

# Issues in Supporting Quality of Service in Mobile Ad Hoc Networks

*M. Scott Corson*

*University of Maryland*

*Institute for Systems Research, University of Maryland, College Park, MD 20742, 301-405-6630, corson@isr.umd.edu*

## Abstract

This paper addresses issues in supporting Quality of Service (QoS)-based delivery in mobile ad hoc networks. It first describes the generic character of mobile ad hoc networks (a.k.a. mobile packet radio networks), some ways in which they differ from non-wireless multihop networks, and the impact these differences have on supporting QoS-based delivery. It then briefly gives some thoughts on supporting QoS in these systems.

## Keywords

Mobile Ad Hoc Networks, Quality of Service

## 1 THE CONTEXT OF MOBILE AD HOC NETWORKING

Providing *strict* QoS guarantees and *robust* service in the face of mobility are *competing* requirements. This competition is amply illustrated every time a cellular user is “dropped” due to mobility. Supporting mobile QoS on the “edge” of a fixed network (where a user is only a *single* wireless hop from a fixed infrastructure as in cellular or wireless LAN technology) is difficult, but supporting QoS without the aid of a fixed infrastructure is more so.

A “mobile ad hoc network” (MANET) can be defined as an autonomous system of mobile routers connected by wireless links, the union of which form an arbitrary graph. MANETs have several salient characteristics:

1) *Dynamic (often rapidly changing) topologies*—Routers are free to move arbitrarily; thus, the network topology—which is typically *multihop*—may change randomly and rapidly at unpredictable times.

2) *Bandwidth-constrained, variable capacity links*—Wireless links will continue to have significantly lower capacity than their hardwired counterparts. The realized throughput of wireless communications—after accounting for the effects of multiple access, fading, noise, and interference conditions, etc.—is often very much less than a radio’s maximum transmission rate.

3) *Power-constrained operation*—Some or all of the nodes in a mobile ad hoc network may rely on batteries for their energy. For these nodes, the most important system design criteria for optimization may be power conservation. Additionally, some envisioned networks (e.g. mobile military networks) may be very *large* (e.g. hundreds or thousands of nodes), which makes the problem of network control even more difficult.

Viewed somewhat abstractly from an end-user, application-level perspective, a communication network is simply one form of a communication “channel”, albeit a somewhat complex one. There are many paths in the channel through which information may flow, each with a potentially different capacity—some, all or none of which may meet a given user’s requirement.

The network channel presented by a mobile ad hoc network differs significantly from that presented by a traditional, multihop, wide area network in two major ways:

1) *Time-varying channel capacity*—In a fixed network, the aggregate network capacity available between any two nodes is static. The task of network resource management is to dynamically allocate fractions of this static capacity to the network’s users. A difficult task, but a simpler one than what follows. In a MANET, the aggregate network capacity available between any two points is continuously changing, and changing simultaneously due to multiple dynamics on differing time scales.

For example, there are fading effects resulting from user movement, shadowing, multiuser interference, all of which change link capacity in a way that cannot be masked by the link layer, and which have a discernible effect at the application level (e.g. audio/video reception). This affects the capacity of a given link or path on a given time scale. Also, radio links first tend to degrade slowly, then—due to the combined effects of coding breakdown and capture-based receiver behavior—simply dropout (referred to as link “failure”). This affects the existence of paths, and results in topological dynamics on a yet slower time scale. Reacting to these network capacity changes on multiple time scales requires protocol action.

2) *Relatively large percentage of capacity required for network control traffic*—A MANET’s time-varying “network channel capacity” presents a more difficult QoS support problem than that faced in non-wireless multihop networks. The differences between largely-static and highly-dynamic networks regarding resource allocation overhead can be thought of in terms of the following connection signaling phases: establishment, tear-down and maintenance. In largely-static networks, connections are established and torn down, but rarely modified once established. If modified, it’s usually due to an end-user request. However, in highly-dynamic networks, while connections are also established and torn down, they typically must be maintained and renegotiated in response to network topology dynamics. In many instances, this maintenance signaling outweighs the initial cost of establishing the connection.

Hence, the amount of network control signaling required to maintain a given

QoS level would appear to be larger (normalized a per-connection basis) than would be required in the non-wireless case. Thus, in a context where the aggregate network capacity is smaller and time-varying, a larger percentage of it must be allocated to control overhead. The places extreme emphasis on the bandwidth efficiency and adaptivity (i.e. low time and communication complexity) of the resource management signaling mechanism. In power-constrained networks, the processing complexity must be kept low as well.

## 2 SUPPORTING QOS IN MANETS—WHAT IS FEASIBLE?

In light of these considerations, what is a reasonable approach to providing QoS? Clearly, the approach should be *adaptive*, but *how* should one adapt? It is very easy—particularly in the context of mobile ad hoc networking—to get entangled in specifically the issue one is trying to avoid. Take, for example, congestion avoidance. How many complex algorithms—with additional complexity added specifically to *avoid* congestion—end up causing *more* congestion, due to their *own overhead*, than the simpler algorithms they replace?

An answer to that question can be given—indirectly—by considering another problem, viz. store-and-forward datagram routing. In the context of fixed networks, it is long-established practice to use some form of shortest-path routing for small to moderately-sized networks, and to resort to some form of hierarchical routing for large networks. Adaptive shortest-path routing algorithms require a minimum amount of communication complexity to permit a shortest-path computation. For these algorithms, this translates—in one form or another—into what can be termed “far-reaching” control message propagation\*. For small to moderately-sized, quasi-static networks, this complexity translates into a communication overhead that is acceptable (i.e. only a small fraction of network bandwidth is required for overhead).

However, in dynamic multihop networks, a large amount of protocol adaptation is required simply to handle topological changes. If the network is bandwidth-constrained, then the resultant routing control overhead may occupy a large fraction of the network’s capacity, preventing transmission of data packets and consequently hurting overall data routing performance. To combat this problem, it is possible to design a “simpler” algorithm—i.e. one that supports a “looser” method of routing, where not even shortest-path routing is attempted. The Temporally-Ordered Routing Algorithm (TORA) (Park *et al.* 1997) is an example of one such algorithm which has *less* communication complexity than a more complex algorithm (e.g. ideal link-state routing) that supports shortest-path routing, yet results in *better* routing performance in

---

\*Far-reaching message propagation results from the need to keep the shortest-path distance estimates up-to-date at each router. While some topological changes may cause only localized algorithmic reaction, others may require many or all of the nodes in the network to be informed—hence the term “far-reaching”. For a more in-depth discussion, see (Park *et al.* 1997).

large, dynamic, bandwidth-constrained networks because of its lower communication overhead (Park *et al.* 1997). The conclusion here is that, in some cases, it is possible to obtain better performance by using algorithms which—at first glance—appear to solve a problem less optimally, yet which, because of the competition for network bandwidth between data and control overhead, are really *more efficient* in terms of delivering more data bits per overhead bit expended.

The previous discussion, while at the outset appearing to discuss routing, is really a discussion on congestion—caused by routing—in MANETs. This points to a QoS architecture which seeks to appropriately balance network and application-level adaptivity to minimize aggregate QoS-related signaling to support a given QoS. Two approaches seem to warrant exploration:

1) **Minimal “Guaranteed” QoS**—Essentially, in this approach a receiver specifies its *needs* (a minimal, required QoS level) and its *wants* (a desired QoS level). A signal may be hierarchically-encoded, and additional levels of fidelity (above a base level) may be requested by a connection and delivered by the network when possible as in (Corson *et al.* 1995). The network does not inform the end-users when it cannot meet the desired QoS level, and the end-users’ application is expected to adapt to any QoS fluctuation between the minimal and desired QoS levels. This approach performs end-to-end admission control and management for the minimal requested QoS level only. Thus both network and application must adapt in this approach: the network adapts to changing channel conditions while trying to deliver the desired QoS, and the application adapts to any QoS fluctuations within its prescribed limits.

2) **“Probabilistic” QoS**—The preceding approach assumes that voice and video (stream traffic) will be a significant fraction of the network traffic, which may, in fact, *not* be the case. In this light, a much more efficient approach to supporting QoS would be a probabilistic, prioritized QoS model involving the use of stateless virtual circuits (VC) at the network level. Such an approach permits instant reaction to routing changes, greatly reduces network control complexity and, through usage of a priority-based scheduling policy at the multiple access level as in HIPERLAN, nearly all VC-based packets make it through the network unimpeded.

## REFERENCES

- Park, V. and Corson, M.S. (1997), A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks, in *Proc. INFOCOM*, Kobe, Japan.
- Park, V. and Corson, M.S. (1997), A Performance Comparison of TORA and Ideal Link State Routing, *submitted to Mobicom*, Budapest, Hungary.
- Corson, M.S. and Batsell, S. (1995), A Reservation-Based Multicast (RBM) Routing Protocol for Mobile Networks: Initial Route Construction Phase, *ACM/Baltzer Wireless Networks Journal*, Vol. 1, No. 4.