Chapter 1

Introduction

Recent progress in mobile wireless networking has provided a major impetus towards the development of self-created, self-organized and rapidly deployable network architecture referred to as Mobile Ad Hoc Networks (MANETs). Although the application area of MANETs was initially proposed for environments of battle-field communications and disaster recovery, the evolution of the Multimedia Technology and the commercial interest of companies to reach widely civilian applications, have made Quality of Service (QoS) in such networks an avoidable task. In the recent years, the efforts of bringing QoS in wire-based networks are becoming a reality. Many ideas inherited from the wire-based networks can be ported onto MANETs if they take into consideration the Bandwidth Constraints, Dynamic Topology and the Constrained Processing and Storing Capabilities of MANETs.

Many different protocols have been proposed to support QoS in MANETs, each based on different assumptions and intuitions. Current proposals in the literature have attempted to provide QoS support in terms of delay and bandwidth. The bandwidth, jitter, packet delivery ratio and end-to-end delay metrics are taken into consideration as QoS constraints for the proposed protocol. The performance metrics are used to show the efficiency and effectiveness of protocol and to perform a comparison study with other prominent QoS aware routing protocols.

MANET is a developing area of research. These networks do not rely on extraneous hardware, which makes them an ideal candidate for rescue
Routing is a core problem in networks for sending data from one node to another. Wireless Ad Hoc network which is also called Mobile Ad Hoc multi-hop wireless network, is a collection of wireless mobile hosts forming a temporary network without the aid of any established infrastructure or centralized administration [1]. MANETs are characterized by a dynamic, multi-hop, rapidly changing topology. Such networks are aimed to provide communication capabilities to areas where limited or no communication infrastructures exist.

MANETs can also be deployed to allow the communication devices to form a dynamic and temporary network among them. A Mobile Ad Hoc network is receiving attention due to several salient characteristics such as: 1) Dynamic topologies 2) Bandwidth-constrained links 3) Energy constrained operation 4) Limited physical security. Therefore the routing protocols for wired networks cannot be directly used for wireless networks. Some examples of the possible uses of Ad Hoc networking include students using laptop computers to participate in an interactive lecture, business associates sharing information during a meeting, soldiers relaying information for situational awareness on the battlefield and emergency disaster relief personnel coordinating efforts after a hurricane or earthquake. A MANET uses multi-hop routing instead of a static network infrastructure to provide network connectivity. Several routing protocols have been proposed for MANETs. Initially, in this work, number of ways of classification or categorization of existing routing protocols and the survey on the performance comparison of important protocols such as AODV [2], DSR [3] and DSDV [4] are presented. A number of researchers
have done the performance analysis of Ad Hoc Routing protocols over the years using different simulators like NS2, Qualnet, Opnet etc.

From the survey of QoS Models and Protocols of the IP network such as IntServ, DiffServ, RSVP, FQMM (Flexible QoS Model for MANET) etc., it is observed that they have affected the evolution of models and protocols in the Wireless Ad Hoc world. Upon comparing the performances of both routing and QoS routing protocols, results show that QoS enabled routing protocol shows a significant improvement in QoS metrics over the normal routing protocol.

Recent studies in Ad Hoc routing show that choosing the route based on the first arrived Route Request (RREQ) packet without considering the route stability leads to frequent route failures [5] and that using signal strength often yields more reliable routes [6]. Routing overhead incurred in a network can be attributed to the number of control packets generated in maintaining and discovering the routes.

In a multi-hop Ad Hoc wireless network, nodes are mostly powered by a battery. It has been shown that the energy or lifetime [7] of any node is limited by its residual battery power [8]. Every packet processed (received/forwarded) by a node consumes a part of the limited battery power. Repeated use of a single node/path for data transmission results in total depletion of battery power on a single node or all the nodes along the path. This leads to data losses as the nodes fail along the path. If the node’s residual battery power was taken into consideration for the path selection, the occurrence of the said possibility could have been minimized. These observations have been made in the past and protocols like, ‘Minimum Battery Cost Routing’, ‘Min-Max Battery Routing’, ‘Conditional Max-Min Battery Capacity Routing’, ‘Maximum Residual Packet Capacity (MRPC)’, ‘Power Aware Source Routing (PSR)’ etc., have been developed [9] taking this factor into account.
1.1 Overview of Mobile Ad Hoc Networks (MANETs)

A MANET consists of mobile platforms (e.g., a router with multiple hosts and wireless communications devices) herein simply referred to as “nodes” which are free to move about arbitrarily. The nodes may be located in or on airplanes, ships, trucks, cars, perhaps even on people or very small devices, and there may be multiple hosts per router. A MANET is an autonomous system of mobile nodes. The system may operate in isolation, or may have gateways to and interface with a fixed network. In the latter operational mode, it is typically envisioned to operate as a “stub” network connecting to a fixed internetwork. Stub networks carry traffic originating at and/or destined for internal nodes, but do not permit exogenous traffic to “transit” through the stub network.

MANET nodes are equipped with wireless transmitters and receivers using antennas which may be omnidirectional (broadcast), highly-directional (point-to-point), possibly steerable, or some combination thereof. At a given point in time, depending on the nodes’ positions and their transmitter and receiver coverage patterns, transmission power levels and co-channel interference levels, a wireless connectivity in the form of a random, multi-hop graph or “Ad Hoc” network exists between the nodes. This Ad Hoc topology may change with time as the nodes move or adjust their transmission and reception parameters [10]. A typical MANET is shown in the Figure 1.1.
1.2 MANET’s Salient Characteristics

a) **Dynamic topologies:** Nodes are free to move arbitrarily; thus, the network topology which is typically multi-hop, may change randomly and rapidly at unpredictable times, and may consist of both bidirectional and unidirectional links. The Figure 1.2 shows the reachability of node A.

![Figure 1.2: Dynamic Topological changes](image)

b) **Bandwidth-constrained, variable capacity links:** Wireless links will continue to have significantly lower capacity than their hardwired counterparts. In addition, the realized throughput of
wireless communications after accounting for the effects of multiple access, fading, noise, and interference conditions, etc. is often much less than a radio’s maximum transmission rate. One effect of the relatively low to moderate link capacities is that congestion is typically the norm rather than the exception, i.e., aggregate application demand will likely approach or exceed network capacity frequently. As the mobile network is often simply an extension of the fixed network infrastructure, Mobile Ad Hoc users will demand similar services. These demands will continue to increase as multimedia computing and collaborative networking applications rise.

c) **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.

d) **Limited physical security:** Mobile wireless networks are generally more prone to physical security threats than are fixed-cable nets. The increased possibility of eavesdropping, spoofing, and denial-of-service attacks should be carefully considered. Existing link security techniques are often applied within wireless networks to reduce security threats. As a benefit, the decentralized nature of network control in MANETs provides additional robustness against the single points of failure of more centralized approaches.

### 1.3 Applications of MANETs

The field of wireless networking emerges from the integration of personal computing, cellular technology, and internet. This is due to the increasing interactions between communication and computing, which are changing information access from “anytime, anywhere” into “all the time, everywhere”. At present, a large variety of networks exist, ranging from the well-known infrastructure of cellular networks to non-infrastructure wireless Ad Hoc networks. With the increase of portable devices as well as
progress in wireless communication, Ad Hoc networking is gaining importance with the increasing number of widespread applications.

Ad Hoc networking [11] can be applied anywhere where there is little or no communication infrastructure, or the existing infrastructure is expensive or inconvenient to use. Ad Hoc networking allows the devices to maintain connections to the network as well as easily adding and removing devices to and from the network. The set of applications for MANETs is diverse, ranging from large-scale, mobile, highly dynamic networks, to small, static networks that are constrained by power sources. Besides the legacy applications that move from traditional infrastructure environment into the Ad Hoc context, a great deal of new services can and will be generated for the new environment. Typical applications include:

a) **Military battlefield**: Military equipment now routinely contains some sort of computer equipment. Ad Hoc networking would allow the military to take advantage of common place network technology to maintain an information network among the soldiers, vehicles, and military information headquarters. The basic techniques of Ad Hoc network came from this field.

b) **Commercial sector**: Ad Hoc network can be used in emergency/rescue operations for disaster relief efforts, e.g., in fire, flood, or earthquake. Emergency rescue operations must take place where non-existing or damaged communications infrastructure and rapid deployment of a communication network is needed. Information is relayed from one rescue team member to another over a small handheld device. Other commercial scenarios include e.g., ship-to-ship Ad Hoc mobile communication, law enforcement, etc.

c) **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, for example, conference or classroom. Another appropriate local level application might be in home networks
where devices can communicate directly to exchange information. Similarly in other civilian environments like taxicab, sports stadium, boat and small aircraft, mobile Ad Hoc communications will have many applications.

d) **Personal Area Network (PAN):** Short-range MANET can simplify the intercommunication between various mobile devices (such as a PDA, a laptop, and a cellular phone). Tedious wired cables are replaced with wireless connections. Such an Ad Hoc network can also extend the access to the Internet or other networks by mechanisms, for example, Wireless LAN (WLAN), GPRS (Global Packet Radio Service), and UMTS (Universal Mobile Telecommunication Systems). The PAN is potentially a promising application field of MANET in the future pervasive computing context.

### 1.4 Routing Classification in Ad Hoc Networks

Routing in wireless Ad Hoc networks is clearly different from routing found in traditional infrastructure networks. Routing in Ad Hoc networks needs to take into account many factors including topology, selection of routing path and routing overhead, and it must find a path quickly and efficiently. Ad Hoc networks generally have lower available resources as compared to infrastructure networks and hence there is a need for optimal routing. Also, the highly dynamic nature of these networks means that routing protocols have to be specifically designed for them, thus motivating the study of protocols that aims at achieving routing stability. Designing a routing protocol for Ad Hoc networks is challenging because of the need to take into account two contradictory factors:

a) A node needs to know at least the “reachability” information to its neighbors for determining a packet route.

b) The network topology can change quite often.
Furthermore, as the number of network nodes can be large, finding a route to the destinations also requires large and frequent exchange of routing control information among the nodes. Thus, the amount of update traffic can be quite high, and it is even higher when the network includes high mobility nodes, which can impact the route overhead of routing protocols in such a way that there might be no bandwidth leftover for the transmission of data packets [12]. In wireless Ad Hoc networks, the communication range of a node is often limited and not all nodes can directly communicate with one another. Nodes are required to relay packets on behalf of other nodes to allow communication across the network. Since there is no pre-determined topology or configuration of fixed routes, an Ad Hoc routing protocol is used to dynamically discover and maintain up-to-date routes between communicating nodes.

1.4.1 Proactive versus Reactive Approaches

Ad Hoc routing protocols may generally be categorized as being either proactive or on-demand (reactive) according to their routing strategy [13]. Proactive protocols require that nodes in a wireless Ad Hoc network should keep track of routes to all possible destinations so that when a packet needs to be forwarded, the route is already known and can be used immediately. Any changes in topology are propagated through the network, so that all nodes know of those changes in topology. Examples include “destination-sequenced distance-vector” (DSDV) routing [14], “Cluster-head Gateway Switch Routing Protocol” (CGSR) [15], and “Wireless Routing Protocol” (WRP) [16].

On-demand protocols only attempt to build routes when desired by the source node so that the network topology is detected as needed (on-demand). When a node wants to send packets to some destination but has no routes to the destination, it initiates a route discovery process within the network. Once a route is established, it is maintained by a route maintenance procedure until the destination becomes inaccessible or until the route is no longer needed. Examples include “Ad Hoc on-demand
distance vector routing” (AODV) [17][18], “Dynamic Source Routing” (DSR) [19], and “Cluster Based Routing protocol” (CBRP) [20]. Proactive protocols have the advantage that new communications with arbitrary destinations experience minimal delay, but suffer from the disadvantage of the additional control overhead to update routing information at all nodes. To cope with this shortcoming, reactive protocols adopt the inverse approach by finding a route to a destination only when needed. Reactive protocols often consume much less bandwidth than proactive protocols, but they will typically experience a long delay for discovering a route to a destination prior to the actual communication. However, because reactive routing protocols need to broadcast route requests, they may also generate excessive traffic if route discovery is required frequently. Table 1.1 shows the distinction between Proactive and Reactive routing protocols.

Table 1.1: Proactive vs. Reactive Routing

<table>
<thead>
<tr>
<th>Differences</th>
<th>Proactive</th>
<th>Reactive</th>
</tr>
</thead>
</table>
| 1. Routing  | a. Every node maintains and updates topology information periodically  
              b. Routes are established before a transmission request is made | Routes are established after a transmission request is made (on-demand) |
| 2. Advantages | No delay needed to establish route | Less control overhead at nodes |
| 3. Disadvantages | a. Need dedicated memory to store long term routing information  
                        b. Additional control overhead to update information at all nodes | Experience a long delay for establishing route |
1.4.2 Routing Protocols Overview

The Figure 1.3 shows the classification of MANET routing protocols.

![Classification of Routing Protocols](image)

Figure 1.3: Classification of Routing Protocols

a) Destination Sequenced Distance Vector (DSDV)

DSDV [21][22][23] is a Proactive routing protocol that solves the major problem associated with the Distance Vector routing of wired networks i.e., Count-to-infinity, by using Destination sequence numbers. Destination sequence number is the sequence number as originally stamped by the destination. The DSDV protocol requires each mobile station to advertise, to each of its current neighbours, its own routing table (for instance, by broadcasting its entries). The entries in this list may change fairly dynamically over time, so the advertisement must be made often enough to ensure that every mobile computer can almost always locate every other mobile computer. In addition, each mobile computer agrees to relay data packets to other computers upon request. At all instants, the DSDV protocol guarantees loop-free paths to each destination.
Routes with more recent sequence numbers are always preferred as the basis for making forwarding decisions, but not necessarily advertised. Of the paths with the same sequence number, those with the smallest metric will be used.

The routing updates are sent in two ways: a “full dump” and an “incremental update”. A full dump sends the full routing table to the neighbours and could span many packets whereas, in an incremental update only those entries from the routing table are sent that have a metric change since the last update and it must fit in a packet. When the network is relatively stable, incremental updates are sent to avoid extra traffic and full dump are relatively infrequent. In a fast changing network, incremental packets can grow big, so full dumps will be more frequent.

The updates can be time triggered (periodic) or event triggered. When any new or substantially modified route information is received by a Mobile Host, the new information will be retransmitted soon (subject to constraints imposed for damping route fluctuations) [22]. When a stabilized route shows a different metric for some destination, it would likely constitute a significant change that needs to be advertised after stabilization. If a new sequence number for a route is received, but the metric stays the same, that would be unlikely to be considered as a significant change. Newly recorded routes are scheduled for immediate advertisement to the current Mobile Host’s neighbours. Routes which show an improved metric are scheduled for advertisement at a time which depends on the average settling time for routes to the particular destination under consideration.

A broken link is described by a metric of infinity (i.e., any value greater than the maximum allowed metric). When a link to a next hop has broken, any route through that next hop is immediately assigned infinity metric and assigned an updated sequence number. Since this qualifies as a substantial route change, such modified routes are immediately disclosed in a broadcast routing information packet.
b) **Wireless Routing Protocol (WRP)**

WRP [11][21][22][23] belongs to the class of path-finding algorithm with the exception of avoiding the count-to-infinity [24] problem by forcing each node to perform consistency checks of predecessor information reported by all its neighbours. The novel part of this protocol is it achieves loop freedom. Each node maintains 4 tables: Distance table, Routing table, Link cost table & Message retransmission list table. Link changes are propagated using update messages sent between neighbouring nodes. Hello messages are periodically exchanged between neighbours.

In WRP, a node checks the consistency of its neighbours after detecting any link change. A consistency check helps to eliminate loops and speed up convergence. One shortcoming of WRP is that it needs large memory storage and computing resources to maintain several tables. Moreover, as a proactive routing protocol, it has a limited scalability and is not suitable for large MANETs.

c) **Cluster-head Gateway Switch Routing Protocol (CGSR)**

In CGSR [21][22][23][25], mobile nodes are grouped into clusters and each cluster has a cluster head. A cluster head can control a group of Ad Hoc hosts and clustering provides framework for network separation (among clusters), channel access, routing and also bandwidth allocation. It uses DSDV as the underlying routing algorithm and each node maintains a cluster member table and a routing table. By forming several clusters, this protocol achieves a distributed processing mechanism in the network. However, one drawback of this protocol is that, frequent change or selection of cluster heads might be resource hungry and it might affect the routing performance. Since CGSR uses DSDV protocol as the underlying routing scheme, it has the same overhead as DSDV. However, it modifies DSDV by using a hierarchical cluster-head-to-gateway routing approach to route traffic from source to destination.
d) Dynamic Source Routing (DSR)

DSR [21][22][23][26] is a reactive protocol i.e., it doesn’t use periodic advertisements. It computes the routes when necessary and then maintains them. Source routing is a routing technique in which the sender of a packet determines the complete sequence of nodes through which the packet has to pass; the sender explicitly lists this route in the packet’s header, identifying each forwarding “hop” by the address of the next node to which to transmit the packet on its way to the destination host.

There are two significant stages in working of DSR: Route Discovery and Route Maintenance. A host initiating a route discovery broadcasts a route request packet which may be received by those hosts within wireless transmission range of it. The route request packet identifies the host, referred to as the target of the route discovery, for which the route is requested. If the route discovery is successful the initiating host receives a route reply packet listing a sequence of network hops through which it may reach the target. In addition to the address of the original initiator of the request and the target of the request, each route request packet contains a route record, in which is accumulated a record of the sequence of hops taken by the route request packet as it is propagated through the network during this route discovery.

While a host is using any source route, it monitors the continued correct operation of that route. This monitoring of the correct operation of a route in use is called route maintenance. When route maintenance detects a problem with a route in use, route discovery may be used again to discover a new, correct route to the destination. To optimize route discovery process, DSR uses cache memory efficiently. Suppose a host receives a route request packet for which it is not the target and is not already listed in the route record in the packet, and for which the pair (initiator address, request id) is not found in its list of recently seen requests; if the host has a route cache entry for the target of the request, it may append this cached route to the accumulated route record in the packet, and may return this route in a route reply packet to the initiator.
without propagating (re-broadcasting) the route request. The delay for route discovery and the total number of packets transmitted can be reduced by allowing data to be piggybacked on route request packets.

DSR uses no periodic routing advertisement messages, thereby reducing network bandwidth overhead, particularly during periods when little or no significant host movement is taking place. DSR has a unique advantage by virtue of source routing. As the route is part of the packet itself, routing loops, either short-lived or long-lived, cannot be formed as they can be immediately detected and eliminated.

e) Ad Hoc On-demand Distance Vector (AODV)

AODV [21][22][23][27] is essentially a combination of both DSR and DSDV. It borrows the basic on-demand mechanism of Route Discovery and Route Maintenance from DSR, use of hop-by-hop routing, sequence numbers, and periodic beacons from DSDV. It uses destination sequence numbers to ensure loop freedom at all times and there by avoids the Bellman-Ford "count-to-infinity" problem. AODV offers quick convergence when the Ad Hoc network topology changes. Route Requests (RREQs), Route Replies (RREPs), and Route Errors (RERRs) are the message types defined by AODV. These message types are received via UDP, and normal IP header processing applies.

As long as the endpoints of a communication connection have valid routes to each other, AODV does not play any role. When a route to a new destination is needed, the node broadcasts a RREQ to find a route to the destination. A route can be determined when the RREQ reaches either the destination itself, or an intermediate node with a ‘fresh enough’ route to the destination. A ‘fresh enough’ route is a valid route entry for the destination whose associated sequence number is at least as great as that contained in the RREQ. The route is made available by unicasting a RREP back to the origination of the RREQ. Each node receiving the request caches a route back to the originator of the request, so that the RREP can
be unicast from the destination along a path to that originator, or likewise from any intermediate node that is able to satisfy the request.

If intermediate nodes reply to every transmission of a given RREQ, the destination does not receive any copies of it. In this situation, the destination does not learn of a route to the originating node. This could make the destination to initiate a route discovery (for example, if the originator is attempting to establish a TCP session). In order that the destinations learn of routes to the originating node, the originating node should set the “gratuitous RREP” (‘G’) flag in the RREQ if for any reason the destination is likely to need a route to the originating node. If in response to a RREQ with the ‘G’ flag set, an intermediate node returns a RREP, it MUST also unicast a gratuitous RREP to the destination node.

Nodes monitor the link status of next hops in active routes. In order to maintain routes, AODV normally requires that each node periodically transmit a HELLO message, with a default rate of once per every second. Failure to receive three consecutive HELLO messages from a neighbour is taken as an indication that the link to the neighbour in question is down. When a link break in an active route is detected, a RERR message is used to notify other nodes that the loss of that link has occurred. The RERR message indicates those destinations which are now unreachable due to the loss of the link. In order to enable this reporting mechanism, each node keeps a “precursor list”, containing the IP address for each of its neighbours that are likely to use it as a next hop towards the destination that is now unreachable.

f) **Temporary-Ordered Routing Algorithm (TORA)**

TORA [21][22][23] is highly adaptive, loop-free, distributed routing algorithm based on the concept of link reversal. It is proposed to operate in a highly dynamic mobile networking environment. It is source initiated and provides multiple routes for any desired source/destination pair. This algorithm requires the need for synchronized clocks. There are 3 basic
functions of the protocol, namely route creation, route maintenance and route erasure.

TORA defines a parameter termed height. Height is a measure of the distance of the responding node’s distance up to the required destination node. In the route discovery phase, this parameter is returned to the querying node.

g) Associativity Based Routing Protocol (ABR)

ABR [21][22][23] is designed for Ad Hoc networks where mobile computers act as routers and packet forwarders in a wireless environment with no base stations. ABR is based on the concept of associativity. The protocol is source-initiated, thus there is no need for periodic route updates. In this routing protocol, a route is selected based on the association stability of mobile nodes. Every node in the Ad Hoc network learns its ‘Associativity’ with its surrounding nodes to determine the best route. Stability is determined using ‘associativity ticks’. Association in ABR takes up a few metrics such as link delay, signal strength, power life, route relaying load, period of presence or spatial and temporal characteristics. Routes are only chosen when they have a high degree of associativity or have high associativity ticks. If there are multiple paths with the same overall degree of association stability, the route with the minimum number of hops is selected.

h) Signal Stability Routing (SSR)

Descendent of ABR and ABR predates SSR [21][22]. SSR is similar to ABR but it selects routes based on signal strength between nodes and on a node’s location stability. SSR’s route selection criteria, has effect of choosing routes that have ‘stronger’ connectivity and it can be divided into Dynamic Routing Protocol (DRP) or Static Routing Protocol (SRP). DRP receives all the transmissions and, after processing, it passes those to the SRP. SRP passes the packet to the node’s upper layer stack if it is the destination. Otherwise, it looks for the destination in the routing table and
forwards the packet. If there is no entry in the routing table for that destination, it initiates the route finding process.

1.5 MANET QoS Overview

Quality of Service (QoS) is the performance level of a service offered by the network to the user. The goal of QoS provisioning is to achieve more deterministic network behaviour, so that information carried by the network can be better delivered and network resources can be better utilized. The network needs are governed by the service requirements of end user applications. The network is expected to guarantee a set of measurable pre-specified service attributes to the users in terms of end-to-end performance, such as delay, bandwidth, probability of packet loss, delay variance (jitter), etc. Power consumption is another QoS attribute which is more specific to MANETs [25]. After receiving a service request from the user, the first task is to find suitable loop-free path from the source to the destination that will have the necessary resources available to meet the QoS requirements of the desired service. This process is known as QoS routing.

The QoS provisioning approaches can be broadly classified into two categories: hard QoS and soft QoS approaches. If QoS requirements of a connection are guaranteed to be met for the whole duration of the session, the QoS approach is termed as hard QoS approach. For e.g., air traffic control, nuclear reactor control etc. If the QoS requirements are not guaranteed for the entire session, the QoS approach is termed as soft QoS approach. For e.g., voice, video transmission etc. Given the nature of network dynamics of Ad Hoc wireless networks, it is very difficult to provide hard QoS guarantees to user applications.

1.5.1 QoS Metrics

QoS metrics can be divided into different groups, e.g. additive metrics, multiplicative metrics, and concave metrics. Let $d(n_i, n_j)$ be a metric for
link \((n_i, n_j)\) and \(p = (n_1, n_2, \ldots, n_m)\) be a path between nodes \(n_1\) and \(n_m\). Then the named metrics are defined as follows [28]:

Additive: \(d(p) = d(n_1, n_2) + d(n_2, n_3) + \ldots + d(n_{m-1}, n_m)\) \hspace{1cm} (1.1)

Multiplicative: \(d(p) = d(n_1, n_2) \times d(n_2, n_3) \times \ldots \times d(n_{m-1}, n_m)\) \hspace{1cm} (1.2)

Concave: \(d(p) = \min (d(n_1, n_2), d(n_2, n_3), \ldots, d(n_{m-1}, n_m))\) \hspace{1cm} (1.3)

The most commonly used metrics in QoS networks are bandwidth and delay. Bandwidth (concave) denotes the bandwidth along a certain path and is limited by the link with the lowest bandwidth along this path. Delay (additive) indicates the time between sending out a packet from the source node and reception of this packet at the destination node. The metric cost does not belong to any of the groups above, as it is more abstract and can be defined by any function.

Besides these metrics, there are other interesting metrics for QoS networks. The number of hops (additive) represents the number of links in a path. Jitter (additive) denotes the variation between expected and actual reception time of a packet. Energy (additive) takes the energy needed to send a packet from source to destination into account. Alternatively, energy can also be handled as a concave metric, where a (mobile) node has to provide a certain energy level, to be considered as part of a route. Loss probability (multiplicative) refers to the probability of a packet to be lost on its way to the destination node, e.g. because of collisions, topology changes or weak radio signals. Further QoS metrics include, e.g., signal strength (concave) and distance (additive).

QoS routing protocols utilize subsets of these metrics. In many cases, only single metrics like bandwidth or delay or specific groups of metrics, e.g. additive metrics, are taken into account.

1.5.2 Quality of Service models for Internet

A lot of work is done to support QoS in the Internet. A QoS model for MANET, however, should be able to overcome the challenges of MANET,
e.g. dynamic topology and time-varying link capacity. The QoS model for MANET has extended the traditional Internet models to make them suitable for MANET.

At present Internet applies best effort (BE) IP forwarding, i.e., the network attempts to deliver all traffic as soon as possible within the limits of its abilities, but without guarantees related to throughput, delay or packet loss. It is left up to the end systems to cope with network transport impairments.

Although best effort will remain adequate for most applications, QoS support is required to satisfy the growing need for multimedia over IP, like video streaming or IP telephony. The existing QoS models can be classified into two types according to their fundamental operation:

- The Integrated Services (IntServ) framework provides explicit reservations end-to-end.
- The Differentiated Services (DiffServ) architecture offers hop-by-hop differentiated treatment of packets.

Later, two more QoS models FQMM (Flexible QoS model for MANET) and SWAN (Service Differentiation in Wireless Ad Hoc Network) [29] based on IntServ and DiffServ have been developed.

a) **Integrated Services (IntServ)**

The IntServ approach [30] aims to provide applications with a guaranteed share of bandwidth. IntServ operates on a per-flow basis, and the requested QoS for a flow is either fully granted or denied.

Three main services are provided to the applications:

i. **Guaranteed services** [31] provide an assured amount of bandwidth, strict end-to-end delay bounds, and minimal queuing delay to packets.

ii. **Controlled load** services [32] give a service that is as close as possible to a best-effort service in a lightly loaded network.
iii. Best effort services are characterized by the absence of a QoS specification. The first two service classes use parameters, such as token bucket rate and size, peak data rate, and minimum and maximum packet size. These provide detailed information about the intended packet stream, so that routers are able to produce detailed reservations.

The IntServ approach assumes that an explicit setup mechanism is used to convey resource requests to routers so that they can provide the requested services to flows that require them. Moreover, the signaling must establish and keep the reservation state in order to guarantee the resources promised. Resource ReSerVation Protocol (RSVP) [33] can be used to create and maintain the required flow-specific states in network elements allowing them to provide the requested services. RSVP is a signaling protocol that applications may use to reserve resources for all kinds of flows in an IP network. The network routers respond by explicitly admitting or rejecting RSVP requests.

b) Differentiated Services (DiffServ)
DiffServ approach [34] does not define any signaling mechanisms, but instead, it provides QoS by dividing traffic into a small number of classes and allocating network resources on a per-class basis. The class is marked directly on the packet, in the 6 bit DiffServ Code Point (DSCP) field. The MBZ field stands for Must Be Zero.

DSCP field is part of the original Type of Service (ToS) field in the IP header. The IETF redefined the meaning of the little-used ToS field, splitting it into the 6-bit DSCP field and a 2-bit unused field. The unused field is being allocated to the Explicit Congestion Notification (ECN) mechanisms [35], as shown in Figure 1.4.
The basic goal of the Differentiated Services architecture is to meet the performance requirements of the users. Different traffic classes have different priority levels and scheduling algorithms have to ensure that high priority packets are forwarded before low priority ones.

The DSCP determines the QoS behavior of a packet at a particular node in the network. This is called the per-hop behavior (PHB) and is expressed in terms of the scheduling and drop preference that a packet experiences. From an implementation point of view, the PHB translates to the packet queue used for forwarding, the drop probability in case the queue exceeds a certain limit, the resources (buffers and bandwidth) allocated to each queue, and the frequency at which a queue is serviced.

1.5.3 Quality of Service in Ad Hoc Networks

This section discusses unique issues and difficulties for supporting QoS in a MANET environment and ends up showing the major drawbacks of each of the two QoS architectures described above with respect to these characteristics.

1.5.3.1 Special Issues and Difficulties in MANETs

MANETs differ from the traditional wired Internet infrastructures. The differences introduce difficulties for achieving QoS in such networks. The following list itemizes some of the problems:
a) *Dynamic topologies*: Nodes are free to move arbitrarily; thus, the network topology which is typically multihop - may change randomly and rapidly at unpredictable times, and may consist of both bidirectional and unidirectional links.

b) *Bandwidth-constrained, variable capacity links*: Wireless links will continue to have significantly lower capacity than their hardwired counterparts. In addition, the realized throughput of wireless communications - after accounting for the effects of multiple access, fading, noise, and interference conditions, etc. is often much less than a radio's maximum transmission rate. One effect of the relatively low to moderate link capacities is that congestion is typically the norm rather than the exception, i.e. aggregate application demand is likely approach or exceed network capacity frequently. As the mobile network is often simply an extension of the fixed network infrastructure, MANET users will demand similar services. These demands will continue to increase as multimedia computing and collaborative networking applications go up.

c) *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.

### 1.5.3.2 Drawbacks of the different QoS Models

a) **IntServ in MANETS**

IntServ has the following salient shortcomings in MANET environments:

i. *Scalability*: IntServ provides per-flow granularity, so the amount of state information increases proportionally with the number of flows. This results in a storage and processing overhead on routers, which is a well-known scalability problem of IntServ.
ii. The scalability problem is less likely to occur in current MANETs considering the small number of flows, the limited size of the network and the bandwidth of the wireless links. On the other hand, as the quality of wireless technology increases rapidly, high speed and large size MANETs may be a matter of fact for some day. Though one could argue that whenever large high-performance MANETs are developed in future, processing and storing capabilities will increase as well. In highly dynamic networks such as MANETs this is no promising approach since routes may change very fast and the adaptation process of protocols using a complex handshaking mechanism would just be too slow. Furthermore, the signaling overhead while maintaining the connection is a potential problem as well.

b) DiffServ in MANETS

The main drawbacks of a DiffServ approach in MANETs are listed below:

i. *Soft QoS guarantees:* DiffServ uses a relative-priority scheme to map the quality of service requirements to a service level. This aggregation results in a more scalable but also in more approximate service to user flow.

ii. *SLA (Service Level Agreement):* DiffServ is based on the concept of SLAs. In the Internet an SLA is a kind of contract between a customer and its Internet Service Provider (ISP) that specifies the forwarding service the customer should receive. The Administration of a DiffServ domain must assure that sufficient resources are provisioned to support the SLAs committed by the domain. Moreover, the DiffServ boundary nodes are required to monitor the arriving traffic for each service class and to perform traffic classification and conditioning to enforce the negotiated SLAs.
iii. In general if someone acquires QoS parameters and pays for such parameters, there must be some entity which assures him/her. In a completely Ad Hoc topology where there is no concept of service provider and client and where there are only clients it would be quite difficult to innovate QoS, since there is no obligation from someone to the other. This makes QoS almost infeasible.

iv. Ambiguous core network: The benefit of DiffServ is that only traffic classification and conditioning has to be done at the boundary nodes. This makes quality of service provisioning much easier in the core of the network. In MANETs there is no clear definition of what is the core network because every node is a potential sender, receiver and router. This drawback would again take us back to the IntServ model where several separate flow states are maintained.

1.6 Organization of the thesis

The thesis comprises of 10 chapters, out of which chapters 5, 6, 7, 8 and 9 focus on the contribution. Brief details about each are given below:

Chapter 1 is the introduction chapter. In this chapter, brief introduction to various concepts and terminologies that are used throughout the thesis is given. It also explains the need of research in Routing in wireless mobile Ad Hoc networks which is clearly different from routing found in traditional infrastructure networks.

Chapter 2 is the comprehensive literature survey of the current research trends in MANET QoS aware Routing. The related works in the areas of Unipath and Multipath Routing, existing QoS models, existing QoS aware routing protocol which are single constrained and multi constrained are thoroughly discussed.
Chapter 3 defines the potential research problems with brief introduction and motivation. The drawbacks of the existing solutions and necessity to go for alternative solutions are also highlighted. Different parameters for the evaluation of proposed techniques are discussed along with the platform for implementation.

Chapter 4 highlights working of existing Routing Protocols for MANETs in which, important routing algorithms such as AODV, DSDV and DSR are described. Performance comparison and analysis of some of the important Routing techniques in terms of QoS metrics such as packet delivery ratio and delay are also presented.

Chapter 5 compares the existing single path and multipath routing protocols such as AODV and AOMDV. The general observations from the simulation are the application oriented performance metrics such as Average delay, Packet delivery ratio and Throughput by varying traffic and mobility scenarios. Compared to AODV the overall performance of AOMDV is better in all the aspects. A method to compute the multiple node disjoint paths using path accumulation feature of DSR in AODV is developed. So, multipath routing protocol is taken as the base protocol to improve the QoS parameters.

Chapter 6 describes a method to avoid unnecessary flooding of RREQ messages during route discovery process. In an existing AODV routing, if a source node is going to send data to the destination, the source node broadcast RREQ messages to the neighbor nodes. If a neighbor node is not a destination node, an intermediate node will store the information available in the RREQ messages to the intermediate node Routing table entry and generate a RREP message to send to the source nodes. Next, intermediate node will rebroadcast RREQ messages to the neighbor nodes until the RREQ message finds the destination node. But, in this route discovery process many unnecessary packets get rebroadcasted resulting in packet loss. This in turn increases the packet loss in the network. The
idea here is to reduce the rebroadcasting of unnecessary packets using average time stamp method.

Chapter 7 discusses a technique for computing the percentage route life time by taking life time parameter of AODV. Life Time is the expiry time for the active route. It is also known as the deletion time for an invalid route. TTL carries a time to live (TTL) value that states for how many hops this message should be forwarded. This value is set to a predefined value at the first transmission and increased at retransmissions. Retransmissions occur if no replies are received. The LTR (life time ratio) is the ratio of Life Time to the TTL. The LTR multiplied by 100 gives the PLTR (percentage lifetime ratio) for a route. If the PLTR is bigger than 50%, the intermediate node allows the rebroadcasting of RREQ messages. This improves the QoS up to some extent. This method shows performance degradation for high mobility scenarios. So another method, for computing the percentage route life time dynamically is described. The link life time is calculated at each hop during the route request packet is traversing the path. Each node calculates the life time of the link between itself and previous hop. The Route Life Time (RLT) is the minimum link life time along a routing path. Therefore, the RLT is equal to the minimum of LLTs for a route.

Chapter 8 presents a simple technique for computing the residual energy across each node, during the route discovery process. Since the nodes in the MANET use battery power, the efficient usage of the energy improves the network life time. The network lifetime is defined as the time from the beginning of simulation until first node in the MANET runs out of energy. During route discovery mechanism it is possible to determine which nodes have consumed more energy, so that the routing path with less consumed energy can be utilized for data transmission and such routes can be stored in routing table. It is possible to compute the energy value at different time intervals across every node.
Chapter 9 presents a simple technique for QoS path computation in a sample MANET. While computing the QoS paths, this method takes multiple constraints such as timestamp, route life time and the energy across the nodes. Also the route establishment time is presented along with other results of comparison and possible marginal increase in time is justified.

Chapter 10 presents a summary of the research work carried out and lists all the contributions. It also suggests the directions for further research.