CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

In an ad hoc network, there exists no a central control namely base station or switching centers. The functioning of the network is, hence, dependent on trust and cooperation between nodes. Because the communicating nodes in a MANETs use any relay point present, congestion frequently occurs in such networks. Network congestion can cause severe delay and packet loss and reduce network throughput. If appropriate congestion control is not performed, this could lead to collapse of the network. The main objective of congestion control is to minimize delay and buffer overflow caused by network congestion. This chapter reviews the available literature on congestion control and aims to propose an improved congestion control routing protocol for MANETs.

2.2 QUEUE MANAGEMENT

In queue management, algorithms are designed to handle the length of queues as well as drop packets when necessary. With respect to dropping of packets, Passive Queue Management (PQM) is that which does not drop packets until the buffer is full, and Active Queue Management (AQM) is that which drops packets probabilistically before the buffer is full.
2.2.1 Passive Queue Management

In PQM, no precautionary step is usually taken for rejecting the incoming packets until the queue reaches its maximum length. When the queue reaches its maximum length, all incoming packets are rejected until at least one packet from the queue has been transmitted. A well-known passive technique is tail drop. The tail drop scheme drops most recently arrived packets when the queue is full, giving a signal of congestion to other sources. However, two important drawbacks of a PQM technique are (i) lock-out and (ii) full queue.

- Lock-out

Lock-out occurs when the tail drop allows more flows monopolize the queue space and prevent other flows to enter the queue, resulting in unfair sharing of network resources among the flows.

- Full queues

When the tail drop queue maintains almost full status for a long period of time, high queuing delays can result, while the objective of queue management is to reduce steady-state queue size. End to End delay is important to monitor than high throughput.

Besides tail drop, two alternative queue disciplines are (i) random drop and (ii) drop front. The random drop queue discipline drops a randomly selected packet from the queue when the queue gets full (Hashem et al 1989). The drop front queue discipline (Lakshman 1996) drops the oldest packet at the front of the queue when the queue gets full. Both queue disciplines can solve only lock-out problem but not full-queue problem.
2.2.2 Active Queue Management

In contrast to PQM, AQM enhances routers to detect congestion well in advance and provide implicit feedback mechanism to senders of impending congestion. Such a feedback could alert the sender to reduce their packet transmission rate before a queue becomes full (Braden et al 1998).

The objectives of AQM are as follows:

1) to reduce the number of packets dropped in routers to eliminate global synchronization and to improve network throughput and bandwidth (Willinger et al 1995).

2) to provide low delay interactive services by maintaining the average queue size small.

3) to avoid lock-out by ensuring that there will always be room available for an incoming packet, allowing fair sharing of bandwidth among the competing flows (Floyd 1993).

The Internet Engineering Task Force (IETF) has proposed many AQM algorithms. Random Early Detection (RED) is one of the earliest AQM mechanisms and currently the most widely implemented. RED ensures to keep the queue size small, reduce burstiness and solve the problem of global synchronization in the Internet (Floyd 1995 a). However, few limitations of RED include: (i) it can be complex to configure RED to achieve predictable performance; (ii) improper configuration of RED parameters can result in worst bandwidth utilization.

In “adaptive RED”, introduced by Feng et al (1997), drop probability was modified based on the past history of average queue size. By increasing the drop probability when the average queue size approaches
“maxthresh” and decreasing the drop probability when the average queue size approaches “minthresh”, the adaptive RED algorithm tries to maintain high link utilization.

May et al (1999) have performed a critical evaluation of RED, and their findings are as follows. RED with smaller buffer would not improve significantly the performance of the network, and parameter tuning in RED had no big impact on end to end performance.

Floyd (2000 b) presented Gentle RED, which introduces a second drop probability from maximum probability (max_p) to 1, when the average queue size is greater than “maxthresh” and less than twice the “maxthreshold”. Christiansen et al (2000) evaluated RED experimentally in a laboratory scenario on web traffic with congestion only in the forward path. They concluded that RED offers no clear advantage over tail drop FIFO in terms of per-connection response time for web users. The authors also reported that performance is quite sensitive to the setting of RED parameters.

Misra et al (2000) discussed the difficulties in tuning RED parameters. Hollot et al (2001) focused on oscillations in queue size and used this starting point to recommend values for RED parameters. Kunniyur et al (2001) presented a new scheme for AQM, called Adaptive Virtual Queue (AVQ), and studied its stability in the presence of feedback delays, its ability to maintain small queue lengths and its robustness in the presence of extremely short flows. How accurately their experimental models simulate the real internet and the performance of all the algorithms on the real internet were not addressed by the study.

Li et al (2001) presented a UDP congestion control scheme for Internet real-time multimedia traffic. In this scheme, packet policing and marking using token bucket were applied at the end host while an enhanced
RED was used to differentiate the marked and unmarked packets at the routers. However, this scheme would need the support from routers.

A modification to RED, called adaptive RED, was proposed by Floyd et al (2001c). The objective of this scheme was to use target queue utilization value to adjust marking or dropping probability by using additive increase/multiplicative decrease mechanism on every predefined interval. If the queue was smaller than the target queue, the marking probability was decreased multiplicatively.

La et al (2002) demonstrated that the presence of UDP traffic has a fundamental impact on the dynamics of the congestion control system. Therefore, modeling TCP and UDP with RED is an important step towards improving the network efficiency and the service provided to Internet users. Ranjan et al (2002) proposed to construct a nonlinear dynamic model of the system that employs the average queue size captured from the queue dynamics in the RED gateway. A discrete model is derived in order to capture the interactions of the TCP and UDP congestion control algorithms with the RED queuing mechanism.

Fu et al (2003) proposed two mechanisms to improve TCP by earlier reaction to link overload: a Distributed Link RED (LRED) and an adaptive pacing strategy. Both methods used the probability for a drop based on the observation of the number of transmission attempts needed on the MAC layer. The limitation of LRED is that the packet retries are more frequent and the probability to drop a packet is also increased.

Medina et al (2007) proposed Neighborhood RED (NRED), which is an extension to the RED mechanism, with the goal of ensuring fair bandwidth allocation across flows in the networks. NRED provides improvements but suffers from limitations that can make its broad
implementation difficult. In particular, NRED relies on the careful tuning of eight parameters, most of them sensitive to topology and traffic patterns. Also, added communication overhead is required for notifying neighboring nodes of detected congestion, which leads to performance decrease in heavy traffic situations. Furthermore, the packet marking mechanism proposed in NRED can lead to additional unfairness across flows sharing common nodes.

Francini (2011) presented algorithmic steps for a new AQM scheme called Periodic Early Detection (PED). Like RED, PED derives its packet drop decisions from the monitoring of average queue length. However, PED differs radically from RED in the handling of the packet drop rate. While RED continuously adjusts the drop rate to the evolution of the queue length, PED applies drop rate adjustments only at the end of fixed time intervals and only if the current rate is clearly off range.

Congestion control in communication networks has been of high research interest in recent decades. AQM has been shown to be an efficient way to control network congestion. However, the existing AQM algorithms suffer from robustness and the need for parameter tuning. Wang et al (2012) proposed a novel scheme called Model Predictive Controller for AQM (MPAQM). The proposed algorithm is able to speed up computations and responses than existing AQM algorithms. This protocol is not considering time-domain constraints on drop probability and queue length.

2.3 ROUTING IN MANETS

Routing in a network is the process of finding routes so that nodes in a network can use these routes to communicate from a source node to the desired destination. The routing protocol specifies the rules for finding the routes. The goal of routing is to find a correct and an optimal route to the destination in an efficient manner.
The general algorithm for routing in MANETs is dependent on two key factors (i) the range of transmission of the individual MANET nodes and (ii) the threshold sensing power at the receiving node. When a mobile node moves out of the range of a transmitting node, the packets are dropped and eventually the link breaks. Similarly, when the signal power received at a node is less than the threshold power, the link breaks. Routing protocols are designed to handle both the scenarios with poise.

There are numerous different routing protocols presently proposed for ad hoc networks. Unfortunately, few have been extensively simulated, where as very few have been implemented in an actual ad hoc environment. We believe by analyzing routing schemes a wider knowledge of the problem could be developed. Moreover, it could also be used to either extend existing schemes or to develop new routing solutions.

Though a very large number of published information in the area of routing in MANETs are available, only selected papers that have close relevance to our proposals, namely reduction in control overhead in route discovery in MANETs and adapting to the node mobility are discussed below.

2.3.1 Proactive (or Table-Driven) Routing Protocol

The proactive protocols maintain routing information about each node in the network. The information is updated throughout the network periodically or when topology changes. Each node requires storing their routing information.

Two of the proactive methods are,

- Destination Sequenced Distance Vector routing (DSDV) (Perkins and Bhagwat 1994)
- Source Tree Adaptive Routing (STAR) (Aceves and Spohn 1999)

### 2.3.2 Reactive (or On-Demand) Routing Protocol

The reactive routing protocols look for the routes that are created as and when required. When a source wants to communicate to a destination, it invokes the route discovery mechanisms to find the path to the destination.

Examples of this type are

- Dynamic Source Routing (DSR) (Johnson 2007)
- Ad-Hoc On-demand Distance Vector (AODV) (Perkins et al 2007)

These classes of routing protocols are available but choosing best out of them is very difficult as one may be performing well in one type of scenario, the other may work well in other type of scenario. The design of routing protocols in MANETs must deal with a number of challenges due to the constraints and unique characteristics of MANETs. However, due to high overhead associated with the proactive routing protocols, only reactive routing protocols have been considered in this research. The rest of the section describes the main functionality of the traditional reactive routing protocol for MANET that has been widely investigated and analyzed, namely AODV.

#### 2.3.2.1 Ad hoc on demand distance vector routing protocol

The specification of AODV is available in Request for Comment (RFC) 3561 Perkins et al (2007). AODV is developed by taking several features from DSR and DSDV and combining these features into a new
protocol. The AODV protocol works similar to DSR in that a route is only built when a route is required and works similar to DSDV in that route requests contain sequence numbers. AODV also uses HELLO packets to get information about neighbour nodes. This HELLO packet is used to determine the local neighbourhood; these packets are also used to detect link breaks in the neighbourhood. The link break is detected if the current node does not receive a HELLO packet within a configurable amount of time from a previously known neighbour. In this case, that neighbour will be removed from the neighbourhood. When a route is required, the local node will broadcast a Route Request (RREQ) packet. The RREQ packet will be forwarded by all neighbours and will eventually reach every node in the network, assuming the network is connected.

A network is considered connected if a route exists between each pair of nodes in the network. When the destination node receives a first RREQ packet, a Route Reply (RREP) packet will be generated and unicast back along the path that the successful RREQ packet had taken. This reverse path is available because the RREQ packet is modified at each hop to include the previous node. This means that when the RREQ packet reaches the destination a full path back to the source is included in the RREQ packet.

The destination node generates RREP only on the reception of the first RREQ. The subsequent duplicate RREQ packets received by the destination node are simply dropped. AODV maintains the routing table based on expiration times. The larger the network the longer the expiration times must be. When a route is cached by a node the route will have an expiration time associated. If that expiration time is reached and no further packets have been relayed along that route, then that route will be deleted from the routing table.
2.4 BROADCASTING ALGORITHM IN MANETS

Broadcast is the process of sending a message from one node to all other nodes in an ad hoc network. It is a fundamental operation for communication in ad hoc networks as it allows for the update of network information and route discovery as well as other operations. This section presents a comprehensive review of existing localized solutions on broadcast, where only local knowledge is required. The following section highlights the use of broadcast in route discovery. Broadcast forms the basis of all communications in ad hoc networks. The simplest form of broadcast in an ad hoc network is referred to as blind flooding. In blind flooding, a node transmits a packet, which is received by all neighbouring nodes that are within the transmission range. Upon receiving a broadcast packet, each node determines if it has transmitted the packet before. If not, then the packet is transmitted. This process allows for a broadcast packet to be disseminated throughout the ad hoc network. Blind flooding terminates when all nodes have received and transmitted the packet being broadcast at least once. As all nodes participate in the broadcast, blind flooding suffers from the Broadcast Storm Problem (Ni et al 1999).

The broadcast storm problem states that, in a CSMA/CA network, blind flooding is extremely costly and may result in the following:

**Redundant rebroadcasts** occur when a node decides to rebroadcast a message to its neighbours; however, all neighbours have already received the message. Thus the transmission is redundant and useless.

**Medium contention** occurs when neighbouring nodes receive a broadcast message and decides to rebroadcast the message. These nodes must contend with each other for the broadcast medium.
**Packet collision:** Because of the lack of the back-off mechanism, RTS/CTS dialog, and the absence of CD, collisions are more likely to occur and result in lost or corrupted messages.

In the dynamic nature of ad hoc networks, a centralized approach is neither desirable nor feasible as the cost of obtaining global topology information and maintaining the broadcast tree is restrictive in terms of overhead. Localized approaches are distributed in nature and require only localized neighbour topology information. Compared with globalized approaches, localized approaches are able to adapt better to the topology change in ad hoc networks. Thus localized approaches are more appropriate to be applied in ad hoc network environments and are the focus of this section.

### 2.4.1 Optimized Broadcast with Fixed Radius

This section explores the optimized broadcast mechanisms that utilize a fixed transmission range and focus on the reduction of unnecessary rebroadcasts, which lead to the broadcast storm problem. Figure 2.1 shows the different methods of optimizing the broadcast to reduce broadcast storm problem.

![Figure 2.1 Types of Broadcast Optimizations](image)
2.4.1.1 Heuristic-based broadcasting

Heuristic-based broadcasting methods require careful selection of parameters and thresholds, which are closely related to ad hoc network environments. Their performance is highly dependent on the selected parameters and their thresholds in the heuristic. Ni et al (1999), Tseng et al (2003) proposed several heuristics on broadcast:

**Counter-based:** The decision of rebroadcast is based upon a threshold value for the number of duplicate packets received by the broadcasting node. If the number of duplicate packets is less than the threshold value, then the node will rebroadcast. Otherwise, it will not rebroadcast. An expected additional coverage function may be defined, which shows that the more times a host has heard the same broadcast packet, the less additional coverage the host contributes if it rebroadcasts the packet.

**Distance/location-based:** The heuristic may involve distance in a relative sense – physical distance between nodes or the transmission power required. Each node is equipped with a GPS device. Given the distance or location of broadcasting nodes, it is possible to calculate the expected additional coverage (in terms of area) a node may contribute by rebroadcast.

**Probability-based:** The decision of rebroadcast is based upon a random probability. This probability may be as simple as flipping a coin or it may be more complex involving probabilities that include parameters such as node density, duplicate packets received, battery power or a node’s participation/benevolence within the network.
2.4.1.2 Neighbour coverage based protocols

In Neighbour Coverage-Based (NCB) broadcast, nodes periodically broadcast beacon messages to advertise their own existence and also to discover the existence of neighbouring nodes within the transmission range (one hop). Beacon messages may typically contain the broadcasting node’s address and the neighbouring nodes addresses that the node may be aware of. Thus, the information of neighbour topology within two hops can be obtained. The use of neighbour information allows the link state topology of nodes to be determined. It is also useful in situations like indoor applications where GPS may not work. The exchange of beacon messages allows for attaching additional information about neighbouring nodes. The additional information may include a node’s remaining battery power, any user-based constraint, physical coordinates acquired through a GPS device, Signal-to-Noise Ratio (SNR) measurements (acquired from the MAC layer), and possible device characteristics such as maximum broadcast power.

2.4.1.3 Dominating sets–based broadcasting

Wu and Li (1999, 2001), describe a simple and efficient distributed algorithm for determining a Connected Dominating Set (CDS). CDS may be used to limit the broadcast storm problem, by limiting broadcasting nodes to those gateway nodes. A dominating set is the set that any node in the network either belongs to the set or is the direct neighbour of some node in the set.

2.4.1.4 Cluster-based broadcasting

Clustering (Gerla and Tsai 1995) is the process of grouping nodes together into clusters (groups). A representative of each cluster is called the cluster head. A cluster encompasses all nodes within a cluster head’s transmission range. Nodes that belong to a cluster, but are not the cluster
head, are called ordinary nodes. Often nodes may belong to more than one cluster. These nodes are called gateway nodes. Only cluster head nodes and gateway nodes are responsible for propagating messages.

2.5 **CONTROL OVERHEAD REDUCTION IN ROUTE DISCOVERY**

The literature pertaining to the control of routing overheads generated during route discovery is discussed in this section.

The Signal Stability Adaptive protocol (SSA) (Dibe et al 1997) tries to discover longer-lived routes based on signal strength and location stability. Each link is differentiated as strong and weak according to the average signal strength at which packets are heard. The location stability criteria further biases the protocol toward choosing a path which has existed for a longer period of time. Beacons are sent periodically by each host for its neighbours to measure these criteria. Each host maintains a signal stability table. On an active link failure, SSA broadcast a “local” RREQ packet with a small hop limit, in hope of rebuilding the route with little effort. Also, the initiator of RREQ should set up a timer, so that if the broken route cannot be rebuilt within the timeout period, it can send a normal route error packet to the source node so that a “global” RREQ packet can be sent.

Associativity-Based Routing (ABR) (Toh 1997) uses pilot signal to determine link stability. Every node sends pilot signal periodically. When a node receives this pilot signal from its neighbour nodes, it records pilot signal received. If it receives these pilot signals from a neighbour continuously and the number of continuous pilot signals beyond certain threshes hold, it considers the link between them as stable link. Route search in ABR is different from SSA. In ABR, it searches all possible routes to find a route that contains more strong links.
The Cluster-Based Routing Protocol (CBRP) for MANET was initially suggested by Jiang et al (1999). In CBRP, the network topology is divided into several disjoint overlapping clusters. Each cluster elects one node as the cluster-head. The cluster-head of each cluster is responsible for forwarding RREQ packets on behalf of its members. Cluster-heads communicate with each other by gateway nodes. A gateway is a node that has two or more cluster-heads as it neighbours.

Goff et al (2001) checked the link reliability with signal strength. Each time a link failure is detected, instead of repairing this link, they will try to find an alternative path for the one including this broken link.

There have been a few proposals for establishing a virtual back bone infrastructure over which routing takes place. The design of virtual back bone is based on Connected Dominating Set (CDS) (Wu and Li 2001, Wu 2002). A CDS is a set of nodes such that every node in the network is either in the set or is the neighbour of a node in the set. In CDS-based routing, only nodes in the dominating set are privileged to forward the RREQ packet. Undoubtedly, the efficiency of the CDS approach depends on the process of finding and maintaining a CDS and the size of the corresponding sub-network. Unfortunately, the problem of finding minimum CDS is shown to be NP-complete.

Moghim et al (2002) proposed an improvement to AODV which tries to reduce routing overhead by using source routes in route request and reply packets, so that the routing information of AODV node increases and routing tables are widely used instead of route discovery flooding. However, usage of source routes may cause problems, especially in large network with stale routes, due to loss of data packets and also pollution of routing tables in other nodes. The probability of stale route existence will become less since AODV uses sequence numbers, stale routes can be recognized and removed
from the routing entries if they don’t need to be updated. As source routes are not used in data packets, usage of stale routes will not pollute other routing tables. So the source routing problems is not encountered in AODV. However, using of source routes even in route request and reply packets can still cause the problem of scalability in huge networks.

The Gossip-based Ad Hoc Routing (Hass et al 2002) proposed a probabilistic route discovery refer to as GOSSIP3 \((p, k, m)\). The parameter \(k\) is the number of hops at which to start forwarding an RREQ packet with probability \(p\). In this protocol, a node that originally did not broadcast a received RREQ packet, but then did not get the packet from at least \(m\) other nodes within some time interval, broadcast the packet with probability \(p = 1\) immediately after the timeout. The authors have incorporated the GOSSIP algorithm in AODV routing protocol. However, in heavily loaded networks the improvement of the gossip based routing algorithm is limited.

Advanced Signal Strength based Link Stability estimation Model (ASBM) is proposed in by Lim et al (2002) by enhancing SBM. Link stability in SBM is decided only with signal strength. In ASBM, Differentiated Signal Strength is added (DSS) as a parameter. DSS indicates the signal strength is going stronger or weaker. If it becomes stronger, it means that two nodes will be closer and the link between them would have longer lifetime. In SSA, only a link, the signal strength of that exceeds certain limit, are considered as a stable link. Whereas ASBM, considers both strong signal links and weak signal links that are coming closer as stable links.

Srinath et al (2002), on the other hand, proposed to maintain the so called neighbour power list and power difference table so as to initiate backup route switching mechanism when signal strength between two adjacent nodes in an active link is weakening. However, maintaining two tables requires
periodical HELLO messages interchange, and thus imposes a considerable burden on the network.

Cartigny and Simplot (2003) proposed an algorithm which combine the advantages of both probabilistic and distance method to privilege the retransmission by nodes that are located at the radio border of the sender. The value of probability $P$ is determined by the information collected from the node’s neighbours and the constant value $K$ which are efficiency parameters to achieve high reachability.

Tseng et al (2003) have proposed an adaptive counter based scheme in which each node dynamically adjusts its threshold value $C$ based on local neighbours information. The fixed threshold $C$ is computed based on a function $C (n)$, where $n$ is the number of neighbours of the node. In this approach the value of $n$ can be achieved through periodic exchange of ‘HELLO’ packets among mobile nodes.

Li and Mohapatra (2003) proposed a Positional Attribute based Next-hop Determination Approach (PANDA) which uses positional attributes such as location and velocity information to determine the rebroadcast delay time, while aiming at finding a longer-lived route with a smaller number of hops to the destination. The basic idea of PANDA is to discriminate neighbouring nodes as good or bad candidates for the next hop on the basis of positional attributes that are of interest. These attributes can be relative distance, link lifetime estimation, and transmission power consumption. The discrimination is done at the downstream node side. Good candidates will use shorter rebroadcast delay, while bad candidates use longer delay such that the good candidates usually go first. Since good candidates usually go before bad ones, a better route in terms of metrics such as hop count, delay, power consumption, or residual battery, can be found. It is assumed that each mobile node is equipped with Global Positioning System (GPS) so that it is aware of
its geographical location and velocity information. To let the downstream
nodes learn the previous-hop node’s location and velocity information, they
assume that this information is carried with the route-request message in each
hop. Upon receipt of a route request packet, an intermediate node can
compare its own location and velocity with that of the previous-hop node and
then determine the rebroadcast delay according to the algorithm it uses,
namely, PANDA-LO (Location Only), PANDA-LV (Location and Velocity),
or PANDA-TP (Transmission Power). This complete protocol depends on
GPS, which do not go well with congested networks.

Gwalani (2003) proposed a new modified AODV, which was called
AODV with Path Accumulation (AODV-PA). AODV-PA combines the route
discovery process of AODV and DSR. When the RREQ or RREP messages
are generated or forwarded by the nodes in the network, each node appends its
own address on these messages, then the RREQ or RREP messages are sent
out again. Each node also updates its routing table with all the information
contained in the control messages. AODV-PA sets up a lot of reverse routes.
AODV-PA’s method has reduced RREQ’s number in the network and it has
lower delay and lower routing load than original AODV. However, it has to
set up too many reverse routes, this would bring in DSR’s disadvantage, and
results in limited performance.

Two independently proposed protocols (Goff et al 2003, Boukerche
and Zhang 2004) make use of link status information of routes that is
piggybacked in the RREP packets to notify the source so that it can plan for
the right timing to route reconstruction. In the meantime, nodes on the route
also need to perform link quality monitoring and confirm that with
ping-pong process for possible link breakage. As soon as a potential link
breakage is identified, the source will be warned so that it can preemptively
initiate route reconstruction procedure. Note that there always exists the possibility of false identifying link breakage which may lead to flooding.

Crisostomo et al (2004) proposed to use nodes’ position and mobility information, and apply such information to predict vulnerable links and potential backup nodes. Special hardware like GPS is required to assist the implementation of such protocol.

Zhang and Agrawal (2005 a) have described a dynamic probabilistic broadcast scheme which is a combination of the probabilistic and counter based approaches. The scheme is implemented for route discovery process using AODV as base routing protocol. The rebroadcast probability \( P \) is dynamically adjusted according to the value of the local packet counter at each mobile node. Therefore, the value of \( P \) changes when the node moves to a different neighbourhood; for example, in sparse areas, the rebroadcast probability is large compared to heavy traffic areas. To suppress the effect of using packet counter as density estimates, two constant values \( d \) and \( d1 \) are used to increment or decrement the rebroadcast probability. However, the critical question is how to determine the optimal value of the constants \( d \) and \( d1 \). The value of a packet counter does not necessarily correspond to the exact number of neighbours from the current host, since some of its neighbours might have suppressed their rebroadcasts according to their local rebroadcast probability. On the other hand, the decision to rebroadcast is made after a random delay.

Zhang and Agarwal (2005 b) have implemented an approach that uses the concept of gossip and Connected Dominating Sets (CDS). But the construction of minimal dominating set is not required. Instead of that, the authors categorized mobile hosts into four groups according to their neighbourhood information. For each group, there is a specified value of probability so that the nodes with more neighbours are given higher
probability, while the nodes with fewer neighbours are given lower probability.

Neighbour Stability Routing (NSR) (Chen and Lee 2005) algorithm selects the most historically and accumulatively stable mobile nodes to form a path between the source node and destination node. The relative stability is then propagated from the collective data by all the nodes along a path. The cumulative collective data, or stability factor, reflects the historical neighbourhood stability among neighbours. When a node or segment on the path is down, NSR will dynamically find an alternative most stable path.

Qi et al (2005) proposed a routing protocol, taking the node power as the major consideration when selecting paths. In their protocol, either link failure or a too low value in node power may trigger route maintenance. Also, they check the path quality periodically for unreliable links.

Sengul and Kravets (2006) proposed a local recovery algorithm and combined that with the DSR protocol. In case of link breakage and no alternative route in the route cache, the protocol issues a local recovery request looking for potential node for path detour. If any packet successfully reaches the destination through the newly discovered route, the destination node would then notify the source node for such a route change by returning a notification message.

Tsai et al (2006) proposed keeping track of the Signal to Noise Ratio (SNR) of links so that a new route request would be initiated once the SNR is below some pre-defined threshold to reduce the possibility of link breaking. However, mistaken in prediction cases, their method may result in additional route reestablishment. Even if the prediction is correct, local route repair may well be sufficient instead of complete new route construction.
Abdulai et al (2007) proposed a Two-P Scheme which is named as Adaptive Probabilistic Broadcasting, where the nodes are logically categorized into two groups based on their number of neighbours. A node is categorized as a member of group 1, if the number of neighbours \( n \) of this node is less than or equal to \( n_{\text{avg}} \), otherwise it is categorized as a member of group 2. Nodes in group 1 are in sparse regions of the network and as such are assigned high forwarding probability, \( p_1 \). Nodes in group 2 are located in heavy traffic regions and are assigned a low forwarding probability, \( p_2 \). To identify and categorize mobile nodes based on their local topological characteristics, they have first conducted extensive simulations to determine the average number of 1-hop neighbours of nodes in the various regions of the network. From the simulation results, they determine \( n_{\text{avg}} \) that denotes the average number of neighbours over all the nodes in the network. This \( n_{\text{avg}} \) may not be the right parameter to decide on the rebroadcast probability as the node and link capacity, network topology and other related network characteristics may vary from network to network and time to time.

Ramachandran et al (2007) proposed a new adaptive cross layer design, where an adaptive decision making of RREQ forwarding is made in accordance with speed of mobile nodes. The distance between the transmission range boundary of the transmitting node and the known maximum transmission range (db), the safe distance from transmission range boundary \( d_s \) that is calculated using the node’s speed information and minimum route life time as specified by the source node \( t_1 \), are used by the node to make a decision on forwarding the RREQ. This approach suffers with increasing delay and overhead by frequent cross layer interactions.

Yu et al (2007) proposed a dynamic route repairing protocol that repairs a broken route using information provided by nodes overhearing the main route communication. When links go down, their protocol intelligently
replaces these failed links or nodes with backup ones that are adjacent to the main route.

Shi et al (2007) related the link reliability to the physical condition of the two end nodes and provide a new method to evaluate the link reliability. Based on that, they proposed a new Link Reliability-aware Route Maintenance mechanism (LRRM). LRRM checks the link reliability for every hop and actively. If necessary, LRRM will replace those weak nodes with more robust ones, which consequently keep the link healthy and contribute in traffic balancing too. This local recovery, with only a few nodes involved, is always triggered before a path breaks in order to avoid costly path reestablishment.

Taleb et al (2007), Wang et al (2007), Yang and Huang (2008) proposed an approach to assess the stability of link by using Link Expiration Time (LET). The significance of this approach is that estimation of Link Expiration Time is that it does not require Global Positioning System (GPS) devices. The routing protocols integrate the evaluation of LET into on-demand routing algorithms, such as DSR or AODV, for discovering stable route. For every node that can obtain GPS information by itself, no periodic message exchange is needed, which can reduce much control overhead. The performance of the protocol is not appealing in high mobility scenarios, because as the speed of nodes increases, the number of path breaks increases, consequently resulting in higher control overhead.

Sakhaee et al (2008) proposed a self-adaptive and mobility-aware path selection in mobile ad-hoc networks. To be aware of a mobility of node, Doppler value is calculated based on the Doppler shift which can be obtained through the forwarding of route request packet like DSR for assessing the stability. The limitation of this protocol is that it cannot perform on high mobility scenarios.
Moh (2009) proposed a link quality aware routing protocol for MANETs that exploits the strong links by forwarding the RREQ packet with the highest Signal to Noise Ratio (SNR) among the multiple RREQ packets received during route discovery. In case there are RREQ packets within $\delta$ dB, the threshold set for SINR (where $\delta = 1$) from the highest SINR, the first-arrived one among them is chosen to cope with the dynamic behavior of SINR. Any node that has received an RREQ receives successive RREQ packets until the predetermined RREQ waiting time expires; afterwards, RREQ packets for the route discovery are ignored. Compared to the conventional protocols such as AODV, in which only the first-arrived RREQ is forwarded and the others are ignored, the proposed scheme may not have the minimum hop-count route but the one with more number of hops (links). The found route is a reliable path with high data rate because it consists of strong links, resulting in high performance as well as robust routing. The performance of the protocol is not appealing in high mobility scenarios.

Yang et al (2009) proposed a hybrid routing protocol called Improved AODV (IAODV) which integrates the two features: Multipath and Path Accumulation. It is equipped with source routing characteristics, namely the path accumulation technique, which permits nodes listening to routing messages to acquire knowledge about routes to other nodes without initiating route request discovery themselves. This method decreases the required transmissions, but increases the routing packet size. IAODV has another important feature, Multipath, to keep the alternative paths valid. When a data packet is forwarded, the node updates the timer for the used route, thus keeping the path valid, until it breaks. In this protocol, the alternative routes are not used, unless the primary route fails. In order to keep these routes alive, the time period of the route is extended. Extension of the route is considered valid without any data packet using it. Setting this parameter to a value that exceeds the expected lifetime of the path, which should keep the path valid in
order to use it when the primary path fails and to prevent the accumulation of invalid paths in the routing table is a critical issue.

Wang and Liu (2009) proposed a protocol to avoid excessive energy consumption and high computation complexity; metrics for route decision are piggybacked in the header of route packets such as RREQ and RREP in AODV protocol. The route selection metrics consist of interference level and energy level to evade areas with excessive interference and lack of energy. Due to energy level integrated into the route selection scheme, the XLR protocol has evaded areas lack of energy to transmit traffic packets. Hence, energy of the entire network can be consumed symmetrically. Therefore, the network can avoid draining off their power prematurely and the network lifetime can be prolonged effectively.

Wang et al (2009) discussed the instability which may arise when reactive routing protocols interact with the IEEE 802.11 MAC protocol. In particular, several erratic behaviours of AODV routing protocol in a congested ad hoc network are demonstrated. A cross-layer solution is proposed based on an Adaptive Bulk Trigger policy and a Dynamic Window Selection scheme. The authors proposed a solution to enhance the link-failure tolerability of reactive routing protocols and provide prioritized channel access based on routing demands.

Soliman and Otaibi (2009) proposed an algorithm to enhance the Local Repair (LR) phase of AODV by using a preemptive mechanism to detect potential link failures and to find in advance some alternative links. By eliminating the use of regular and costly control messages such as RREQ, RREP, and RERR, their protocol is able to reduce substantially the control overhead generated by the repair process. A Neighbour Activity Table (NAT) that records the information about the nodes located two-hop away is maintained to reduce the overheads. The maintenance of such a NAT depends
on nodes of an active route monitoring activities of the other active routes, which make their protocol sensitive to the number of conversations. To be more specific, in cases of fewer conversation-pairs, the NAT table may not be able to populate enough alternative links for future path detouring. This approach did not investigate effect of high-mobility nodes on the steadiness of a link over its lifetime.

Hu et al (2010) proposed an improvement of the Route Discovery Process in AODV. They suggest appending only the second node’s address on RREQ is appropriate (unlike DSR). After the source node generated RREQ messages, the second node receives them and appends its own address on these route discovery messages. So the updated RREQ packet contains the source node and the second node’s information. As these updated RREQ messages are broadcasted, every intermediate node that does not have a route to the destination node forwards these RREQ packets. When an intermediate node receives a RREQ packet, it updates the route to the source node and the second node. New entry is made in the routing table for the source node or the second node, if there is no one existed yet. If a route entry for the source node or the second node does exist, and if the hop count to the source node or the second node is less than the previously known hop count to that node, the routing table entry is updated for that node. The major problem here is, route tables are too big and include too many routes, when the ad hoc network’s topology is used RREQ and RREP messages to update route table.

Yassein et al (2010) presents the Smart Probabilistic Broadcasting (SPB) as a new probabilistic method to improve the performance of existing on-demand routing protocols by reducing the RREQ overhead during the rout discovery operation. When hearing a broadcast RREQ packet at node X for the first time, the node compares its neighbour by average of minimum neighbours (avg$_{min}$), average neighbours (avg) and average of maximum
neighbours \( \text{avg}_{\text{max}} \): if the node has a number of neighbours \( n \) less than \( \text{avg}_{\text{min}} \), this implies that the node is in a high sparse region. Then, the node rebroadcasts the packet according to probability \( p_1 \). However, the probability \( p_2 \) is selected if the number of neighbours \( n \) satisfies \( \text{avg}_{\text{max}} \leq n < \text{avg} \): this implies that the node is in a medium sparse region. Otherwise, the value of probability \( p_3 \) is chosen if the node is in a medium density region and the number of neighbours \( n \) meets the following this condition: \( \text{avg} \leq n < \text{avg}_{\text{max}} \). Finally, the value of probability \( p_4 \) is chosen if the number of neighbours \( n \) is \( n \geq \text{avg}_{\text{max}} \). This implies that the node is in a high density region. Even though the proposed method arrives at a smart probabilistic value to rebroadcast, it involves few time consuming calculation and require a global knowledge about the network. This improvement applied over the pure broadcasting, performs better in terms of reachability and saved rebroadcast, it can be further enhanced to work relatively better in congested networks.

Cheng (2010) proposed integrating the Dynamic Link Break Avoidance (DLBA) and Dynamic Path Shortening (DPS) mechanisms into a modified protocol by developing a pair of parameters to determine the timing for activating DLBA or DPS so that the two algorithms can work cooperatively and complementarily together. The DLBA mechanism operates by monitoring the quality of communication links and inserts an intermediate node in between the two end nodes of some breaking link. The DPS mechanism operates by recognizing potential redundant intermediate nodes and either removes or replaces the nodes from the established path. There are two types of path shortening scenarios. In type I scenario, some redundant nodes are directly removed from the active route and in type II scenario, some redundant nodes in the active route is replaced with a node that is not on the active route. This approach did not consider the node mobility effects and packet transmitting speed in determining the threshold values.
Hu et al (2011) introduced a routing algorithm which enhances the stability and the continuity of communications in MANETs. Communication stability is ensured by choosing the most stable route which bases on the computation of the Link Expiration Time (LET). The route with the longest LET is considered as the most stable. Then the reactive calculation of LET, asynchronous mobility information and LET update, and alternative route pre-discovery based on the Critical LET (CL) zone are proposed to further enhance the adaptability of stability-oriented routing to the dynamics of network and ensure the continuity of communications. This method only deals link stability mechanism. The problem of link stability is periodic exchange stability message, which increases consumption of network resources and increase the opportunity for congestion.

2.6 CONGESTION CONTROL PROTOCOLS

2.6.1 Dealing with Route Failures

In a typical MANETs, route failures occur frequently. The amount of time required for discovering a new route is quite large in standard TCP congestion control mechanisms. For a long time, no packets would be delivered and no acknowledgment may be received, causing the TCP sender to reduce its window size dramatically, even though, in fact, no real congestion situation might exist.

The first approach to deal with congestion control problem in MANET is called TCP-Feedback (TCP-F), proposed by Chandran et al (1998). There are situations where severe performance degradation is likely when wrong route failure notifications are sent and unnecessary route discoveries are performed. This observation is quite fundamental and should generally be taken into account in MANETs.
Liu et al (2001b) presented a solution Ad hoc Transmission Control Protocol (ATCP) to several TCP problems in MANETs. ATCP not only handles route failures but is also intended to handle longer periods of disconnection and to distinguish congestion-related losses from other losses. For the latter, Explicit Congestion Notification (ECN) messages are used. The ATCP had two limitations: (i) it heavily depends on ECN messages to recognize a congested network; (ii) when ATCP is in “loss” state; all unacknowledged packets in TCP’s buffers are retransmitted unnecessarily.

Wang and Zhang (2002) proposed a pure end-to-end mechanism, called TCP with Detection of Out-of-Order and Response (TCPDOOR). This method consists of two mechanisms (congestion control and instant recovery). By experimental results, the authors have found that the combined methods did not bring better performance. To achieve better performance, the author recommended that temporarily disable congestion control and perform instant recovery.

Anantharaman et al (2004) performed an analysis of TCP behavior in MANETs. They concluded that the schemes proposed are only able to deal with TCP’s problems like breaking links and route failures.

Klemm et al (2005) proposed an approach for anticipating route failures, based on signal strength-based link management. They point out that the 802.11 MAC cannot identify link breaks correctly.

2.6.2 Coping with Wireless Losses

A wireless link is much more prone to random packet losses than a wire-line connection. Such losses are detrimental to a transport protocol’s performance, as well as are misinterpreted as congestion-induced packet drops.
Gunes and Vlahovic (2002) introduced three states in TCP senders. Their approach was called TCP with Restricted Congestion Window Enlargement (TCP/RCWE) and based on the Explicit Link Failure Notification (ELFN) mechanism. A basic limitation is that it was compared only to standard TCP without ELFN.

Fu et al (2002) proposed a TCP-friendly transport protocol namely Ad hoc Transmission Control Protocol (ADTCP) for ad hoc wireless networks. It uses two metrics, inter-packet delay difference and short-term throughput, to detect network congestion. These two metrics are combined to gain a more robust congestion indicator. A problem with ADTCP is that it tends to keep a large window despite frequent routing changes and link losses, thus mostly underperforms than ordinary TCP.

Chen et al (2003 a) proposed a scheme where the TCP sender could limit the size of its maximum transmission window to the Bandwidth-Delay Product (BDP) of the path in multi-hop networks. However, the small transmission window leads to high Additive-Increase/Multiplicative-Decrease (AIMD) costs, and the TCP source is also unable to detect packet loss as it does not receive sufficient duplicate ACKs, resulting in significant throughput deterioration. In contrast with non-TCP variants, a majority of the proposals are modified versions of the legacy TCP protocol.

The Ad hoc Transport Protocol (ATP) proposed by Sundaresan et al (2005) is a rate-based, network-supported transport protocol for MANETs with end-to-end congestion control. The authors considered TCP’s mechanisms as inappropriate for ad hoc networks. Thus, ATP was designed as an “antithesis” to TCP. ATP strictly separates congestion control from reliability mechanisms and requires only limited feedback from the receiver. The intermediate nodes piggyback the maximum queuing delay along the
route on the packets passing by. This information is then used to determine the appropriate rate at the source node.

DeOliveira et al (2005) proposed a dynamic adaptive acknowledgment strategy for TCP over multihop wireless networks. It used the measured roundtrip times to distinguish between congestion and medium losses. They examined the performance of their approach with background traffic and found better results in detecting loss accurately. But if the traffic was highly dynamic, the performance of this algorithm was too slow.

Elrakabawy (2005) proposed a strategy to proactively detect the incipient congestion by relying on some measured metrics at the source, such as the variation of measured round trip times and short-term throughput. The proposed protocol is not able to deal link failures and route failures.

Chen et al (2006) presented a model for the joint design of congestion control, routing and scheduling for ad hoc wireless networks by extending the framework of network utility maximization and applying dual-based decompositions. They proposed a sub-gradient algorithm that is not only distributed spatially but more interestingly decomposes the system problem vertically into three protocol layers, where congestion control, routing and scheduling jointly solve the network utility maximization problem. This provides a general technique to carry out optimization-based network designs in a time-varying environment. This algorithm did not consider node mobility-related issues. Scheduling is always challenging for ad hoc networks; further, optimization is required to enhance the performance gain from cross-layer design involving link layers.

multi-hop wireless networks. The proposed scheme attempts to determine the actual cause of MAC-layer packet loss and then coordinates the congestion-control response among the MAC, network, and transport protocols. The congestion control efforts are invoked at all intermediate and source nodes along the upstream paths directed from the wireless link experiencing the congestion-induced packet drop.

Li et al (2008) proposed a novel effective datagram-oriented end-to-end reliable transport protocol, called Datagram Transport Protocol (DTPA), for ad hoc networks. They used two techniques, namely fixed-window-based flow control and a bit-vector-based Selective Acknowledgment (SACK) strategy, with which the ACK packets contain a vector of bits representing the reception status of the set of packets that were transmitted earlier. As a result, the DTPA protocol can proactively avoid generating heavy congestion. This DTPA, however, does not address issues related to route breakages, and it also experiences severe unfairness among competing flows in ad hoc networks.

Scheuermann et al (2008) proposed a novel way of accomplishing congestion control in wireless multi-hop networks: implicit hop-by-hop congestion control. It is based on the insight that an input rate exceeding the optimal output rate of a node or network area even on a short term will be detrimental to the performance of a wireless multi-hop network. This mechanism exploits the wireless broadcast medium in order to gain the necessary information for a backpressure mechanism that reliability limits the number of packets to one per flow and hop, and thereby implicitly avoids network congestion. This proposed work only dealt congestion problem, but not addressed mobility related problems.
2.6.3 Managing a Shared Medium

Cordeiro et al (2003) proposed an approach called Contention-based Path Selection (COPAS), which focuses on the capture problem of TCP in MANETs. The COPAS constructs two disjoint routes, which is based on congestion measurements performed during the discovery process. A basic demerit of this concept is that it perfectly opposes the Symmetric Route Pinning technique in the Atra Framework.

Ye et al (2004) proposed two methods, CCAR (Centralized Congestion Aware Routing) and DCAR (Distributed Congestion Aware Routing), to extend the routing in order to separate flows. In CCAR, every node has total knowledge about source, destination, and route of every single TCP flow. In DCAR, every node locally calculates a congestion weight representing its local load situation and broadcasts it to its neighbours. The problem with CCAR is overhead for broadcasting and is potentially outdated. The problem with DCAR is failure in short path.

Zhai et al (2004 a) proposed four mechanisms to reduce the impact of inter-flow and intra-flow contention on throughput and fairness in MANETs. Their approach was called Optimum Packet scheduling for Each Traffic flow (OPET). The idea of OPET appears very promising. However, the backward pressure scheme might have some problems in networks in that it requires a potentially long period of silence for the next hop to allow continuing with the transmission. This scheme might prolong the time until broken links are detected.

Papanastasiou and Khaoua (2004) proposed a scheme called Slow Congestion Avoidance (SCA). It is intended to reduce the number of packets in the network without putting hard constraints on the maximum window size, like in CWL. Similar to the ideas realized in SCA, Nahm et al (2005)
proposed to reduce the rate of congestion window growth of TCP. They call their scheme Fractional Window increment (FeW). A problem in both approaches is that it is not yet clear during short connections how much of data might suffer from the slower congestion window growth and the resulting slower convergence.

The performance of TCP in wireless networks is further diminished in IEEE 802.11-based ad hoc networks due to the unnecessary actions of on-demand routing protocols when interference-induced packet drops happen at the MAC layer. Kliazovich and Granelli (2006) addressed the problem of performance degradation of transport layer protocols due to congestion in wireless multi-hop LANs. Following the analysis of available solutions to this problem, a Cross-Layer Congestion Avoidance scheme (C3TCP) was presented, which has been able to obtain higher performance by gathering capacity information such as bandwidth and delay at the link layer. The method requires the introduction of an additional module within the protocol stack of the mobile node, which has been able to adjust the outgoing data-stream based on capacity measurements.

Yu et al (2008) proposed a joint design of congestion and contention control for wireless ad hoc networks. It provided a decoupled and dual-decomposable convex formulation, based on which sub-gradient-based cross-layer algorithms were derived to solve the dual problem in a distributed fashion for non-logarithmic utilities. These algorithms decompose vertically in two layers, the network layer where sources adjust their end-to-end rates, and the MAC layer where links update persistence probabilities. These two layers interact and are coordinated through link prices.

In the recent years, bandwidth and traffic control have emerged as issues of great importance in ad hoc wireless networks, requiring sophisticated managing techniques. Antonopoulos et al (2010) proposed a
novel, lightweight and efficient cross-layer architecture for congestion control in wireless ad hoc networks. This design approach provides critical features that suit the characteristics of ad hoc wireless networks. However, the architecture did not consider QoS demands of different applications or data flows.

Congestion control in wireless mesh networks is, if anything, harder than in wired networks. Rangwala et al (2011) implemented fair and efficient rate control for mesh networks that yielded nearly optimal throughputs. This work has included short-lived flows and mobility.

2.6.4 Alternative Protocol Designs

A variety of wireless peculiarities have been shown to be detrimental to TCP’s end-to-end way of performing congestion control in mobile ad hoc networks. Consequently, several researchers have attempted to develop reliable transmission protocols that are specifically tailored to cope with the characteristics of MANETs. Although these approaches have reported to have obtained a broad range of improvements, this necessarily comes at the cost of TCP compatibility. Moreover, most approaches are also limited to “clean” environments where no other transport protocols are used.

An approach called Application-Controlled Transport Protocol for MANET was proposed by Liu and Singh (1999 a). It was based on the observation that TCP’s throughput in MANETs is very low, while UDP achieves reasonable throughput, but suffers from a high packet loss rate. Therefore, Application-Controlled Transport Protocol’s approach alone cannot probably be sufficient to save a MANET from severe congestion problems.
Floyd et al (2000 b) noticed that TCP is not suited for transporting delay-sensitive data such as voice and video traffic. With the increase in the demand for such multimedia applications, transporting this type of data effectively becomes an important issue. The UDP protocol provides a viable alternative and has been widely adopted as the default transport protocol by many real-time applications and protocols, such as RTP. Although UDP does not perform congestion control, researchers have advocated that applications or protocols that use UDP should also implement some form of congestion control.

Load-aware Routing (LWR) is another scheme based on balancing the load throughout the network proposed by Yi et al (2001). The main idea of LWR is to drop the redundant RREQ in the intermediate node based on some local information like node queue size and channel utilization. It reduces the effects of the broadcast storm problem but it uses a predefined threshold to compute utilization, which cannot always give the right topology of the network. A Dynamic Load-Aware Routing protocol (DLAR) was proposed by Lee et al (2001). In this approach, a node with low routing load is favored to be included in the routing path during the route discovery phase. It still has a large overhead because of the request flooding problem in the route discovery phase.

The first representative of the alternative protocol category is the Explicit Rate-based Flow Control Framework in MANET (EXACT) by Chen et al (2004 b). The EXACT may be used reliably (TCP-EXACT) or unreliably (UDP-EXACT). It is rate-based and is supported by the intermediate nodes. All the nodes determine their current bandwidth to their neighbours and calculate local fair bandwidth shares for all flows. The limitations on EXACT might be imposed by the fact that it requires explicit state information for each flow in each intermediate node.
Lu et al (2003) found that AODV is ineffective under stressful network traffic situations. They proposed a modified version of AODV called Congestion-Aware Distance Vector (CADV), which favors nodes with short queuing delays in adding into the route to the destination. The problem of CADV is that it is not congestion-adaptive. It offers no remedy when an existing route becomes heavily congested.

A transport protocol for MANETs to minimize the number of required packet retransmissions and congestion control was designed by Anastasi et al (2005). It uses a window mechanism with a tightly limited maximum window size. A few limitations are, if the network tried highly dynamic scenarios, it would become more complex; additionally, for time-critical applications, the higher latency introduced by the protocol might be a problem.

Different solutions have been proposed to provide reliable packet delivery in MANETs. Most of the existing proposals in MANETs focus on the modification of congestion control algorithms in the transport protocol. These proposals may be divided into two categories: non-TCP variants and TCP variants. Zhai et al (2005) and Sundaresan et al (2005) proposed a few non-TCP variants, where in the source adjusts the transmission rate based on explicit feedback from intermediate nodes along the path. Although relatively accurate congestion information can be obtained, the algorithms in these schemes cannot retain the end-to-end semantics of a transport protocol. Moreover, they require complicated mathematical computation and incur excessive network overhead.

In the Congestion Adaptive Routing (CAR) proposed by Tran et al (2006), each node that appears on a route will alert the previous node whenever congestion is likely to happen. Once this happens, the previous node uses a bypass route to take a diversion in order to avoid the potential
congestion to the first non-congested node on the route. This still has a limitation because of PQM to monitor congestion. Therefore, CAR’s PQM approach cannot probably be sufficient to estimate network’s dynamic congestion.

Philip et al (2006) determined a rate-distortion optimized streaming of packetized media, which would minimize decoded video distortion at the receiver. UDP is a commonly used transport layer protocol for multimedia streaming applications. It does not employ any flow control scheme in response to network congestion, and therefore it can be a burden for other users on the network and, ultimately, lower its service quality. To overcome this limitation, a Real-time Transport Protocol (RTP) and Real-Time Control Protocol (RTCP) can be adopted on the top of UDP in multimedia streaming applications. That is, the RTP/RTCP layer supplements the functions of UDP by correcting out-of-order data and controlling the volume of data transmitted by senders for congestion control. However, these actions rely on periodic reports between the sender’s and receiver’s RTCP.

In streaming applications, timely delivery is more critical than reliability. This is why streaming applications typically employ unreliable UDP as transport protocol instead of reliable TCP, and TCP-Friendly Rate Control (TFRC) is implemented on top of UDP on the application layer. An interesting alternative to UDP is Datagram Congestion Control Protocol (DCCP). DCCP is a transport protocol that offers congestion control and many other useful features for streaming applications, including partial checksums (Kohler et al 2006). DCCP has not gained wide popularity yet, and it is not supported by most of the currently available IP protocol stack implementations.

A web service-based middleware for mobile applications is a promising platform to accelerate the application development for mobile
systems. Currently, the web service protocol stack is using by default the Simple Object Access Protocol (SOAP) on top of HTTP. Since HTTP is using TCP, the round-trip time of a web service call in mobile networks is high due to the connection establishment of TCP (three-way handshake) and the congestion and error control mechanisms (slow-start phase). Some applications do not require delivery guarantees of TCP. For these applications, an unreliable transport of SOAP messages over the UDP is a natural choice. Gehlen et al (2006) presented the architecture and the realization of a reliable UDP SOAP binding for a mobile web service-based middleware. The reliable UDP binding performs approximately four times faster than the default HTTP binding.

Using audio streaming in MANETs is the biggest challenge that significantly affects the throughput, end-to-end delay, and jitter of audio streams in MANETs. Daniel et al (2008) presented a novel Audio Rate Cognition (ARC) protocol that uses Artificial Neural Networks (ANNs) to adapt the audio transmission rate to changing conditions in a MANET. The ARC protocol equipped with the ANN topology appears to be the most accurate for high-quality audio transmissions. The ARC does not include larger screening experiments and the use of nonlinear modeling as well as analysis of fairness for competing ARC flows.

The transmission of real-time traffic without any control mechanism, such as UDP, can easily cause congestion in a network and also result in unfairness and throughput degradation for TCP traffic. Although a number of works had focused on congestion control in unreliable transport protocol to enable the real-time traffic friendly coexisting with the TCP traffic in a network, most of them still have shortcomings, including security problem and needing support from routers, and a few of them are designed for high-speed networks. Xu et al (2010) proposed a new transport protocol,
FAST DCCP, using congestion control for real-time traffic in high-speed networks. This protocol can satisfy the real-time requirements of multimedia data, but also achieves a high throughput in high-speed networks as well as a good fairness performance. However, this protocol did not study the effectiveness of usage in different multimedia applications.

The basic DCCP is seen as a candidate for transport layer protocol for multimedia networks. Shekhar et al (2010) modified an existing wired network implementation of DCCP in order to enable it to function in wireless MANETs. This modified DCCP can be suited to multimedia applications in MANETs. The DCCP, however, cannot perform TCP-based traffic and also cannot address causes of congestion in wireless scenarios.

Luo et al (2010) proposed an Adaptive Rate Datagram Protocol (ARDP) for transporting a real-time and short message data in multi-hop Ad-hoc networks, which employs a congestion control scheme for UDP. The significance of ARDP is that it uses ACK to measure RTT. Unlike TCP, ARDP does not carry out a retransmission; therefore, ACK is used only for congestion control mechanism. If there are more than three packets dropped for every 10 packets, the congestion control will be triggered to slow down the current transmission rate. The ARDP does not interact with upper layer’s application in multi-hop ad hoc networks.

The DCCP shows much promise of becoming a protocol of choice for real-time applications. However, only a small number of academic papers claim to its performance and its various nuances. Wilson et al (2011) described the effects of queuing, and in particular queue sizes, has on DCCP when Congestion Control Identifier 3(CCID3) is selected as the congestion control mechanism. It was found that employing small fixed-sized queues on the network led to lower packet latencies but higher volumes of packet loss as a result of the queue size reaching its maximum threshold more frequently.
Alternatively, when large queue sizes were used, the number of packet loss events reduced significantly; however, packet latency values increased. The authors suggested that new contributions are added to the various DCCP standards.

The use of UDP protocols for real-time applications over the Internet has been increasing. Nonetheless, it is highly important to have an accurate model for UDP streams since the UDP protocol does not react to network congestion. Larissa et al (2012) introduced a simple model composed of four states that accurately reproduces the patterns existing in real UDP flows. The approach is used to characterize the marginal distributions of UDP flows so that the distributions and range of parameters used in the model can be defined. The proposed model is quite accurate and its use is recommended to generate synthetic data in packet-level simulation experiments. This model did not address large-scale network-related issues.

Baboo et al (2012) presented a genetic algorithmic approach to the congestion-aware routing problem in MANETs. The proposed congestion-aware routing fitness function algorithm is capable of curing all the infeasible chromosomes with an adaptive repair function. This gives an improved quality of solution and enhanced rate of convergence. This algorithm addressed congestion-related issues but it could not discuss mobility problems.

Kuo et al (2013) proposed Cross-layer Overlay for Multimedia Environment on multimedia streaming on wireless Peer-to-Peer ad hoc network (COME-P2P). This integrates the routing protocol with P2P protocol for adapting real-time service to the dynamic wireless network. The proposed scheme can provide a stable routing path for high data rate real-time video service. This protocol has good performance on playback continuity for live and high quality streaming in slow moving speed. However, when the moving
speed is fast, this protocol cannot work well because Distributed Hash Table (DHT) changes too frequently.

2.7 CONCLUSION

Congestion control in MANET is considered as a transport layer issue. To control congestion, different approaches have been proposed over the years. In this chapter, several types of TCP-based congestion control mechanisms are discussed and the related works are highlighted.

Several issues in MANETs, including channel error, very frequent route disconnections, power and bandwidth constraints, and congestion, are the main causes for high packet loss and long delay.

Understanding these aspects in more detail might help to decide which direction future research should put its main focus on. Especially in the field of non-TCP compatible, network-supported protocols, there is still a lot of potential for the development of new protocol designs.

The basic problem with UDP is that it does not include congestion control mechanism and UDP connections have no way of detecting or reacting to network congestion. It is necessary to find a way of using UDP protocol to resolve congestion along the route and improve the overall network performance and to reduce the drop rate and delay. This work has the aim to propose ways to detect congestion well in advance at the MAC layer and take necessary action against congestion at the network layer. The following chapters discuss about the new proposal.