

A REVIEW AND COMPARISON OF QUALITY OF SERVICE ROUTING IN WIRELESS AD HOC NETWORKS

Sunita Prasad¹ and Zaheeruddin²

¹Centre for Development of Advanced Computing, NOIDA, India
sunitaprasad@cdacnoida.in

²Department of Electrical Engineering, Jamia Millia Islamia, India
zaheer_2k@hotmail.com

ABSTRACT

Quality of Service (QoS) guarantees must be supported in a network that intends to carry real time and multimedia traffic. IETF RFC 2386 defines QoS as a set of service requirements to be met by the network while transporting a packet stream from source to the destination. The dynamic network topology and wireless bandwidth sharing makes QoS provisioning far more challenging in wireless networks as compared to the wired counterparts. The support for the QoS services is underpinned by QoS routing. A QoS routing protocol selects network routes with sufficient resources for the satisfaction of the requested QoS parameters. The goal of QoS routing is to satisfy the QoS requirements for each admitted connection, while achieving global efficiency in resource utilization. The problem of QoS routing with multiple additive constraints is known to be NP-hard. This requires the QoS dynamics to be fully understood before it can be implemented in wireless ad hoc networks. The paper discusses the issues involved in QoS routing and presents an overview and comparison of some existing QoS routing protocols. The article concludes with some open issues for further investigation.

KEYWORDS

Wireless ad Hoc Networks (WANET), Quality of Service (QoS), QoS Routing, Multi-Constrained Path (MCP) problem.

1. INTRODUCTION

A Wireless Ad hoc Network (WANET) consists of a collection of mobile nodes connected by wireless links which can be created on-the-fly without using any infrastructure or administrative support [1]. These networks are characterized by *self-organization and autonomy*. The wireless ad hoc network has been attracting increasing attention of researchers owing to its good performance and special application scenarios. For example, it can be used in military operations for fast deployment of troops in hostile and unknown environments, search and rescue operations for communication in areas having little or no wireless infrastructure support, disaster relief operation where the existing infrastructure is destroyed or left inoperable and commercial application like enabling communication in exhibitions, conferences and large gatherings. Examples of wireless ad hoc networks are Zigbee and Bluetooth networks. The perception that wireless ad hoc networks are simply a wired network with cables replaced with antennas is a common misconception. The unique characteristic of wireless ad hoc networks like dynamic topology and resource constraint distinguishes it from wired networks and necessitates the need of special solutions in these networks [2].

With the advancement of technology, the wireless and portable computers and devices are becoming more powerful and capable. There is a growing desire for the wireless networks to support real time multimedia applications. Such applications require the network to provide guarantees on the Quality of Service (QoS). QoS is the ability of the network to provide some

level of assurance for consistent network data delivery [3]. The network is expected to guarantee a set of measurable prespecified service attributes to the users in terms of delay, delay variance (jitter), bandwidth, probability of packet loss, etc. Power consumption and service coverage area are the two other QoS attributes that are more specific to WANETs [14]. The goal of QoS is to provide some level of predictability and control in the network behavior and at the same time achieve global efficiency in resource utilization.

However, the current WANET architecture is not adequate for the support of more demanding applications. With Internet as the basic model, ad hoc networks have been initially considered only for best-effort services. QoS provisioning in wired networks is based on two approaches: (1) Overprovisioning and (2) Network Traffic Engineering. The first method consists of offering huge amount of resources to accommodate all the demanding applications. This method is difficult to implement in wired networks and completely infeasible in wireless networks where there is a scarcity of resources. The second method classifies the ongoing traffic in the network and processes them according to some set of rules with the joint goals of good user performance and efficient use of network resources. These solutions cannot be directly applied to wireless ad hoc networks because of their unique characteristics. Thus, in the last few years, QoS for ad hoc networks has emerged as an active area of research [8][13][16][17][18][21].

The support for QoS services in the network depends to a large extent on QoS routing. QoS routing is a routing mechanism under which paths that satisfy the QoS constraints are determined based on the knowledge of resource availability in the network and the QoS requirements of the flows. The objective of QoS routing is to find a feasible path between a source–destination pair, if one exists, that optimizes the use of network resources and satisfies the required QoS guarantees. It is evident that QoS routing is a constrained combinatorial optimization problem [12]. [4] presents an exhaustive review on QoS routing methodologies for ad hoc networks. This paper presents an overview and comparison of some existing QoS routing protocols. We also discuss the future trends in providing QoS guarantees in the network. The paper is organized as follows- Section 2 reviews the QoS routing and challenges in providing QoS guarantees in wireless ad hoc networks. Section 3 presents the QoS metrics. Section 4 analyzes some of the QoS aware routing protocols. Finally the last section concludes the paper and gives some future direction.

2. QoS ROUTING

The support of QoS services is underpinned by QoS routing. Many routing protocols for wireless networks such as AODV [5] and DSR [6] use best effort routing where all nodes within the range compete for the shared channel. No guarantees or predictions can be given here on when a node is allowed to send. In contrast, QoS routing is a mechanism under which the feasible paths provide QoS guarantees. The goal of QoS routing is to identify paths that have sufficient resources to satisfy a set of constraints and at the same time achieve global efficiency in resource utilization. The path selection in a wireless ad hoc network must be realized in an automatic and distributed way.

QoS routing consists of two parts – routing algorithm and routing protocol [7]. The routing algorithm is used to compute appropriate feasible paths while the routing protocol consists of all actions that inform the individual nodes with a consistent and updated view of the network. The routing protocol must typically define the information that is helpful in taking routing decisions and the way the information is communicated to the nodes and encoded in the routing table. The behavior of the routing protocol drives the network dynamics and critically affects the performance of the routing algorithm. The routing table is the local database of routing information and defines for each destination and for each path, the cost associated on the selection of the specific node as the next hop to forward data to the desired destination. The routing algorithm makes use of this information to actually select the paths and forward data along them.

The routing protocol can follow two different design approaches – bottom-up and top-down [8]. Many classical routing protocols follow a top-down approach. In a typical top-down design, a centralized algorithm is implemented in a distributed system. The centralized algorithm calculates the shortest path to the destination based on the knowledge of the global state of the network. Thus, the modification of the centralized algorithm has to take into consideration the intrinsic limitations of the distributed architecture in terms of limited state knowledge and delays in the propagation of the information. As a result of these modifications, several properties of the centralized algorithm may be rendered unsuitable for the wireless network. This approach requires asserting some general formal properties of the system.

In bottom-up approach, there are two main steps (1) A well defined protocol specification of individual nodes for interaction with one another and with the environment, and (2) The evolution of global behavior of the network as a result of these local level interactions. The bottom-up design approach is generally more flexible, scalable and capable of adapting to variety of situations. The negative aspect is that it is hard to state the formal properties and the expected behavior of the system.

Routing Protocol: The routing protocol presents correct and updated information of the topology and states of the link. The information about the state can be at three levels – local, global and aggregated [9]. In the local state representation each node maintains an up-to-date information of the node state and the link state. Distance vector protocols make use of the local state representation. In global state representation each node maintains a complete topological database of the network. The link state protocols exploit this strategy. The totality of the local state information for all nodes constitutes the global state information and is constructed by exchanging the local state information for every node among all the network nodes at appropriate intervals. The global state is just an approximation of the actual state. The distribution of the state information can highly increase the communication overhead. The global state algorithms converge faster but require more CPU power and memory than the local state algorithms. To reduce the protocol overhead associated with the frequent distribution of update, aggregated global state is proposed. It is obtained by first partitioning the network into hierarchical clusters and then aggregating the information. Such aggregation represents partially true global state [4] but scales better.

Routing Algorithm: The routing algorithm is static and is used to make a routing decision to find a feasible path based on the collected information. The classification of the routing algorithms can be based on many criteria.

Based on behavior:

Reactive protocols gather routing information only in response to an event which may be due to the requirement of a new route or due to the failure of an existing route. Examples are ad hoc on demand distance vector (AODV) [5] routing protocol, and dynamic source routing (DSR) [6]. Using this approach, better route can be computed as the knowledge of most recent link characteristics and the exact QoS requirements is available. Path route caching is also not required. The negative aspect is that overhead is high due to per-request processing.

In proactive protocols, the information is constantly gathered so that it is readily available when it is required. Examples are destination sequenced distance vector (DSDV) [10] routing and wireless routing protocol (WRP). The advantage of this approach is that the proactive gathering of routing information can be used to build statistical estimates of the relevant aspects of the network dynamics that can be used to learn and adapt with continuity the local routing policies. On the other hand, it is infeasible to build sound estimates when using a purely reactive strategy since there is no continuity of information gathering.

Both of above approaches focus on finding the shortest path between the source and the destination by considering the node status and network configuration when a route is desired. Constraint based routing approach use metrics other than the shortest path to find a suitable and

feasible route. Examples are Associativity-based routing (ABR) [11] which takes into account the node's signal strength to choose the path.

Based on Selection Rule:

The deterministic algorithms use a deterministic selection rule applied to the information contained in the routing table to decide the next hop. Usually it relies on the greedy selection of best routing alternative. The probabilistic algorithms make use of the probabilistic selection rule. It results in suboptimal choices but spreads the traffic across different concurrent paths resulting in load balancing. The probabilistic scheme requires more computational power and memory resources to process each packet and maintain the necessary routing information. The advantage is that due to a certain level of randomness in the selection rule, it adds robustness and flexibility to the routing system to better cope up with the network variability.

Based on routing decision:

In source routing, a feasible path is locally computed at the source node using the global information stored at the node. The path information is then disseminated to other nodes. In this, loop free path is guaranteed. However, global state information is required to be maintained at each node. In distributed or hop-by-hop routing, the source as well as other nodes is involved in path computation by identifying the adjacent nodes to which the packet must be forwarded. In this, the inconsistency in the routing tables may create loops. Hierarchical routing uses the partial global information to determine a feasible path using source routing.

2.1. Challenges in QoS Routing

QoS routing needs to take into account both the applications requirements and the availability of network resources. As a result, QoS routing in ad hoc networks impose great challenges.

1. **Dynamic topology:** The topology of wireless ad hoc networks may vary with time. This leads to imprecise network state information at the nodes and thus makes it difficult to provide QoS guarantees. In fact, when the network topology changes at a fast rate, it would not be possible to provide any QoS guarantees.
2. **Bandwidth sharing:** The bandwidth in a wireless network is limited and is also shared by other nodes in the network. The transmission from a node not only consumes local resources but also consumes the bandwidth of the neighbors within the contention range. Thus resource allocation for QoS is difficult.
3. **Resource constraint:** Mobile nodes have limited power supply. This limited power supply should be used in a manner that prolongs the lifetime of the battery. If the battery power is used blindly, mobile nodes will fail quickly which will affect the network availability and functionality.
4. **Lack of centralized control:** The absence of any centralized control demands the routing protocols to be self-creating and self-organizing. Further, the protocols must also be distributed in nature.

The QoS flows require certain resources to be reserved. The imprecise network state information and lack of centralized control makes it difficult to provide Hard QoS. In other words, QoS requirements are not guaranteed to be met for the entire session. The focus of the entire research community is to provide Soft QoS [13] or better than best-effort service in wireless ad hoc networks. Soft QoS, implies that the failure to meet the QoS requirement is allowed. It means that after connection setup, there may exist transient periods of time when the QoS specification is not honored. However, a minimum level of QoS satisfaction must be guaranteed by the network. The QoS satisfaction is quantified as the ratio of total disruption time to the total connection time. This ratio should not be higher than a threshold [14]. The actual QoS for the session may vary between a range specified by a minimum and maximum. The mobility of the nodes adds another dimension of difficulty. If the topology changes too frequently, then providing even Soft QoS is not possible. Therefore, combinatorial stability [12]

must be met to provide QoS guarantees. It means that given a specific time window, topology changes occur sufficiently slowly to allow successful propagation of topology updates as necessary.

3. QoS METRICS

QoS constraints are specified in terms of end-to-end performance, such as delay, bandwidth, probability of packet loss, delay jitter, etc. Although loss probability, cost, and delay jitter are useful QoS metrics, delay and bandwidth are the two most important QoS metrics. In general, the QoS metrics could be concave, multiplicative, or additive. Let m be the performance metric for the link (u, v) connecting node u to node v , and path $(u, u_1, u_2, \dots, u_k, v)$ a sequence of links for the path P from u to v . Three types of constraints on the path can be identified [15]:

1. *Additive constraint*: A constraint is additive if

$$m(P) = m(u, u_1) + m(u_1, u_2) + \dots + m(u_k, v)$$

For example, the end-to-end delay is an additive constraint because it consists of the summation of delays for each link along the path.

2. *Multiplicative constraint*: A constraint is multiplicative if

$$m(P) = m(u, u_1)m(u_1, u_2) \dots m(u_k, v)$$

The probability of a packet sent from a node u to reach a node v , is multiplicative, because it is the product of individual probabilities along the path.

3. *Concave constraint*: A constraint is concave if

$$m(P) = \min\{m(u, u_1), m(u_1, u_2), \dots, m(u_k, v)\}$$

The bandwidth requirement for a path between node u and v is concave because it consists of the minimum bandwidth between the links along the path. The basic QoS routing problem can be classified into four classes depending on the metric [9]. If the path metric is optimized, then the problem is called link optimization for concave metric and path optimization for additive/multiplicative metric. Similarly if the path metric is constrained, then the problem is called link constrained for concave and path constrained for additive / multiplicative metric. Possible composite routing problems can be obtained from these four basic QoS routing. Examples are link constrained link optimization, link constrained path optimization, path constrained path optimization, etc. The composite QoS routing problems involving two or more additive/multiplicative metrics is NP complete whereas the rest is solvable in polynomial time [15]. To find the path with multiple constraints, the commonly used approach is sequential filtering where the paths based on the primary constraint are selected. The primary set is optimized by eliminating the paths not satisfying the secondary constraint. The formal definition of Multi-Constrained Path (MCP) problem is given later. The MCP problem can have two or more conflicting objectives which leads to *nondominated* solutions [16].

3.1. Multi Constrained Path (MCP) Problem

Consider a network topology modeled by a graph $G = (V, E)$ where V is the set of nodes and E is the set of edges. Each link $(u, v) \in E$ is characterized by m additive QoS metrics. An m -dimensional non-negative weight vector is associated with the link $w_i(u, v)$ for $i = 1, 2, \dots, m$. Let the m positive path constraints be given by L . The MCP problem is given as

$$w_i(P) = \sum_{(u,v) \in P} w_i(u, v) \leq L_i \quad (1)$$

where $i = 1, 2, \dots, m$. The paths that satisfy the above condition are feasible paths.

3.2. Multi Constrained Optimal Path (MCOP) Problem

There may be many paths that satisfy the constraint specified by Eq. (1). The problem that additionally optimizes some length function $l(P)$ is called the Multi-Constrained

Optimal Path. Formally, the problem is to find a path P from a source node s to a destination node d such that

$$w_i(P^*) \leq L_i \text{ for } i = 1, 2, \dots, m$$

where $l(P^*) \leq l(P) \quad \forall P$

3.3. Path Dominance

Consider the number of constraints as 2 (i.e. $m=2$). Let there be two paths P_1 and P_2 from the source to the destination. Each path is characterized by a path weight vector $(w_1(P_1), w_2(P_1))$ and $(w_1(P_2), w_2(P_2))$. If P_1 is shorter than P_2 , then $w_i(P_1) < w_i(P_2)$ for all $1 \leq i \leq m$. In that case, any path from the source to the destination that uses P_1 will always be shorter than P_2 . In another scenario, $w_i(P_1) \leq w_i(P_2)$ for some indices i and $w_j(P_1) > w_j(P_2)$ for atleast one index j . In this case, the two points, are said to be nondominated paths. In summary, path P is called nondominated, if there does not exist a path P' for which $w_i(P') \leq w_i(P)$ for all link weight components i , except for at least one j for which $w_j(P') < w_j(P)$. The set of nondominated solution constitute the pareto optimal set as shown in fig 1.

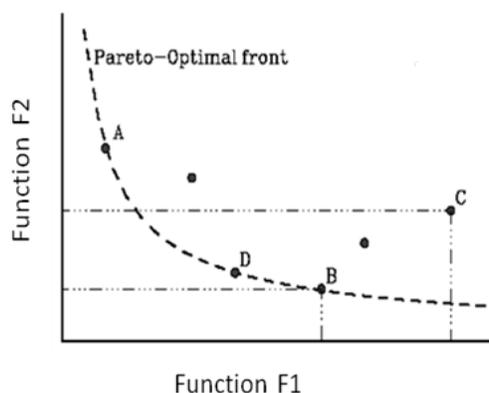


Fig. 1 Pareto set for MCP and MCOP problem

4. QoS AWARE ROUTING PROTOCOLS

Any routing protocol can be verified for its correctness and effectiveness. This requires a formal characterization and quantification of the routing protocol. We provide the analysis of some QoS aware routing protocols (illustrated in Table 1) based on the following factors.

- (a) Type of algorithm
 - a. Unicast – This has exactly one sender and one receiver
 - b. Multicast – In this, the number of senders may be one or more. The senders address the packets to multiple destinations.
 - c. Broadcast – In this, the packets are addressed to all the nodes in the network.
- (b) Network Architecture:
 - a. Flat – In this all nodes are at equal level and maintain the routing table. Thus every node is equally responsible for routing.
 - b. Hierarchical – The nodes designated as clusterheads form a virtual backbone and maintain the routing table.
- (c) Quality of Service
 - a. Constraints – This specifies the number of constraints used for route selection. The routing algorithm may be single or multiconstrained.

- b. Metrics – The QoS requirements are specified in terms of metrics. The metrics may be additive, multiplicative or concave. The most commonly used metric are bandwidth and delay.
- (d) Routing protocol complexity
 - a. Communication complexity – This relates to the exchange of information to have an up-to-date information of the network topology and assist in finding a feasible path satisfying the constraints.
 - b. Memory Complexity – This denotes the amount of memory that is required to be stored to determine the state information.
- (e) Routing Algorithm
 - a. Routing Type :Source Routing, Distributed Routing, Hierarchical Routing
 - b. Route Discovery: Proactive or Table Driven, Reactive or On Demand, Hybrid Routing
 - c. Routing Complexity – This indicates the number of messages needed to discover a feasible route satisfying the constraints.
 - d. Routing Overhead – This specifies the control packets to be exchanged for finding a feasible path.
 - e. Route Caching – This parameter indicates whether the routing algorithm allows route caching or not. If route caching is not allowed then all the communication will be direct from source to destination and the intermediate nodes are not allowed to send the route reply packets.
- (f) Resource Estimation – This estimates the availability of resources in the network. The information is used to find a feasible path satisfying the constraints.
- (g) Resource Reservation – This specifies whether the routing algorithm just determines a feasible path or takes into account the reservation of the path. The reservation may be hard or soft.
- (h) Route Maintenance – The mobility of nodes may cause frequent route breakages. The methods used for route maintenance are
 - a. Route Prediction – This specifies whether the algorithm has a prediction scheme to predict the breakage of route.
 - b. Redundant Routes – The algorithm may find more than one route for route maintenance. One route acts as the primary route and the other routes act as secondary routes.

4.1. Core Extraction Distributed Ad Hoc Routing (CEDAR) [17]

CEDAR is a routing protocols designed for small to medium sized networks. It dynamically establishes a core set for route set up, QoS provisioning, routing data and route maintenance. The link states of stable high bandwidth links are propagated to the core nodes. Route computation is on demand and is performed using only local state. Two assumptions are made in CEDAR: (1) The MAC/link layer can estimate the available link bandwidth and (2) The network consist of tens to hundreds of nodes. The key components in CEDAR are:

Core extraction – A set of nodes is elected to form the core that maintains the local topology, topology and available information exchange, perform route discovery and route maintenance on behalf of all the nodes in its domain. A greedy algorithm is used to create an approximate dominating set proactively. All hosts in the set are either member of the core or one-hop neighbors of core hosts.

State propagation - The bandwidth availability information of stable links is disseminated to other nodes using increase and decrease waves mechanism. The waves are generated when an estimate of the available bandwidth of a core node is changed by a certain amount. The information about the small changes in available bandwidth is kept locally whereas the relatively stable bandwidth information is propagated among the core hosts. The increase in the available bandwidth of the core node is represented by increase waves. This information is

propagated periodically. The decrease wave which represents the decrease in available bandwidth is propagated immediately to prevent the overestimation of the available bandwidth by the core nodes.

Route computation - The route computation first establishes a core path from the domain of the source to the domain of the destination. For route selection, the shortest widest route is chosen among all the feasible routes using a two-phase Dijkstra algorithm. The directional information provided by the core path is used iteratively to find a partial route from the source to the domain of the furthest possible node in the core path satisfying the bandwidth constraints. This node then becomes the source for the next iteration.

Route maintenance - The algorithm uses two strategies for route maintenance - rerouting and repairing. In rerouting, the source is notified of the failure and a new route is computed to reach the destination. In repairing, the link is locally repaired at the point of failure by the surrounding nodes.

The routing overhead in CEDAR is very low due to the presence of core nodes. The increase and decrease wave mechanism for link state propagation also ensures that the available bandwidth information is propagated without incurring a high overhead.

4.2. Ticket Based Probing (TBP) [13]

TBP is a distributed ticket based routing proposed by Chen and Nahrstedt. The basic idea of the TBP is to utilize tickets to limit the number of paths searched during route discovery. A ticket is the permission to search a single path. The tickets are used to find delay constrained or bandwidth constrained routes. In this, when a feasible route is to be established between the source and the destination then the source sends a limited number of probes (routing messages) to some neighboring nodes. Each probe contains at least one ticket. When the connection requirements are tighter, the probes may carry more than one ticket. The total number of tickets for path discovery is constant. At an intermediate node, a probe with more than one ticket may be split into multiple probes with tickets distributed between them. Each probe searches for a different downstream subpath. The decision to split the probe and to which neighbors the probes should be forwarded is taken on the basis of its available state information. TBP is based on the assumption that stable links tend to remain stable in contrast to the transient links. Each node i collects statistical information about delay $D(t)$, bandwidth $B(t)$ and cost $C(t)$ for all other nodes in the network using distance vector. Along with the delay and bandwidth each node also maintains the associated variation $\Delta D(t)$ and $\Delta B(t)$ by which the next reported value will differ with the current one. The protocol searches for a least cost delay constrained path or least cost bandwidth constrained path. For this purpose, it defines two types of tickets - yellow and green. The number of these tickets is based on the imprecise information and look for feasible paths or least constrained path. In the case of route failures, TBP utilizes three mechanisms - path rerouting, path redundancy, and path repairing. Rerouting requires that the source node be informed of a path failure. In path rerouting, the source node is informed about the route failure which then initiates a route discovery to find a new path to the destination. In path redundancy, multiple routes are established from the source to the destination. One path may serve as the primary path and the other paths may be used as backup paths. The resources are reserved only along the primary path to reduce the wastage of resources. In path repairing mechanism, TBP tries to repair the route at the point of failure.

4.3. Ad Hoc QoS On-Demand Routing (AQOR) [18]

Ad hoc QoS On-demand Routing (AQOR) is an on demand QoS aware routing protocol. It integrates (1) bandwidth estimation and end-to-end delay measurement in the route discovery process, (2) bandwidth reservation, and (3) adaptive route recovery. The protocol uses limited flooding to discover the best route available in terms of smallest end-to-end delay and bandwidth guarantee.

The bandwidth estimation is accomplished by disseminating the traffic information to neighbors through periodic announcement packets, called Hello packets. Each node i includes the self traffic in the hello packets. Typically, it consists of three types of traffic (1) Self-traffic ($B_i(\text{self})$) i.e. the traffic between the node i and its neighbors. This traffic indicates the bandwidth consumed by the traffic transmitted or received by node i , (2) Neighborhood traffic ($B_i(\text{neighborhood})$) i.e. the total traffic between i 's neighbors and (3) Boundary traffic ($B_i(\text{boundary})$) i.e. the total traffic between i 's neighbors and nodes that are outside i 's range. The sum of the neighbors' traffic of a node is estimated as the total traffic affecting the node

$$B_i(\text{agg}) = B_i(\text{self}) + B_i(\text{neighborhood}) + B_i(\text{boundary})$$

However, the estimated traffic can be larger than the real overall traffic. The available bandwidth is thus a lower bound on the real available bandwidth and imposes stringent bandwidth admission control threshold.

The route discovery process is activated when a route is desired. The route request packet includes both bandwidth and end-to-end delay constraints. Upon receiving the route request packet, the intermediate nodes perform bandwidth admission hop-by-hop. If the intermediate nodes have sufficient bandwidth, the request is accepted and the node adds the route in its routing table with an expiration time. The status of the node is set as explored and the request is rebroadcast to the next hop. The node remains in the explored status for a period of $2D$ where D is the end-to-end delay. If the reply packet does not arrive within the expiration time, the entry is deleted. The reply packet that arrive late are ignored to reduce overhead and exclude invalid information from the routing table. The smallest delay route with sufficient bandwidth is chosen as the route satisfying the QoS constraints.

The bandwidth reservation is made along the route discovered, but it is activated only when the data flow passes through the route. The reservation is soft and if the node does not receive data packets for a certain time interval, the node immediately invalidates the reservation. The adaptive route recovery consists of QoS violation detection. The end-to-end QoS violations are caused either by congestion or route breakage. These violations are detected by the destination which then initiates a destination initiated route recovery. The end-to-end delay violation is also detected by the destination by monitoring the delay of the arriving packets. If the delay exceeds the maximum delay requirement, QoS recovery is triggered. The route failure or network partition is detected by the neighbor lost detection mechanism. The non arrival of the Hello message in time indicates a route failure. When a neighbor lost is detected, the source is notified about the break which then initiates the reroute process. The QoS violations are detected by the destination.

4.4. Bandwidth Estimation QoS Routing (BEQR) [21]

The aim of this protocol is to provide Soft QoS or better-than-best effort service rather than guaranteed service. The design of the protocol is based on two schemes (1) feedback scheme i.e. providing feedback about the available bandwidth to the application, and (2) admission scheme i.e. admit a flow with the requested bandwidth. Both of these require the estimation of the bandwidth. Thus, bandwidth estimation is the key in the design of BEQR. The route discovery function of this protocol is based on AODV [5] with a modified packet format. The residual bandwidth can be estimated either by listening to the channel and calculating the ratio of free and busy times or by appending the node's current bandwidth with that of its 1-hop neighbors to AODV's periodic hello messages. AODV's route request packets (RREQ) include additional information about the used scheme and either the bandwidth constraint or the minimum of bandwidth constraint and detected bandwidth on the partial path.

4.5. Adaptive Proportional Routing (APR) [23]

The authors in [23] argue that majority of the QoS routing schemes obtain a global view of the network state. This gives rise to prohibitive communication and processing overheads. APR

proposes a localized approach to QoS routing. In this, the source nodes infer the network QoS state on flow blocking statistics collected locally and perform flow routing using this localized view of the network QoS state. For each source-destination pair, the protocol sets up one or multiple explicit routed paths *a priori* using MPLS (Multi-Protocol Label Switching). These paths are referred as candidate paths. Each flow routed along the candidate path has a certain probability of being blocked. The virtual capacity of this path is computed using the knowledge of capacity and blocking rates. APR tries to equally distribute the flows among the available paths w.r.t the virtual capacity of each path selecting the shortest paths. No QoS information is exchanged between the nodes; thus reducing the protocol overhead.

4.6. QoS Enabled AODV (QAODV) [24]

QoS enabled AODV was proposed by Perkins *et al.* [24]. To support QoS, the protocol extends the formats of the RouteRequest and RouteReply packets. This necessitates the need to modify the routing table structure. Four new fields are appended in the routing table for QoS support – (1) Maximum delay (2) Minimum available bandwidth (3) List of sources requesting delay guarantees and (4) List of sources requesting bandwidth guarantees. The QoS constraint is specified in the RouteRequest packets. The intermediate nodes forward the RouteRequest packets if the requested QoS parameter can be satisfied, otherwise the packet is dropped. The actual parameter value that can be satisfied for the path is recorded in the routing table. The intermediate nodes may generate the QoS-LOST message packet if the node detects that the required QoS constraint cannot be satisfied any more. This message is transmitted to all the nodes along the path.

5. CROSS LAYER DESIGN

The traditional network architectures are designed using layering approach. In this approach, the entire network communication functionality is divided into modules known as layers. Each layer fulfils a limited and well defined purpose. Every layer of the system is designed separately and is independent of the application. Every layer offers services to the layer above it and also accepts the services of the layer below it. The information exchange between the layers is only through interfaces which are limited. Thus, the information exchange and coupling between the layers is kept as low as possible. The advantage of this approach is modularity and simplicity. Also the same network architecture can be used by many applications. The negative aspect is that the protocols designed using the layered approach is not optimal for any application. The wireless network characteristics are quite different from the wired networks; so are the challenges. The wireless channel characteristics generally affect all the OSI layers. Optimizing each layer individually to fix the problem leads to unsatisfactory results. It is argued in [19] that it is hard to achieve design goals like energy efficiency and QoS using the traditional layered approach. In other words, a cross layer design (CLD) is needed to achieve the optimal results. Cross layer design seeks to enhance the performance of the system by jointly optimizing multiple protocol layers [22]. The extreme design alternative for CLD is to have a complete layerless approach i.e. collapsing the entire stack to obtain completely integrated protocol architecture. In the other approach some layers of the protocol stack can be merged to obtain the desired results. Since the modular approach has proven itself over time, in yet another approach for CLD, the layers are kept intact and the information is shared between the layers either directly or through a database. Cross layer design is becoming an integral part of several wireless standards. For example 3G standards such as CDMA 2000, Broadband Radio Access Network (BRAN) of HiperLAN2, High Speed Downlink Packet Access (HS-DPA) of 3G Partnership Project and IEEE Study Group Mobile Broadband Wireless Access Networks [2] are based on cross layer design. In designing an architecture using cross layer approach, one has to keep in mind that CLDs without solid architecture guidelines can inevitably lead to “Spaghetti Design”[20]. CLD with tight coupling between the layers becomes hard to review and redesign. The change in one subsystem implies changes in other parts as everything is

connected. This may lead to unpredictable systems as it is hard to foresee the impact of modifications.

5. CONCLUSION

Wireless ad hoc networks are likely to be the center of future communication. Although it is very difficult but providing QoS guarantees has become essential for the operation of today's multimedia wireless networks. This paper presented an overview of the QoS routing protocols and outlined the challenges that make QoS routing difficult in wireless ad hoc networks. We also presented an extensive review of some current existing protocols. However, several important issues remain to be addressed before QoS in wireless ad hoc networks becomes a reality. The energy constraint is of principle interest in wireless ad hoc networks. The future direction for research is to take into consideration the battery constraint while providing QoS. Designing such QoS protocols that optimize multiple objectives is computationally intractable. Thus, a multiobjective protocol architecture design for providing QoS and minimizing energy dissipation has to be thoroughly investigated. The current paradigm shift is towards cross layer optimization to provide energy efficient QoS solutions. Also providing QoS in broadcasting and multicasting has found little attention in the literature.

REFERENCES

- [1] Wu, J. & Stojmenovic, I., (2004) "Ad Hoc Networks", *IEEE Computer*, Vol. 37, No. 2, pp. 29-31.
- [2] B. Tavli and W. Heinzelman, (2006) "Mobile Ad Hoc Networks" Springer Publishers.
- [3] Crawley, E., "A Framework of QoS Based Routing in the Internet", RFC 2386, <http://www.ietf.org/rfc/rfc2386.txt>
- [4] Chen, S., (1999) "Routing Support for Providing Guaranteed End-to-End Quality of Service", Ph. D Thesis, University of Illinois, Urbana-Champaign.
- [5] Perkins, C. E., Royer, E. M. & Das, S. R., (2002) "Ad Hoc On-Demand Distance Vector (AODV) Routing", draft-ietf-manet-aodv-10.txt, IETF MANET working group.
- [6] Johnson, D.B., Maltz, D.A. & Hu, Y.C., (2002) "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks", draft-ietf-manet-dsr-07.txt, IETF MANET working group.
- [7] Masip-Bruin, X. & Kuipers, F., *et al.*, (2006) "Research Challenges in QoS Routing", Elsevier Journal of Computer Communication, Vol. 29, No. 5, pp. 563-581.
- [8] Farooq, M., & Di Caro, G.A., (2008) "Routing Protocols for Next Generation Networks Inspired by Collective Behaviors of Insect Societies: An Overview", *Swarm Intelligence (Natural Computing Series)*, Springer, pp. 101-160.
- [9] Chen, S., & Nahrstedt, K., (1998) "An Overview of Quality of Service Routing for Next Generation High Speed Networks: Problems and Solutions", *IEEE Network*, Vol. 12, No. 6, pp. 64-79.
- [10] Perkins, C.E., & Bhagwat, P., (1994) "Highly dynamic destination-sequenced distance vector routing (DSDV) for mobile computers", *Computer Communication Review*, Vol. 24, No. 4, pp. 234-244.
- [11] Toh, C.K., (1997) "Associativity Based Routing for Ad Hoc Networks" *Wireless Personal Communications*, Vol. 4, No. 2, pp. 1-36.
- [12] Chakrabarti, S., (2003) "Quality of Service in Mobile Ad Hoc Networks", *Handbook of Ad Hoc Wireless Networks*, CRC press.
- [13] Chen S., & Nahrstedt, K., (1999) "Distributed Quality of Service in Ad Hoc Networks", *IEEE Journal on Selected Areas in Communication*, Vol. 17, No. 8, 1999, pp 1-18.

- [14] Mahapatra, P., Li, J., & Gui, C., (2003) "QoS in Mobile Ad Hoc Networks", IEEE Wireless Communications, Vol. 10, No. 3, pp. 44-52.
- [15] Wang Z., & Crowcraft, J., (2002) "Quality of Service Routing for Supporting Multimedia Applications", IEEE Journal on Selected Areas in Communications, Vol. 14, No. 7, pp. 244-256.
- [16] Meigham P., & Kuipers, F., (2004) "Concept of Exact QoS Routing Algorithms", IEEE/ACM Transactions on Networking, Vol. 12, No. 5, pp. 851-864.
- [17] Sivakumar, R., Sinha, P., & Bharghavan, V., (1999) "CEDAR: A Core-Extraction Distributed Ad Hoc Routing Algorithm", IEEE Journal on Selected Areas in Communications, Vol. 17, No. 8, pp. 1454-1465.
- [18] Xue Q., & Ganz, A., (2003) "Ad Hoc QoS On Demand Routing (AQOR) in Mobile Ad Hoc Networks", Elsevier Journal of Parallel and Distributed Computing, Vol. 63, No. 3, pp. 154-165.
- [19] Heinzelman, W., (2000) "Application Specific Protocol Architectures for Wireless Networks", Ph.D Thesis, Massachusetts Institute of Technology, 2000
- [20] Kawadia V., and Kumar, P.R., (2005) "A Cautionary Perspective on Cross Layer Design", IEEE Wireless Communication, Vol. 12, No. 1, pp. 3-11.
- [21] Chen L., & Heinzelman, W., (2005) "QoS Aware Routing Based on Bandwidth Estimation for Mobile Ad Hoc Networks", IEEE Journal on Selected Areas in Communication, Vol. 23, No. 3, pp. 561-572.
- [22] Van der Schaar M., & Sai Shankar, N., (2005) "Cross layer wireless multimedia transmission – challenges, principles and new paradigms", IEEE Wireless Communication, Vol. 12, No. 4, pp. 50-58.
- [23] Nelakuditi, S., Zhi-Li Zhang, Tsang, Rose P., & David H. C. Du, (2002) "Adaptive Proportional Routing: A Localized QoS Routing Approach", IEEE/ACM Transactions of Networking, Vol 10, No. 6, pp. 790-804.
- [24] Perkins, C., & Royer, E., (2003) "Quality of Service for Ad Hoc on-demand Distance Vector Routing", Internet Draft, draft-perkins-manet-aodvqos-02-txt.

Authors

Sunita Prasad received M.Tech degree in Information Technology Guru Gobind Singh Indrapratha University in 2006. She is currently pursuing her Ph.D from Deptt. of Electrical Engg., Faculty of Engineering and Technology, Jamia Millia Islamia (A Central University). Presently she is working in Centre for Development of Advanced Computing (CDAC), NOIDA as a Sr. Lecturer. Her research interest includes network optimization, evolutionary algorithms and adaptive routing. She is a member of ACM, ACEEE and IETE India.



Prof. Zaheeruddin received the B. Sc. Engg. Degree in Electrical and M. Sc. Engg. Degree in Electronics & Communication from Aligarh Muslim University (AMU), Aligarh (UP) in 1982 and 1988 respectively, and a Ph. D Degree in Computer Science from Jawaharlal Nehru University (JNU), New Delhi in 2002. **Zaheeruddin** is currently Professor in the Department of Electrical Engineering, Faculty of Engineering & Technology, Jamia Millia Islamia (A Central University), New Delhi, India since 2003. He is reviewer of many International journals such as IEEE Transactions on Systems, Man, and Cybernetics, Part-A, Applied Soft Computing (Elsevier), and International Journal of Modelling and Simulation (ACTA Press/IASTED). His research interest includes Soft Computing, Fuzzy Logic, Neural Networks, Genetic Algorithms, Neuro-fuzzy Computing, Effects of Environmental Pollution on Humans He is a life member of World Federation of Soft Computing (U.S.A) and Fellow of IETE (India). He is also member of CSI (India), ISTE, ISCA and ICTP. He has published about thirty research papers in the areas of Soft Computing, Fuzzy Logic, Neural Networks, Neuro-fuzzy Computing, and Noise Pollution.



Table 1 : Comparison of Routing Protocols

Parameters	TBP [13]	AQOR [18]	CEDAR [17]	BEQR [21]	APR [23]	QAODV [24]
Type of Algorithm	Unicast	Unicast	Unicast	Unicast	Unicast	Unicast
Network Architecture	Flat	Flat	Hierarchical	Flat	Flat	Flat
Quality of Service						
a. Constraints	Single constrained	Multi-constrained	Single Constrained	Single Constrained	Single Constrained	Single Constrained
b. Metrics	BW or Delay	BW and Delay	BW	BW	BW	BW or delay
Routing Protocol Complexity						
a. Communication Complexity	$O(n)$	$O(n)$ per second	$O(n)$	$O(n)$	Not Available	$O(n)$
b. Space Complexity	$O(n^2)$	$O(n)$	$O(n^2)$	$O(n)$	$O(p)$	$O(n)$
Routing Algorithm						
a. Routing Type	Distributive	Distributive	Distributive	Distributive	Source	Distributive
b. Route Discovery	Reactive	Reactive	Reactive	Reactive	Reactive	Reactive
c. Routing Complexity	$O(t \times n)$	$O(n)$	$O(n)$	$O(2n)$	Not Available	$O(2n)$
d. Routing Overhead	Limited flooding of RREQs	Full flooding of RREQs	Limited: Link state information distributed only among core nodes	Full Flooding of RREQs	Not Available	Full flooding of RREQs
e. Route Caching	No	No	Yes	No	Yes	No
Resource Estimation	No	Yes	No	Yes	No	No
Resource Reservation	Yes (soft)	Yes (soft)	Yes (soft)	No	No	No
Route Maintenance						
a. Route Prediction	No	No	No	No	No	No
b. Redundant Routes	Yes	No	No	No	No	No

Where t = no. of tickets

n = no. of nodes

p = no. of candidate paths connected to the node