The routing messages are forwarded when there is a link change as in link state protocol. On receiving a routing message, the node updates its topology table when the sequence number of the message is newer than the sequence number stored in the table. After this process the node reconstructs its routing table and broadcasts the information to its neighbours.

Complexity comparison: Complexity of GSR scheme is studied and compared to DBF, LS. The complexity is studied under five aspects.

*Computation complexity:* The number of computation steps for a node to perform a routing computation after an update message is received.

*Memory complexity:* The memory space required to store the routing information.

*Data complexity:* The size of the control packet exchanged by a node in each time slot.

*Packet complexity:* The average number of routing packets exchanged by a node in each time slot.

*Convergence time:* The time required to detect a link change.

GSR and LS have the same memory and computation complexity, since both maintain the topology of the whole network and use Dijkstra algorithm to compute the shortest path. LS transmit one short packet for each link update. On the other hand, GSR and DBF transmit a fixed number of update tables using longer packets to optimize the throughput. Convergence time of GSR is superior to that of DBF. If shorter update interval is used, GSR can converge as fast as LS.
Advantages: Global state routing protocol avoids flooding of message.

Disadvantages: The drawback of GSR is that it uses large size update message. The latency of LS change propagation depends on the update period.

2.4.1.4. Fish eye state routing (FSR)

Fisheye is based on hierarchical routing protocol. It uses the fish eye technique, proposed by Kleinrock and Steven 1971, to reduce the size of information required to represent graphical data. The eye of the fish captures in high detail the pixels near the focal point. The detail increases as the distance of the focal point increases. The fish eye approach maintains accurate distance and path quality information about the immediate neighbourhood of node.

Fisheye state routing introduces the multi level fish eye scope to reduce routing update overhead in large network. The nodes exchange link state entries with a frequency based on destination and distance. From the link state entries the nodes, construct the topology map of the entire network and compute optimal routes. The scope of the fish eye is based on the central node. The scope is defined in terms of nodes that can be reached in certain number of hops. The central node has most accurate information about all the nodes in white circle. Even though a node does not have accurate information about distant nodes, the packets are routed correctly because the route information becomes more accurate as the packets move closer to the destination.
Advantages: Fisheye state routing scales well for larger networks where mobility is high and bandwidth is low. By choosing proper scope levels and radius size, FSR provides a flexible solution to maintain accurate routes in ad-hoc networks (Guangyu et al. 1997).

2.4.1.5. Hierarchical state routing (HSR)

HSR (Iwata et al. 1999) is based on multilevel clustering and partitioning of mobile nodes. The network is partitioned into clusters and a cluster head is elected based on cluster-based algorithm. These cluster heads organize themselves as clusters. Nodes in the physical cluster broadcast their link information to each other. The cluster head summaries these information and sends it to the neighbouring cluster heads through gateway.

These cluster heads are members of cluster on a higher level and they exchange their link information as well as low level information among each other. A node at higher level floods the information to its lower level. Each node has a hierarchical address. A gateway can be reached from root via more than one path. A gateway can have more than one hierarchical address. Hierarchical address ensures delivery from anywhere in the network to the host.

The nodes are partitioned into logical subnet and each node is assigned logical address subnet host. Each of this subnet has a location management server (LMS). The nodes within the subnet must register their logical address with the location
management server. The LMS advertise their hierarchical address to top levels and the information is sent to all LMS too. The transport layer sends a packet to network layer specifying the logical address of destination. The network layer will then find the hierarchical address and then forwards the packet. The destination LMS forwards the packet to destination. Once the source and destination know the hierarchical address of each other they can bypass LMS and communicate directly. Both logical and hierarchical addresses are used for routing and are adaptable to network changes.

*Advantages:* Hierarchical state routing ensures delivery of packets anywhere in the network to host.

2.4.1.6. Zone based hierarchical link state routing protocol (ZHLS)

Zone routing protocol is a neighbour selection based protocol, designed to combine the advantages of proactive and reactive routing strategies. In this protocol, the network is divided into non-overlapping zones. Each node must be aware of the node connectivity of its own zone and the zone connectivity of the whole network (Joa-Ng, and Lu 1999). The routing is performed in two levels

- local node levels
- global zone

The hierarchical characteristics of the protocol reduce communication overhead and storage requirement of routing information in large network. There are no cluster heads in this protocol. The zone level topology is more stable and reduces
the amount of routing messages that are globally propagated. ZHLS provides efficient bandwidth and can accommodate itself to changing network topology.

In ZHLS, it is necessary to keep track of physical location to determine the affiliated zone. For this purpose, a Global Positioning System can be used.

Node level specifies how nodes within a zone are connected. A virtual link exists between two zones if at least one node of a zone is physically connected to a node in another zone. There are two types of packets viz., Node link state packets and Zone link state packets. Node link state packets contain the neighbouring node information and are propagated within the zone.

Zone link state packets contain zone information and are propagated globally. Therefore, each node has full node connectivity knowledge in its zone and only zone connectivity information about other zones in the network. Given Zone id and node id, the packet is routed based on zone id till it reaches the correct zone and then the packet is routed within the zone using the node id. A <zone id, node id> of the destination is sufficient for routing.

Advantages: ZHLS is adaptable to changing topologies. The Zone level topology is more stable and reduces the amount of routing messages that are globally propagated. ZHLS provides efficient use of the bandwidth and can accommodate itself to changing network topology.
2.4.1.7. Cluster gateway switch routing protocol (CGSR)

Ad-hoc network consists of identical set of nodes that can move freely, independently and communicate with other nodes via wireless link. Such nodes can be represented as a set of cluster by grouping them together which are in close proximity with one another. Cluster heads form a backbone and can be used to route packets to the nodes within the cluster. Clusters are formed by diffusing node identities along their links. Different heuristics employ different policies to elect their cluster head (Chiang, 1997). CGSR is a clustered multihop wireless network with several routing schemes. Cluster head controls the group of ad-hoc nodes. A framework for code separation, channel access, routing and bandwidth allocation can be achieved. The disadvantage of CGSR is frequent changes of cluster head, which may affect the performance of routing protocol. Instead of reselecting a cluster head, a least cluster change (LCC) clustering algorithm is introduced. Using LCC cluster head changes are made only when two cluster head comes into contact or when a node moves out of contact of all cluster heads.

CGSR uses DSDV as an underlying routing scheme and has the same overhead as that of DSDV. CGSR uses a hierarchical routing approach to route traffic from source to destination.

Gateway nodes are nodes which are within the communication range of two or more cluster heads. A packet sent by a node is first routed to its cluster head and then from the cluster head to gateway to another cluster head and so on until it reaches the cluster head of the destination. The packet is then transmitted to the destination. Each
node keeps a cluster member table where it stores the destination cluster head for each mobile node in the network. Each node in the network broadcasts and updates the cluster member tables. In addition, each node maintains a routing table, which specifies how to find the next hop to reach the destination. When a node receives a packet, the node consults its routing table and cluster member table to determine the nearest cluster head to reach the destination. By this method the packet is transmitted to reach the destination.

2. 4.2 Source Initiated Demand driven

In on-demand routing protocols routes are created as and when required. When a source wants to send to a destination, it invokes the route discovery mechanisms to find the path to the destinations. The route remains valid till the destination remains reachable or until the route is no longer needed.

The different types of On Demand driven protocols are:

- Cluster based routing protocols
- Ad-hoc on demand distance vector routing
- Dynamic Source routing protocol
- Temporally ordered routing algorithm
- Associativity Based routing
- Signal Stability routing protocol
2.4.2.1. Cluster based routing protocol (CBRP)

In Cluster based routing protocol (Jiang et al 1997) the nodes are divided into clusters. When a node comes up, it enters into an undecided state, a timer is set and a hello message is broadcasted. When a cluster head receives a hello message it responds with a hello message immediately. When an undecided node gets this message, it sets its state to member. If the timer goes off for undecided node, then it elects itself as cluster head if it has a bidirectional link to neighbours else it remains in this state and repeats the procedure again. Cluster heads should be changed less frequently as possible.

Each node maintains a neighbour table. For each neighbour, the neighbour table of a node contains the status of link (unidirectional or bidirectional) and the state of the neighbour (cluster or member). A cluster head keeps information about all nodes in its cluster and maintains a cluster adjacency table, which specifies the information about all its neighbouring clusters. For each neighbouring cluster, the entry in the table specifies the gateway through which the cluster can be reached and the cluster head of the cluster.

When a source has a data to send, it floods route request packets. On receiving the request, the cluster head checks whether the destination is in its cluster. If present, it sends the request directly to the destination, else it sends the request to all adjacent cluster heads. Since the address of the cluster head is recorded in the packet, the cluster head can discard the request packet if it has already received the packet. The cluster head address is recorded in the packet because the cluster head can discard
request packet if it has already seen. When a destination receives the packet, it replies the back through the route that has been recorded in the request packet. When a source does not receive any reply within the period, it backs off exponentially before trying to send the route request again.

In CBRP, routing is done based on source routing. It uses route shortening so that on receiving a source route packet the node tries to find the farthest node in the route that is its neighbour and sends the packet to that node thus reducing the route. While forwarding the packet if there is a broken link then it sends back an error message to the source and then uses a local repair mechanism. The CBRP has built in local repair mechanism by which if a node finds the next hop is unreachable, it checks if the next hop can be reached through its neighbours or if hop after next hop can be reached through any other neighbour. If any of the two works then the packet can be sent through the repaired path.

2.4.2.2. Ad-hoc on demand distance vector routing (AODV)

AODV (Perkins and Royer, 1999) initiates routing activities only in the presence of data packets in need of route. AODV is a uniform, destination based, reactive protocol. It is implemented by using table driven routing framework and destination sequence numbers into an on demand protocol. AODV is an improvement of DSDV and it minimizes the number of required broadcasts by creating routes on demand basis, as opposed to maintaining a complete list of routes in DSDV algorithm. AODV can be classified as pure on demand acquisition system. The nodes, which are not selected in the path, do not maintain any routing information or do not take part in routing table exchanges. AODV prepares loop-free routes for both unicast
and multicast even while repairing loop broken links. AODV is capable of unicast, broadcast, and multicast communication. Unicast and multicast routes are discovered on demand and use a broadcast route discovery mechanism. There are numerous advantages of combining both unicast and multicast communication ability into the same protocol. A protocol which specifies that route information can be obtained when searching a multicast route can also be used to increase unicast routing knowledge.

AODV primary objectives:

- To provide unicast, broadcast, and multicast capabilities to all nodes in ad-hoc network.
- To minimize the broadcast of control packets.
- To disseminate information about link breakages to neighbouring nodes that utilizes the link.

When a source node has to send some message to the destination and if there is no valid path, then it initiates a path discovery process to locate the other node. It broadcasts a route request packet to its neighbours which then forwards the request to its neighbours and so on until it reaches the destination or any intermediate node with a relatively new route to the destination node. AODV uses destination sequence numbers to ensure that all routes are loop free and contains the most recent information. Each node maintains its sequence number and broadcast id. A route request is initiated by incrementing the broadcast id and it includes the node’s IP address as well. Along with its own sequence number and broadcasts id, the source
node includes in the route request the most recent sequence number for the destination. Intermediate nodes do reply only if they have a route to the destination, whose corresponding destination numbers are equal to or greater than that contained in the route request.

While forwarding the route request the intermediate nodes record in the route table the address of the neighbour from which the copy of the broadcast is received, thereby establishing a reverse path. If additional copies of same route request are received, then these packets are discarded. Once the route request reaches the destination, the destination responds by sending a route reply packet back to the neighbour from where it received the route request. As the route reply is routed back along the reverse path the nodes in this path sets forward route entries in their route tables which point to the node from where the route reply had come. These forward entries indicate active forward route. Along with the route entry, a timer is also set which causes deletion of the entry if it is not used within the specified lifetime. Route reply is forwarded along the path established by the route request. AODV supports only symmetric links.

*Route maintenance:* If a source node moves, it is able to reinitiate a route discovery protocol to find a new route to the destination. If a node along the route moves, its upstream neighbour notices it and propagates a link failure notification message to each of the active stream neighbours to inform them about the erasure of the part of the route. These nodes propagate the link failure notification to their upstream neighbours until the source node is reached. The source then reinitiates a route discovery for that destination if a route is still desired.
Hello messages are broadcasted by nodes to inform about each others neighbourhood and they are used to maintain the local connectivity of the node. When a retransmission of packet has to be made the nodes listen to ensure that next hop is within the reach. For the above reasons, hello messages are used to ensure that the next hop is within the communication range.

*Advantages:* There are numerous advantages of combining both unicast and multicast communication ability in the same protocol. A protocol, which specifies that route information, can be obtained when searching a multicast route can also increase unicast routing knowledge. If a node returns a route for a multicast group to some source node, that source node in addition to learning how to reach a multicast group will also have learned a route to the node returning the information. AODV has the advantage of enhanced general routing knowledge. Combining both the protocols simplifies coding.

*Disadvantages:* AODV does not provide guaranteed delivery of data packets. To enhance this service, multicasting algorithm known as Scalable Reliable Multicast (SRM) (Floyd et al 1997) can be used.

2.4.2.3. Dynamic Source routing protocol (DSR)

DSR is a uniform, topology based reactive protocol. It emphasizes aggressive caching and deduction of topology information extracted from source routing headers. When using source routing, each packet to be routed carries in its header the complete
list of nodes through which the packet must pass. An advantage of source routing is that intermediate hops need not maintain routing information in order to route the packets they receive, since the packets carry the necessary routing information.

DSR adapts quickly to routing changes when host movement is frequent, it requires less or no overheads during periods when the host moves less frequently. This protocol performs well in environmental conditions such as host density and movement rates. DSR (Johnson, and Maltz 1996) is an on-demand routing protocol which is based on the concept of source routing. Mobile nodes should maintain a route cache, which specifies the source route, which the mobile should be aware. The entries in the route caches must be continually updated as new routes are learned. This protocol is of two phases: Route maintenance and Route discovery

When a source node has a packet to be sent to the destination, it first checks the route caches for a route to the destination. If it has an unexpired route to the destination then it uses this route to send the packet. If there is no route then the source node initiates a Route discovery by broadcasting a route request packet. This request packet contains the address of the source, the address of the destination and a unique identification number. When a node receives the packet, it checks whether it has a route to the destination. If not it adds its own address to the route record of the packet and forward the packets to the outgoing links. To restrict the number of route requests being forwarded, the nodes forward the route request packet only if the mobile has not seen the request and if the mobile address does not appear in the route record.
A route reply is generated when it reaches the destination. When the packet reaches the destination, the route record specifies the number of hops it has taken. Figure 2.6 explains the formulation of the route record as the route request propagates through the network (Royer and Toh 1999). When a node generating route reply is a destination, it places the route record contained in the route request into route reply. If the responding node is an intermediate node then it appends its cached route-to-route record and generates the route reply. To return the route reply the responding node must have a route to the initiator. If it has a route in the route cache then it uses that route. If symmetric links are provided the node reverses the route in route record. If symmetric links are not supported, the node initiates its own route discovery and piggybacks the route reply on the new route request. Figure 2.7 explains the transmission of the route reply with its associated route record back to the source node.

Figure 2.6 Creation of route record in DSR – Building of the route record during route discovery
Route maintenance is done with route error packets and acknowledgements. Route error packets are transmitted by a node when the data link layer encounters a fatal transmission problem. When a route error packet is received, the hop in error is removed from the nodes, route cache and nodes containing that hop are truncated at that point. Acknowledgements are used to verify the correct operation of links. Such acknowledgments are passive acknowledgements where the mobile node is able to hear the next hop forwarding the packet along the route. Route maintenance can be performed using end-to-end ack rather than hop-by-hop ack. As long as a route exists between two end-to-end host which can communicate, route maintenance is possible. With hop-by-hop ack, the particular hop error can be indicated in the route error packet but with end-to-end ack, the sender assumes that error has occurred in the last hop of the route to the destination.

*Route optimizations:* The study Lee and Gerla 1999 on the simulation study of table driven and on-demand routing protocols for mobile networks explains few
optimization schemes which improve performance and reduce overheads. The following are some of the optimization schemes:

- No propagating route request
- Piggybacking on route discoveries
- Gratuitous route replies
- Gratuitous route errors
- Salvaging
- Snooping

No propagating route request: When the sender originates a request, it sets the time to limit to zero hop, thus allowing the neighbour only to receive the packet. If a neighbour is the destination or has route information to the destination in its cache, it sends a reply to the originator. If no reply is received within the time out period, then the sender floods a route request.

Piggybacking on route discoveries: To eliminate the route acquisition latency, data can be piggybacked on route request packets. If the intermediate node, which has route information to the destination in its cache replies with a route then that node has to construct a data packet and forward it to the destination node in order not to lose any data.

Gratuitous route replies: When a node receives a packet which is not address to itself, then the node refers to the listed source route which has not been traversed yet. If the unprocessed part contains the identification of the node, it realizes that shorter route can be achieved by not visiting the preceding hops in the source route.
Gratuitous route errors: When the source of the broken route receives route error, it piggybacks the received route error on the next route request packet for route discovery. This piggybacking prevents nodes from replying with stale routes.

Salvaging: If an intermediate node of a route detects that the next hop cannot be reached; it searches the route cache for an alternate route. If such a route is found, it substitutes the available route for the stale route in the data header and forwards it. The intermediate node is still required to send a route error back to the sender.

Snooping: When processing data, a node examines the unvisited node in the source route and inserts those routes into its route cache. This snooping enables nodes to have multiple alternate routes for each destination.

Two sources of bandwidth overhead in DSR are route discovery and route maintenance. These occur when new routes are to be discovered or when network topology changes. This overhead can be reduced by caching techniques in each node, at the expense of memory and resources. The remaining source of bandwidth overhead is the source route header included in every packet.

Advantages: Dynamic source routing protocol performs well in environmental conditions such as host density and movement rates. DSR can adapt itself to routing changes when host movement is frequent. It requires less or no overhead during periods when host moves less frequently. DSR does not require periodic transmission of router advertisements or link status packets, thereby reducing the overhead of DSR.
2.4.2.4. Temporally ordered routing algorithm (TORA)

TORA is a uniform destination based reactive protocol designed for use in mobile wireless network (Park, and Corson 1997). Such a network mainly comprises of collection of routers (equipped with wireless transmitters and receivers), which can move freely. The status of communication links between any routers at any instant of time is a function of their positions, transmission power levels, antenna patterns, co-channel interference levels etc. The mobility of the routers and the connectivity factors result in a network with unpredictable changing topology. Congested links are expected characteristics of such a network and wireless links have lower capacity than hardwired links and hence are more prone to congestion. TORA is based on link reversal algorithm. TORA has desirable properties which makes it well suited for wireless networking environment

- Distributed execution
- Provision of loop free routes
- Provision of multiple routes
- Quick establishment of routes
- Minimize algorithmic reactions or communications overhead.

To reduce the communication overhead and to maximize the efficiency, routes are established only when necessary by constructing a directed acyclic graph (DAG) rooted at the destination using a query reply process.
TORA is a highly adaptive loop free distributed routing algorithm based on the concept of link reversal (Park and Corson 1997). TORA can work in a highly dynamic mobile networking environment. It is source initiated and provides multiple routes for any source/destination pair. The concept of TORA is to localize the control message to a very small set of nodes near the occurrence of topological changes. The protocol performs three basic functions: Route creation, Route maintenance and Route erasure.

For route creation and route maintenance, it uses height metric to establish a directed acyclic graph rooted at the destination. Links are assigned directions based on height metric of neighbouring nodes. When there is node mobility the DAG route is broken and route maintenance is necessary to re-establish a DAG rooted at the same destination (Corson and Ephremides 1995).

As shown in figure 2.8 when there is a failure of downstream link, a node generates a reference level, which results in propagation of that reference level by neighbouring nodes, effectively coordinating a structured reaction to failure. Links can be reversed to reflect the changes, thereby adapting to new reference level. This is same as reversing the direction of one or more links when a node has no downstream links.

Timing is an important factor because the height metric is dependent on logical time of link failure. TORA assumes that all nodes have synchronized clocks. TORA metric is quintuple consisting of five elements namely
• Logical time of link failure
• The unique id of the node that defined the new reference level
• A reflection indicator bit.
• A propagation ordering parameter
• The unique id of the node

The first three elements correspond to reference level. A new reference level has to be defined whenever there is a loss in last downstream link due to link failure. TORA’s route erasure phase involves flooding a broadcast clear packet throughout the network to erase invalid routes. TORA uses inter nodal coordination, its instability problem is similar to count to infinity problem in distance vector routing protocol except for oscillations are temporary and route convergence will occur.

TORA is one of the most complicated protocol. With respect to memory requirements, each node must maintain a structure describing the nodes height and the status of all connected links per connection supported by the network (Park and Corson 1997). In terms of CPU and bandwidth requirements, each node must be in constant coordination with neighbouring nodes to detect topology changes and converge. Routing loops can occur when there is a change in topology.

**Advantages:** TORA provides loop free routes, provides multiple routes, establishes route quickly, minimize algorithmic reactions or communication overhead. TORA is best suitable for a large densely packed array of nodes with very low node mobility.
Figure 2.8 TORA

(a) Ad hoc node
Height metric

(1) Link reversal
Link failure

(2) A) Route creation (showing link direction assignment);
B) route maintenance (showing the link reversal phenomenon) in TORA

A

B

C

D

E

F

G

Source

Destination
2.4.2.5. Associativity based routing (ABR)

ABR routes are established based on demand (Toh 1996 and 1997). ABR is a uniform, destination based, reactive protocol. ABR uses end-to-end topology information in route selection, preferring routes that reflect long-lived associations. Only destination vectors are maintained in routing. ABR introduces new routing metrics

- Longevity of a route based on associativity
- Route relaying load of intermediate node supporting existing routes
- Link capacities of selected route.

By associativity, the temporal and connection relationship of mobile host with its neighbours are defined. ABR is free from loops, deadlocks and packet duplicates. ABR defines a new routing for ad-hoc mobile network. This metric is known as degree of association stability. Route selection is based on association stability of mobile nodes. Each node maintains three tables.

Routing table: The nodes which are in the active route of the network stores the route information in the routing table which contains the source and destination id, incoming and outgoing id, hop count to the destination and the number of active routes.

Neighbour table: Each node maintains a neighbour table, which specifies the associativity relationship with surrounding neighbours.
**Seen table:** A seen table is used to prevent a mobile host from processing and forwarding the same Broadcast Query or Localized Query message multiple times.

Each node periodically generates a beacon to specify its existence. When received by a neighbouring node this beacon causes the associativity table to be updated. For each beacon received, the associativity tick of the current node with respect to beaconing node is incremented. Association stability is defined as connection stability of one node with respect to another over time and space. High degree of stability indicates low state of node mobility. Associativity ticks are reset when their neighbours or the node itself move out of proximity. The three phases of ABR are Route discovery, Route reconstruction and Route deletion.

Route discovery is based on broadcast query and await reply (BQ-REPLY) cycle. A node, which desires for a route, broadcasts a BQ message in search of mobiles that have a route to the destination. All nodes on receiving the query (which are not the destination) append their addresses and their associativity ticks with their neighbours along with their QOS information to their query packet. A successor node erases the upstream node neighbour’s associativity tick entries and retains itself the entries concerned with itself and its upstream nodes. By this method the resultant packet arriving at the destination will have the associativity ticks of the nodes along the route to the destination and based on this information the destination can select the best route. When multiple paths have the same number of associativity stability, the one with the smallest number of hops must be selected. The destination sends a reply back to the source along this path. Nodes propagating this reply mark this node as...
valid and all other routes remain inactive, thereby avoiding the duplicate packets arriving at the destination.

Route Reconstruction phase (RRC) consists of partial route discovery, invalid route erasure, valid route updates and new route discovery depending on how nodes move along the route. Movement of source results in BQ-REPLY process. The RN \{1\} message indicates the erasure of route entries associated with downstream nodes. When the destination moves, the upstream node erases its route and determines if the node is reachable by a localized query (LQ \{H\}) process, where H refers to hop count from upstream node to the destination. When the destination receives the LQ packet it REPLY’S with the best partial route, otherwise the initiating node times out and the process backtracks with the next upstream node. RN \{0\} message is used to erase the invalid route of next upstream node and inform this node should invoke the LQ \{H\} process. If this process also results in backtracking, then the LQ process must be discontinued and a new BQ process is initiated at the source.

When the discovered route is not required any more, then the source node initiates a ROUTE DELETE broadcasts so that all nodes along the route updates the routing table.

Data transmission: For efficient channel utilization, ABR uses a simple and short packet header. The data packet header contains only the neighbouring node information rather than all nodes in the route.
Flow control is achieved by use of passive acknowledgements. When node A receives a packet and forwards the packet to the next hop node B and if A hears B transmitting to another node, then this is known as passive acknowledgement. This technique is used in packet radio (Jubin and Tornow 1987). If a node does not receive passive acknowledgement within a time out period, then it retransmits the packet for number of times. If no ack are received after few attempts then the mobile host is considered to be out of radio arrange or powered down and a RRC phase must be invoked.

Both ABR and DSR have less overhead compared to DBF. Increasing the mobility speed makes ABR, more efficient than DSR. If a node in the path is unreachable in ABR the upstream of a migrated node starts the LQ process to find a new partial route without the intervention from source, hence minimizing the transmission of control message. ABR selects a route where nodes in the path are associatively stable, and have a light load. It has a shorter delay and this becomes more obvious as speed increases.

**Advantages:** ABR is free from loops, deadlocks and packet duplicates. ABR defines a new routing for Ad-hoc network. This metric is known as degree of association stability.

**Disadvantages:** Each node maintains three tables viz., routing table, seen table and neighbour table in the network. This becomes a constant overhead even when there is no traffic in the network. ABR demands more power in order to process the beacon message.
2.4.2.6. Signal Stability routing (SSR)

Another on demand protocol is Signal Stability based adaptive Routing protocol (SSR) (Dube 1997). SSR selects the route based on signal strength between nodes and the node’s location stability. Routes have to be chosen based on stronger connectivity. SSR is divided into two routing protocols

- Dynamic Routing Protocol (DRP)
- Static Routing Protocol (SRP)

DRP maintains a signal stability table (SST) and a routing table (RT). SST records the signal strength of each neighbouring node, which is obtained by periodic beacons from link layer of each neighbouring node. Signal strength recorded may be strong or weak. All transmissions received are processed by DRP and after the updating the table, DRP passes the received packet to SRP.

The SRP processes these packets by passing the packet up to the stack for if it is to the intended receiver or looks up for the destination in the routing table and then forwards the packet. If it does not find a route in the RT, a route search process is initiated to find the route. Route requests are propagated throughout the network and they are forwarded to the next hop if it is being received by a strong channel and has not been previously processed (to prevent looping). The destination chooses the first arriving route search packet to send back because it is highly probable that packet has arrived through least congested or shortest path. The DRP then reverses the selected
route and sends a route reply message back to the initiator. The DRP of nodes along this path then performs path update in their RT accordingly.

Route search packets arriving at the destination have chosen a path of strongest signal stability, because the packets would have dropped if they have arrived over weak channel. If the source does not receive a Route-Reply within a particular timeout period then it changes the PREF field in the header so that weak channels are also acceptable, because they may be the only links through which the packets can be propagated.

When a failed link is detected within the network, the intermediate nodes send an error message specifying which channel has failed. The source then initiates another route search process to find a new path to the destination. The source also sends an erase message to notify all nodes of the broken link.

**Advantages:** SSR is found to be more stable and alive for longer period, resulting in fewer route reconstruction. The intermediate nodes cannot send reply to route request sent towards destination and this results in longer delay before a route can be discovered.

### 2.5. Summary

In this chapter the basic concepts in wireless networks such as network architectures, wireless channels, MAC layer protocols and network layer routing protocols were briefly discussed. From the examples of ad-hoc routing protocols
presented in this chapter, it should be evident that the selection of an ad-hoc routing protocol for use in an ad-hoc network requires careful thought. Parameters such as network size, mobility and traffic load have an impact on the suitability of each protocol. This, together with the spontaneous nature of ad-hoc networks, easily turns the selection of a routing protocol into a complicated task. The primary research goal is to evaluate the routing protocols based on different mobility models. The following chapters will summarize the research and contributions towards this goal. To achieve this, simulations are done based on Network Simulator NS-2.