QoS Support in Mobile Ad Hoc Networks

Kui Wu and Janelle Harms

Computing Science Department

University of Alberta, T6G 2H1

wkui@cs.ualberta.ca

Abstract

Due to the bandwidth constraint and dynamic topology of Mobile Ad hoc NETworks (MANET), supporting Quality of Service (QoS) in MANETs is a challenging task. Nowadays, a lot of researches have been done on supporting QoS in the Internet and other network architectures, but most of them are not suitable in the MANET environment. In this paper, we review the current researches on QoS support in MANETs, which include QoS models, resource reservation signaling, QoS routing, and QoS Medium Access Control (MAC). The purpose of this paper is twofold. One is to describe a whole picture of QoS support in MANETs; the other is to present the challenges and stimulate more research interests in this area.

Keyword: Mobile Ad hoc NETworks (MANET), QoS

1 Introduction

Since their emergence in the 1970s, wireless networks have become increasingly popular in the network industry. Wireless networks can provide mobile users with ubiquitous communication capability and information access regardless of locations. Conventional wireless networks are often connected to a wired network (e.g. ATM or Internet) so that the ATM or Internet connections can be extended to mobile users. This kind of wireless network requires a fixed wireline backbone infrastructure. All mobile hosts in a communication cell can reach a base station on the wireline network in one hop. In parallel with the conventional wireless networks, another type of model, based on radio to radio multi-hopping, has neither fixed base stations nor a wired backbone infrastructure. In some application environments, such as battlefield communications, disaster recovery etc., the wired network is not available and multi-hop wireless networks provide the only feasible means for communications and information access. This kind of network is called Mobile Ad hoc NETwork (MANET). It is also expected to play an important role in civilian forums such as campus recreation, conferences, and electronic classrooms etc.

A MANET can be seen as an autonomous system or a multi-hop wireless extension to the Internet. As an autonomous system, it has its own routing protocols and network management mechanisms. As a multi-hop wireless extension to the Internet, it should provide a flexible and seamless access to the Internet. Recently, because of the rising popularity of multimedia applications and potential commercial usage of MANETs, QoS support in MANETs has become an unavoidable task.

A lot of work has been done in supporting QoS in the Internet, but unfortunately none of them can be directly used in MANETs because of the bandwidth constraints and dynamic network topology of MANETs. To support QoS, the link state information such as delay, bandwidth, cost, loss rate, and error rate in the network should be available and manageable. However, getting and managing the link state information in MANETs is very difficult because the quality of a wireless link is apt to change with the surrounding circumstances. Furthermore, the resource limitations and the mobility of hosts make things more complicated. The challenge we face is to implement complex QoS functionality with limited available resources in a dynamic environment.

In the literatures, the researches on QoS support in MANETs include QoS models, QoS resource reservation signaling, QoS routing, and QoS Medium Access Control (MAC). Because all of the researches just discuss a certain aspect of QoS in MANETs, it is difficult for the readers to understand the relationships among the researches. Without a whole picture, it is even impossible to understand and evaluate the importance of a particular method. In this paper, we present a comprehensive introduction to the current work on QoS support in MANETs. We describe how the researches differ from and coordinate with each other to deliver QoS in MANETs. With a big picture in mind, we can evaluate which is essential and critical to QoS support with current techniques in MANETs and which is optional but may improve the performance of QoS in future. The relationships among the QoS researches are as follows.

First of all, a QoS model specifies an architecture in which some kinds of services could be provided in MANETs. It is the system goal we should implement. All other QoS components, such as QoS signaling, QoS routing, and QoS MAC must cooperate together to achieve this goal. What is feasible for supporting QoS in MANETs is the first question we must answer because it will influence the functionality of all other QoS components. For example, if we only want to provide differentiated quality of services, signaling for every flow state is unnecessary.

Second, QoS signaling acts as the control center in QoS support. It coordinates the behaviors of QoS routing, QoS MAC, and other components such as admission control and scheduling. The functionality of QoS signaling is determined by the QoS model.

Third, QoS routing searches for a path with enough resources but does not reserve resources. It is the QoS signaling to reserve resources (if necessary in the QoS model) along the path determined by QoS routing or other routing protocols. So QoS routing enhances the chance that enough resources can be guaranteed when QoS signaling wants to reserve resources. Without QoS routing, QoS signaling can still work but the resource reservation may fail because the selected path may not have enough resources. QoS signaling will work better if coordinating with QoS routing. However, since most QoS routing algorithms are complicated, we must balance the benefits against the cost of QoS routing in the bandwidth constraint MANETs.

Fourth, the QoS MAC protocol is an essential component in QoS support in MANETs. All upper-layer QoS components (QoS routing and QoS signaling) are dependent on and coordinate with the QoS MAC protocol.

Finally, other QoS components in MANETs, such as scheduling and admission control, can be borrowed from other network architectures without or with little modifications.

The organization of the rest of the paper is as follows. In Section 2, we describe the QoS models. Section 3 introduces the resource reservation signaling protocols. Section 4 and 5 introduce the QoS routing and QoS MAC layer protocols respectively. Finally, we summarize the paper in Section 6. Because the researches on QoS in MANETs have inherited relationship with QoS in the Internet, some related work in the Internet is also included for the purpose of reference and comparison.

2 QoS Model

The QoS model specifies the architecture in which some kinds of services could be provided in the network. A QoS model for MANETs should first consider the challenges of MANETs, e.g. dynamic topology and time-varying link capacity. In addition, the potential commercial applications of MANETs require the seamless connection to the Internet. Thus the QoS model for MANETs should also consider the existing QoS architectures in the Internet. In this section, we first introduce the QoS models in the Internet, such as IntServ [BCS94] and DiffServ [Blake98], as the background. Then we describe a newly proposed QoS model by H. Xiao [XSLC00] for MANETs.

2.1 IntServ/RSVP on Wired Networks

The basic idea of the Integrated Service (IntServ) [BCS94] model is that the flow-specific states are kept in every IntServ-enabled router [XN99]. A flow is an application session between a pair of end users. A flow-specific state should include the information about bandwidth requirement, delay bound, and cost etc. of the flow. IntServ proposes two service classes in addition to Best Effort Service. One is Guaranteed Service [SPG97]; the other is Controlled Load Service [Wro97]. The Guaranteed Service is provided for applications requiring fixed delay bound. The Controlled Load Service is for applications requiring reliable and enhanced best effort service. Because every router keeps the flow state information, the quantitative QoS provided by IntServ is for every individual flow.

In an IntServ-enabled router, IntServ is implemented with four main components [XN99]: the signaling protocol, the admission control routine, the classifier, and the packet scheduler. Other components, such as the routing agent and management agent, are the original mechanisms of the routers and can be kept unchanged. The Resource ReSerVation Protocol (RSVP) [BZBHJ97] is used as the signaling protocol to reserve resources in IntServ. Applications with Guaranteed Service or Controlled-Load Service requirements use RSVP to reserve resources before transmission. Admission control is used to decide whether to accept the resource requirement. It is invoked at each router to make a local accept/reject decision at the time that a host requests a real-time service along some paths through the Internet. Admission control notifies the application through RSVP if the QoS requirement can be granted or not. The application can transmit its data packets only after the QoS requirement is accepted. When a router receives a data packet, the classifier will perform a Multi-Field (MF) classification [GM99], which classifies a packet based on multiple fields such as source and destination addresses, source and destination port numbers, Type Of Service (TOS) bits and protocol ID in the IP header. Then the classified packet will be put into a corresponding queue according to the classification result. Finally, the packet scheduler reorders the output queue to meet different QoS requirements.

IntServ/RSVP model is not suitable for MANETs due to the resource limitation in MANETs: 1) the amount of state information increases proportionally with the number of flows (the scalability problem, which is also a problem for current Internet). Keeping flow state information will cost a huge storage and processing overhead for the mobile host whose storage and computing resources are scarce. Although the scalability problem may be not likely to happen in current MANET due to its limited bandwidth and relatively small number of flows compared with wired networks, we argue that it will occur with the development of fast radio technology and potential large number of users in the near future; 2) the RSVP signaling packets will contend for bandwidth with the data packets and consume a substantial percentage of bandwidth in MANETs; 3) every mobile host must perform the processing of admission control, classification, and scheduling. This is a heavy burden for the resource-limited mobile hosts.

2.2 DiffServ

Differentiated Service (DiffServ) [Blake98] is designed to overcome the difficulty of implementing and deploying IntServ and RSVP in the Internet backbone [XN99]. DiffServ just provides a limited number of aggregated classes in order to avoid the scalability problem of IntServ. DiffServ defines the layout of the Type Of Service (TOS) bits in the IP header, called the DS field, and a base set of packet forwarding rules, called Per-Hop-Behavior (PHB) [Blake98]. At the boundary of a network, the boundary routers control the traffic entering the network with classification, marking, policing, and shaping mechanisms. When a data packet enters a DiffServ-enabled domain, a boundary router marks the packet's DS field and the interior routes along the forwarding path forward the packet just based on its DS field. Since the DS field only codes very limited service classes, the processing of the interior routers is very simple and fast. Unlike in IntServ, interior routers in DiffServ do not need to keep per-flow state information.

Many services, such as Premium Service [NJZ99], Assured Service [CF98, IN98], and Olympic Service [NJZ99, HBWW98], can be supported in the DiffServ model. Premium Service is supposed to provide low loss, low delay, low jitter, and end-to-end assured bandwidth service. Assured Service is for applications requiring better reliability than Best Effort Service. Its purpose is to provide guaranteed or at least expected throughput for applications. Furthermore, it is more qualitative-oriented than quantitative-oriented and thus easy to implement. Olympic Service provides three tiers of services: Gold, Silver and Bronze, with decreasing quality [XN99]. Supporting Premium Service is more difficult and almost impossible in current MANETs because the strict requirement of Premium Service is not suitable in the dynamic MANET environment.

DiffServ may be a possible solution to the MANET QoS model because it is lightweight in interior routers. In addition, it provides Assured Service, which is a feasible service context in MANET. However, since DiffServ is designed for fixed wire networks, we still face some challenges to implement DiffServ in MANETs. First, it is ambiguous as to what are the boundary routers in MANETs. Intuitively, the source nodes play the role of the boundary routers. Other nodes along the forwarding paths from the sources to the destinations are interior nodes. But every node should have the functionality as both boundary router and interior router because the source nodes can not be predefined. This arouses a heavy storage cost in every host. Second, the concept of Service Level Agreement (SLA) in the Internet does not exist in MANETs. The SLA is a kind of contract between a customer and its Internet Service Provider (ISP) that specifies the forwarding services the customer should receive [Blake98]. In the Internet, a customer must have a Service Level Agreement (SLA) with its Internet Service Provider (ISP) in order to receive Differentiated Services. The SLA is indispensable because it includes the whole or partial traffic conditioning rules [Blake98], which are used to re-mark traffic streams, discard or shape packets according to the traffic characteristics such as rate and burst size. How to make a SLA in MANETs is difficult because there is no obvious scheme for the mobile nodes to negotiate the traffic rules.

2.3 FQMM

A Flexible QoS Model for MANET (FQMM) is proposed in [XSLC00]. It considers the characteristics of MANETs and tries to take advantage of both the per-flow service granularity in IntServ and the service differentiation in DiffServ.

As in DiffServ, three kinds of nodes (ingress, interior, and egress nodes) are defined in FQMM. An ingress node is a mobile node that sends data. Interior nodes are the nodes that forward data for other nodes. An egress node is a destination node. Note that the role of a mobile node is adaptively changing based on its position and the network traffic.

The provisioning in FQMM, which is used to determine and allocate the resources at various mobile nodes, is a hybrid scheme of per-flow provisioning as in InterServ and per-class provisioning as in DiffServ. FQMM tries to preserve the per-flow granularity for a small portion of traffic in MANET, given that a large amount of the traffic belongs to per aggregate of flows, that is, per-class granularity. A traffic conditioner is placed at the ingress nodes where the traffic originates. It is responsible for re-marking the traffic streams, discarding or shaping packets according to the traffic profile, which describes the temporal properties of a traffic stream such as rate and burst size.

FQMM is the first attempt at proposing a QoS model for MANETs. However, some problems still need be studied. First, how many sessions could be served by per-flow granularity? Without an explicit control on the number of services with per-flow granularity, the scalability problem still exists. Second, just as in DiffServ, the interior nodes forward packets according to a certain PHB that is labeled in the DS field. We argue that it is difficult to code the PHB in the DS field if the PHB includes per-flow granularity considering the DS field is at most 8 bits without extension. Finally, how to make a dynamically negotiated traffic profile is a very difficult problem.

3 QoS Signaling

QoS signaling is used to reserve and release resources, set up, tear down, and renegotiate flows in the networks. Two distinct mechanisms should be included in a QoS signaling system. First, the QoS signaling information must be reliably carried between the routers. Second, the QoS signaling information must be correctly interpreted and the relative processing should be activated. Based on the first mechanism, the QoS signaling system can be divided into in-band signaling and out-of-band signaling. The in-band signaling refers to the fact that control information is carried along with data packets [LC98]; the out-of-band signaling refers to the approach that uses explicit control packets.

3.1 RSVP

Resource reSerVation Protocol (RSVP) [BCS94] is adopted as the signaling system in the Internet. We simply introduce RSVP in order to provide a useful comparison of signaling in the Internet with signaling in MANETs.

RSVP is an out-of-band signaling system. The main motivation for RSVP is to allow efficient support for establishing multicast and unicast connections. For simplicity of explanation, we just describe the unicast case. When a source host wants to send information to a receiver, it first sends a PATH message to the receiver. The PATH message includes the specification of the traffic characteristics such as rate and burst size. Each intermediate router forwards the PATH message to the next hop determined by the routing protocol. Upon getting the PATH message, the receiver sends back a RESV message to the sender. This RESV message includes the resource requirement for the flow. When a router receives the RESV message, it checks if the required resource can be satisfied. If yes, it allocates the resource for the flow, stores the flow state information, and goes on forwarding the RESV message back to the sender. Otherwise, the resource request is rejected and an error message is sent to the receiver.

RSVP has two important characteristics. First, it is the receiver, instead of the sender, that initiates the resource request. Note that different receivers may have different requirements in the multicast case. Second, the flow and reservation information is periodically refreshed. This feature is important in case of link failures. Currently, RSVP has been modified and extended to include more mechanisms, for example, resource reservation for aggregation of flows [GBH97].

RSVP is not suitable for MANET since the signaling overhead of RSVP is heavy for the mobile hosts. The signaling control message will contend with data packets for the channel and cost a large amount of bandwidth. Furthermore, it is not adaptive for the time-varying topology because it has no mechanism to rapidly respond to the topology change in MANETs.

3.2 INSIGNIA

INSIGNIA [LC98, ACLZ99] is an in-band signaling system that supports QoS in MANETs. To our knowledge, it is the first signaling protocol designed solely for MANETs. The signaling control information is carried in the IP option of every IP data packet, which is called the INSIGNIA option. Like RSVP, the service granularity supported by INSIGNIA is per-flow management. Each flow state information is established, restored, adapted and removed over an end-to-end session in response to topology change and end-to-end quality of service condition.

Figure 1 [ACLZ99] shows the position and the role of INSIGNIA in wireless flow management at a mobile host. The packet forwarding module classifies the incoming packets and forwards them to the appropriate modules (routing, INSIGNIA, local applications, and packet scheduling modules) [LC98]. If a received IP packet includes an INSIGNIA option, the control information is forwarded to and processed by the INSIGNIA module. In the meantime, the received packet is delivered to a local application or forwarded to the packet scheduling module according to the destination address in the IP head. If the mobile host is the destination of the packet, the packet is processed by a local application. Otherwise the mobile host will forward the packet to the next hop determined by the MANET routing protocol. Before the packets are sent through the MAC component, a packet scheduling module is used to schedule the output of the flows in order to fairly allocate the resource to different flows. In INSIGNIA, a Weighted Round-Robin (WRR) discipline that takes location dependent channel conditions into account [LBS97] is implemented. Note that a wide variety of scheduling disciplines could be used to realize the packet scheduling.

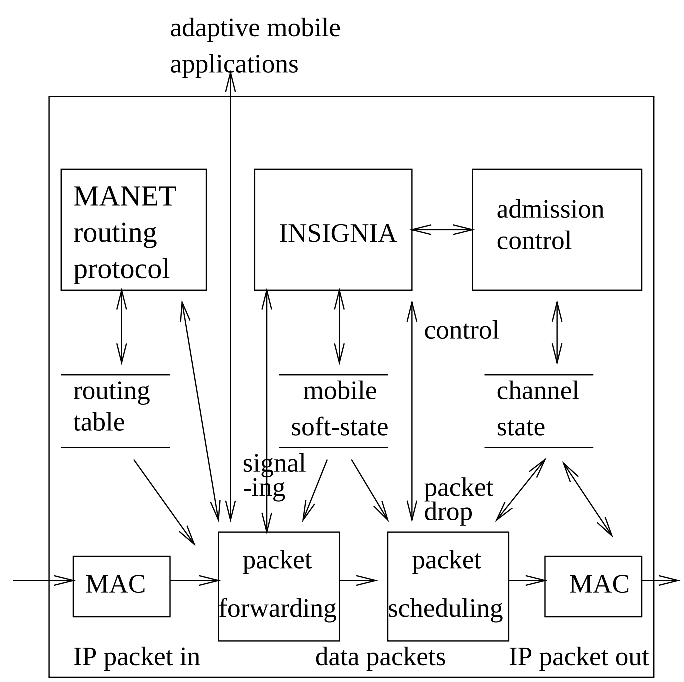

Figure 1: Wireless flow management model at a mobile host [ACLZ99]

The INSIGNIA module is responsible for establishing, restoring, adapting and tearing down real-time flows. It includes fast flow reservation, restoration and adaptation algorithms that are specifically designed to deliver adaptive real-time service in MANETs [LC98]. The flow state information is managed in soft-state method, that is, the flow state information is periodically refreshed by the received signaling information. Coordinating with the admission control module, INSIGNIA allocates bandwidth to the flow if the resource requirement can be satisfied. Otherwise, if the required resource is unavailable, the flow will be degraded to best-effort service. To keep the processing simple and lightweight, INSIGNIA does not send rejection and error messages if the resource request is not satisfied.

For fast responding to the changes in network topology and end-to-end quality of service conditions, INSIGNIA uses QoS reports to inform the source node of the status of the real-time flows. The destination node actively monitors the received flows and calculates QoS statistical results such as loss rate, delay, and throughput etc. The QoS reports are periodically sent to the source node. Through this kind of feedback information, the source node can take corresponding actions to adapt the flows to observed network conditions.

As a whole, INSIGNIA is an effective signaling protocol for MANETs. Coordinating with other network components (viz. routing protocol, scheduling, and admission control), INSIGNIA can efficiently deliver adaptive real-time flows in MANETs. However, since the flow state information should be kept in the mobile hosts, the scalability problem may hinder its deployment in the future.

3.3 Out-of-band Signaling vs. In-band Signaling

Signaling is usually the most complex component in a computer network [Kes97]. This is because signaling must support complex network services and have strict requirements on performance and reliability. For example, in MANETs, the flows should be rapidly established with minimal signaling overhead; active flows should be maintained even if the topology changes and rerouting happens; flow state information should be automatically removed when the session is finished. All of these are non-trivial tasks. In addition, signaling software should be extensible and maintainable because people often want to add new services into the existing network. For example, the mechanism of resource management for aggregation of flows should be provided if we want to support per-class based service in MANETs.

The complexity of the signaling system requires clean solutions to deal with signaling independently. This is the main reason why most signaling systems use explicit out-of-band control messages. Because the control messages do not rely on the transmission of data packets, it is flexible to implement the out-of-band signaling system. Furthermore, the supported services can be rich and powerful. In some cases, it is even impossible to implement some functionality with absolute in-band signaling. For example, in the circumstance that the data flow is totally unidirectional from the source to the destination once it is established, sending a feedback control message back to the source with absolute in-band signaling is impossible during the session of the flow. We use the term absolute in-band signaling to refer that all control information is piggybacked within data packets.

Out-of-band signaling, however, consumes more network bandwidth. In wireless networks, it competes for transmission channel with data packets. Since the signaling messages should have higher priority over data packets, a complex out-of-band signaling system will greatly degrade the performance in bandwidth-constrained MANETs. On the other hand, in-band signaling carries the control information in data packets. It is usually lightweight and simple. Although in-band signaling costs some bandwidth more or less, it will not contend for the transmission channel with data packets since it is included in every data packet. This feature is very important in wireless networks, where the transmission channel is shared by all neighboring hosts.

Due to bandwidth and power constraints, keeping the signaling lightweight and simple is more important than designing a powerful but complex signaling system. At least at present, this should be one of the main principles of designing signaling system in MANETs.

4 QoS Routing Protocols

4.1 The Difficulties of QoS Routing

QoS routing protocols search for routes with sufficient resources for QoS requirements. The QoS routing protocols should work together with resource management to establish paths through the network that meet end-to-end QOS requirements, such as delay or delay jitter bounds, bandwidth demand, or multi-metric constraints. The QoS metrics could be concave or addictive.

Definition Concave and additive QoS metrics [CN98-2]: Let m(i,j) be a QoS metric for link (i,j). For a path P = (s, i, j, ..., l, t), metric m is concave if m(P) = min{m(s,i), m (i,j), ..., m(l,t)}. Metric m is additive if m(P) = m(s,i)+ m (i,j)+ ...+ m(l,t).

The bandwidth metric is concave, i.e., a certain amount of bandwidth must be required on each link along the path. Delay, delay jitter, and cost are additive. Wang et al. proved that if QoS contains at least two additive metrics, then the QoS routing is an NP-complete problem [WC96]. Therefore, searching for the shortest path with minimal cost and finding delay-constraint least-cost paths are NP-complete problems. Because of this reason, we only seek approximated solutions for these problems.

QoS routing is difficult in MANETs. First, the overhead of QoS routing is too high for the bandwidth limited MANETs because the mobile host should have some mechanisms to store and update link state information. We have to balance the benefit of QoS routing against the bandwidth consumption in MANETs. Second, because of the dynamic nature of MANETs, maintaining the precise link state information is very difficult. Third, the traditional meaning that the required QoS should be ensured once a feasible path is established is no longer true. The reserved resource may not be guaranteed because of the mobility-caused path breakage or power depletion of the mobile hosts. QoS routing should rapidly find a feasible new route to recover the service.

Comparing with the abundant work on QoS routing for fixed wire networks [CN98], results for QoS routing in MANETs are relatively scarce due to the above difficulties. Even the question of whether to support QoS routing in MANETs is still a hotly debated issue. However, some promising work on QoS routing in MANETs, such as CEDAR [SSB99], ticket-based probing [CN99], and QoS routing based on bandwidth calculation [LL99], have been done and show good performance. Below we introduce CEDAR as an example to illustrate how to deal with the above difficulties. We also simply introduce the ticket-based probing algorithm.

4.2 CEDAR

P. Sinha et al. proposed a Core-Extraction Distributed Ad-hoc Routing (CEDAR) algorithm [SSB99], which can react effectively to the dynamics in MANETs. It includes three main components: core extraction, link state propagation, and route computation.

4.2.1 Core Extraction

Definition Dominating set of a network: The dominating set of a network is a set of hosts, say DS, such that every host in the network is either in DS or a neighbor of a node in DS. The minimum set of the DS is called the minimum dominating set of the network.

The purpose of core extraction is to elect a set of hosts to form a core of the network by using only local computation and local state. The core of the network is an approximation of a minimum Dominating Set (DS) of the network. Every host in DS is called a core host. Every host not in DS chooses one of its neighbors who are in DS as its dominator. Note that the dominator of a core host is itself. Two core hosts are called nearby core hosts if the distance between them is no more than 3. A path between two core hosts is called a virtual link. The graph that consists of core nodes and virtual links to connect nearby core hosts is called a core graph. A core path is a path in the core graph.

CEDAR presents a distributed algorithm to choose core nodes. When a host loses connectivity with its dominator due to mobility, it either finds a core neighbor as its dominator, or nominates one of its non-core-host neighbors to join the core, or itself joins the core. Further details can be found in [SSB99].

Since flooding in MANETs causes repeated local broadcasts, it is highly unreliable because of the abundance of hidden and exposed hosts [Bha98]. CEDAR proposes the core broadcast mechanism to ensure that each core host does not transmit a broadcast packet to every nearby core host. The core broadcast approach has very low overhead and adapts easily to topology changes. It also provides an efficient way to update link state information.

4.2.2 Link State Propagation

In order to compute the feasible QoS paths in CEDAR, each core host maintains its local topology as well as the link-state corresponding to stable high-bandwidth links further away. Note that it does not keep the link state information of unstable or low-bandwidth further links, because these links are not useful in searching for the QoS routes. To achieve this goal, CEDAR utilizes the increase/decrease waves.

For every link $l = (a,b)$, the host a and b are responsible for monitoring the available bandwidth on the link. When the link l comes up or the bandwidth of the link l increases beyond a given threshold value, host a and b will notify their dominators to initialize a core broadcast for an increase wave, which indicates the stable high-bandwidth link. On the other hand, if the link l breaks down or the bandwidth of the link l decreases beyond a given threshold value, host a and b inform their dominators to initialize a core broadcast for a decrease wave, which indicates the unstable or low-bandwidth link. The increase wave is slow-moving, while the decrease wave is fast-moving. For the same link state, the fast-moving decrease wave will take over and kill the slow-moving increase wave. Finally, the survivable increase wave will propagate the stable high-bandwidth link state information through the cores. In addition, CEDAR provides a mechanism that keeps the decrease wave from propagating throughout the whole network. So the unstable low-bandwidth link states are kept locally.

4.2.3 Route Computation

The QoS route computation in CEDAR includes three main steps: (a) discovering the location of the destination and establishing a core path to the destination, (b) searching for a stable QoS route with the established core path as a directional guideline, and (c) dynamically re-computing QoS routes upon link failures or topology changes.

When a source s wants to send messages to a destination d, it first sends a <s, d, b> triple to its dominator, dom(s), where b is the required bandwidth. If dom(s) can calculate a feasible path to d with its local state information, it responds to s immediately. Otherwise, dom(s) discovers the dom(d) if it does not know the location of d, and simultaneously establishes a core path to d. A core path request message is initialized and core-broadcasted by dom(s). By the virtue of the core broadcast algorithm, the core path request message traverses an implicitly established source routed tree from dom(s), which is typically a breadth-first search tree. Thus the core path is approximately the shortest path in the core graph from dom(s) to dom(d) and provides a good directional guideline for the calculation of QoS routes.

Because dom(s) knows the up-to-date local topology and only some possibly out-of-date link state information about remote stable high-bandwidth links, dom(s) may not be able to calculate a possible path to d with enough required bandwidth based on its own link state knowledge. However, as mentioned before, the core path from dom(s) to dom(d) provides a good directional guideline for the possible QoS routes. Based on its own link state information, dom(s) will try to calculate a route with enough bandwidth to meet the bandwidth requirement to the furthest core node, dom(t), which is on the core path from dom(s) to dom(d). Then dom(s) sends dom(t) a message to notify it of continuing the same computation further. If dom(t) can calculate a route with enough bandwidth to d based on its own link state knowledge, then the computation is finished and a feasible path with enough bandwidth from dom(s) to d is found. Otherwise, dom(t) repeats the same operation as dom(s). The computation will continue along the core path from dom(s) to dom(d) step by step. Finally, at a core node t_n, a feasible path with enough bandwidth from t_n to d is found or no possible path could be produced at t_n. In the first case, the whole feasible path is the concatenation of the partial paths computed by the core nodes s, t, $\ldots t_n$. In the latter case, the bandwidth requirement can not be satisfied and the request is rejected.

CEDAR deals with link failures by two mechanisms: (1) dynamic re-computation of a feasible route at the point of failure, and (2) notification back to the source to activate re-computation of a feasible route at the source. The two mechanisms work in concert to respond to topology changes.

Above we have simply introduced the main techniques in CEDAR. The simulation of CEDAR shows that it can compute good admissible routes with high probability and still adapt effectively with low overhead to the dynamics of the network topology [SSB99]. These characteristics are very important to QoS routing in MANETs.

4.3 Ticket-based Probing

S. Chen and K. Nahrstedt proposed a ticket-based probing algorithm [CN99] for QoS routing in MANETs. The basic idea is using tickets to limit the number of candidate paths. When a source wants to find QoS paths to a destination, it issues probe messages with some tickets. The number of the tickets is based on the available state information. One ticket corresponds to one path searching; and one probe message should carry at least one ticket. So the maximum number of the searched paths is bounded by the tickets issued from the source. When an intermediate host receives a probe message with n tickets, based on its local state information, it decides whether to and how to split the n tickets and where to forward the probe(s). When the destination host receives a probe message, a possible path from the source to the destination is found.

Some questions must be answered in the above route search approach. First, how many tickets should be issued by the source? Second, when an intermediate host receives a probe message with n tickets, it must decide (1) whether and how to split the n tickets and (2) where to forward the probe message(s). What are the ticket-splitting and probe-forwarding rules? Third, how to dynamically maintain the multiple paths?

S. Chen and K. Nahrstedt solved these questions in detail [CN99]. Simply stated, for the first question, more tickets are issued for the connections with tighter or higher requirements. For the second question, the link with larger residual bandwidth gets more tickets. For the third question, the techniques of re-routing, path redundancy, and path repairing are used.

5 QoS MAC Protocols

5.1 The Mechanisms of QoS MAC Protocols

QoS supporting components at upper layers, such as QoS signaling and QoS routing, assume the existence of a MAC protocol, which solves the problems of medium contention, supports reliable unicast communication, and provides resource reservation for real-time traffic in a distributed wireless environment. A lot of MAC protocols [FG95, Karn90, BDSZ94, TG97] have been proposed for wireless networks. Unfortunately, their design goals are usually to solve medium contention and hidden/exposed terminal problems and improve throughput. Most of them do not provide resource reservation and QoS guarantees to real-time traffic.

The first problem that a MAC protocol in wireless networks should solve is the hidden/exposed terminal problem. For convenience to later discussion, we simply describe the problem and the RTS-CTS dialogue as its basic solution. As shown in Figure 2, host A and host C can not hear each other. When A is transmitting a packet to B, C can not sense the transmission from A. Thus C may transmit a packet to B and cause a collision at B. This is the "hidden terminal" problem since A is hidden from C. Similarly, when B is transmitting a packet to C, A can not initiate a transmission to D, since this can potentially cause collisions of the control packets at both B and A, thereby disrupting both transmissions. This is the "exposed terminal" problem since A is exposed to B. An RTS-CTS dialogue can be used to solve the hidden/exposed terminal problem. In Figure 2, when C wants to send a data packet to B, it first sends a Request-To-Send (RTS) message to B. When B receives the RTS, it broadcasts a Clear-To-Send (CTS) message to C and A. When C receives the CTS, it begins to transmit the data packet. Upon receiving the CTS, A will defer its data transmission because it knows B will receive data from C. This method avoids the possible collisions at host B and thus solves the hidden terminal (A is hidden from C) and exposed terminal (A is exposed to B) problems.

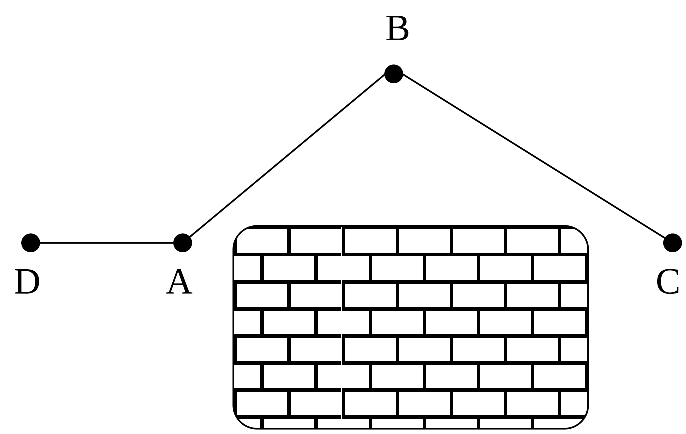

Figure 2: A is hidden from C and exposed to B

Another dialogue frequently used in MAC protocols is the PKT-ACK dialogue, which means the sender sends a data packet (PKT) to the receiver and the receiver immediately responds with an acknowledgment packet (ACK) to the sender if the data packet is correctly received. Failure to receive the ACK will prompt a retransmission after a short timeout within the link level.

Besides dealing with the hidden/exposed terminal problems, a QoS MAC protocol must provide resource reservation and QoS guarantees to real-time traffic. The GAMA/PR protocol [MG98] and the newly proposed Black-Burst (BB) contention mechanism [SK99] can provide QoS guarantees to real-time traffic in a distributed wireless environment. However, they are supposed to work in a wireless LAN in which every host can sense each other's transmission, or in a wireless network without hidden hosts. Below we introduce MACA/PR [LG97], a MAC layer protocol for QoS support in MANETs.

5.2 MACA/PR

C.R. Lin and M. Gerla [LG97] proposed Multiple Access Collision Avoidance with Piggyback Reservation (MACA/PR) for multihop wireless networks. MACA/PR provides rapid and reliable transmission of non-real-time datagrams as well as guaranteed bandwidth support to real-time traffic.

For the transmission of non-real-time datagrams in MACA/PR, a host with a packet to send must first wait for a free "window" in the reservation table (RT), which records all reserved send and receive windows of any station within the transmission range. It then waits for an additional random time on the order of a single hop round trip delay. If it senses that the channel is free, it proceeds with RTS-CTS-PKT-ACK dialogue for a successful packet transmission. If the channel is busy, it waits until the channel becomes idle and repeats the above procedure.

For the transmission of real-time packets, the behavior of MACA/PR is different. In order to transmit the first data packet of a real-time connection, the sender S initiates an

RTS-CTS dialogue and then proceeds with PKT-ACK dialogues if the CTS is received. For subsequent data packets (not the first one) of a real-time connection, only PKT-ACK dialogues are needed. Note that if the sender fails to receive several ACKs, it restarts the connection with the RTS-CTS dialogue again. MACA/PR does not retransmit the real-time packets after collision.

In order to reserve bandwidth for real-time traffic, the real-time scheduling information is carried in the headers of PKTs and ACKs. The sender S piggybacks the reservation information for its next data packet transmission on the current data packet (PKT). The intended receiver D inserts the reservation in its reservation table (RT) and confirms it with the ACK to the sender. The neighbors of the receiver D will defer their transmission once receiving the ACK. In addition, from the ACK, they also know the next scheduled receiving time of D and avoid transmission at the time when D is scheduled to receive the next data packet from S. The real-time packets are protected from hidden hosts by the propagation and maintenance of reservation tables (RT) among neighbors, not by the RTS-CTS dialogues. Thus, through the piggybacked reservation information and the maintenance of the reservation tables, the bandwidth is reserved and guaranteed for the real-time traffic.

6 Summary

In this paper, we review the current researches on QoS support in MANETs. Although these researches focus on different problems, they are highly related to each other and have to deal with some common difficulties, which include mobility, limited bandwidth and power consumption, and broadcast characteristic of radio transmission.

Because of the mobility of the hosts, an established path may break and the reserved resources may not be available again. Since mobile hosts could move in an arbitrary way, the traditional meaning that some performance metrics must be guaranteed once a request is accepted is no longer true in MANETs. In addition, the topology changes could also affect the available bandwidth. How to minimize the influence of the hosts' movement should be the first consideration for QoS support in MANETs.

Because of the bandwidth and power limitation, the cost on providing QoS should be controlled in a reasonable range. Compared with the work on QoS support in wired networks, our goal is not to design strong but complex QoS mechanisms. Instead, finding a way to support sort of QoS service with little burden on the mobile hosts will be a main principle in current MANETs.

Because of the broadcast characteristic of radio transmission, abundant hidden/ exposed terminals may exist, which makes reliable transmission and end-to-end delay control very difficult. Since the transmission between neighboring nodes could interfere with each other, the available bandwidth in a mobile host changes with the surrounding environment, such as how many neighboring hosts are contending for the transmission channel, how many neighboring hosts are moving away or how many non-neighboring nodes are moving near etc. All these factors will increase the complexity of QoS support in the MANET environment.

As a whole, the above difficulties constitute the main challenges for QoS support in MANETs.

References

[ACLZ99] G-S. Ahn, A. T. Campbell, S-B. Lee, and X. Zhang, "INSIGNIA," *Internet Draft*, draft-ietf-manet-insignia-01.txt, Oct. 1999.

[BCS94] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture- an Overview," *IETF RFC1663*, June 1994

[BDSZ94] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A Media Access Protocol for Wireless LANs," *Proc. of ACM SIGCOMM'94*, pp.212-225, 1994.

[Bha98] V. Bharghavan, "Performance of multiple access protocols in wireless packet networks," *International Performance and Dependability Symposium*, Durham, North Carolina, Sept. 1998.

[Blake98] S. Blake, "An Architecture for Differentiated Services," *IETF RFC2475*, December 1998.

[BZBHJ97] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource reSerVation Protocol (RSVP)- Version 1 Functional Specification," *RFC 2205*, Sept. 1997.

[CF98] D. D. Clark and W. Fang, "Explicit Allocation of Best-effort Packet Delivery Service," *IEEE/ACM Transactions on Networking*. Vol. 6, No. 4, pg. 362-373, August 1998.

[CN98] S. Chen and K. Nahrstedt, "An Overview of Quality-of-Service Routing for the Next Generation High-Speed Networks: Problems and Solutions, " *IEEE Network*, Special Issue on Transmission and Distribution of Digital Video, Nov./Dec. 1998.

[CN98-2] S. Chen and K. Nahrstedt, "Distributed Quality-of-Service Routing in High-Speed Networks Based on Selective Probing," *23rd Annual Conference on Local Area Networks (LCN'98)*, October 1998.

[CN99] S. Chen and K. Nahrstedt, "Distributed Quality-of-Service Routing in Ad-Hoc Networks," *IEEE Journal on Special Areas in Communications*, Vol. 17, No. 8, August 1999.

[FG95] Fullmer and J.J. Garcia-Luna-Aceves, "Floor Acquisition Multiple Access (FAMA) for Packet-radio Networks," *Proc. SIGCOMM'95*, Cambridge, MA, pp. 262-273, 1995.

[GBH97] R. Guerin, S. Blake, and S. Herzog, "Aggregating RSVP-based QoS Requests," *Internet draft*, draft-guerin-aggreg-RSVP-00.txt, Nov. 1997.

[GM99] P. Gupta and N. McKeown, "Packet classification on multiple fields," *Proceeding of ACM SIGCOMM'99 conference on Applications, technologies, architectures, and protocols for computer communications*, ACM, Cambridge, MA, Aug. 1999.

[HBWW98] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB Group," *Internet Draft*, draft-ietf-diffserv-af-03.txt, Nov. 1998.

[IN98] J. Ibanez and K. Nichols, "Preliminary Simulation Evaluation of an Assured Service," *Internet Draft*, draft-ibanez-diffserv-assured-eval-00.txt, August 1998

[Karn90] P. Karn, "MACA- A New Channel Access Method for Packet Radio," *Proc. ARRL/CRRL Amateur Radio Ninth Computer Networking Conf.*, pp. 134-140, ARRL, 1990.

[Kes97] S. Keshav, "*An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network*," Addison-Wesley professional computing series, 1997.

[LBS97] S. Lu, V. Bharghavan, and R. Srikant, "Fair Scheduling in Wireless Packet Networks," *Proceedings of ACM SIGCOMM*, San Francisco, California, 1997.

[LC98] S-B. Lee and A.T. Campbell, "INSIGNIA: In-band Signaling Support for QOS in Mobile Ad Hoc Networks," *Proc of 5th International Workshop on Mobile Multimedia Communications (MoMuC,98)*, Berlin, Germany, October 1998.

[LG97] C.R. Lin and M. Gerla, " MACA/PR: An Asynchronous Multimedia Multihop Wireless Network," *Proceedings of IEEE INFOCOM'97*, 1997.

[LL99] R. Lin and J.S. Liu, "QoS Routing in Ad Hoc Wireless Networks," *IEEE Journal on Selected Areas in Communications*, August 1999.

[MG98] Muir and J.J. Garcia-Luna-Aceves, "An Efficient Packet-sensing MAC Protocol for Wireless Networks," *ACM Jounal on Mobile Networks and Applications*, Vol. 3, No. 2, pg. 221-234, August 1998.

[NJZ99] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet," *IETF RFC2638*, July 1999.

[SK99] J.L. Sobrinho and A. S. Krishnakumar, "Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks," *IEEE Journal on Special Areas in Communications*, Vol. 17, No. 8, August 1999.

[SPG97] S. Shenker, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service," *RFC 2212*, Sept. 1997.

[SSB99] P. Sinha, R. Sivakumar, and V. Bharghavan, "CEDAR: a Core-Extraction Distributed Ad hoc Routing algorithm," *IEEE Infocom'99*, New York, NY. March 1999

[TG97] Talucci and M. Gerla, "MACA-BI (MACA By Invitation). A Wireless MAC Protocol for High Speed Ad hoc Networking," *Proceedings of IEEE ICUPC'97*, 1997.

[XN99] Xipeng Xiao and Lionel M. Ni, "Internet QoS: a Big Picture," *IEEE Network Magazine*, March 1999.

[XSLC00] H. Xiao , W.K.G. Seah, A. Lo, and K. C. Chua, "A Flexible Quality of Service Model for Mobile Ad-Hoc Networks," *IEEE VTC2000-spring*, Tokyo, Japan, May 2000.

[WC96] Z. Wang and J. Crowcroft, "QoS routing for supporting resource reservation," *JSAC*, Sept. 1996.

[Wro97] J. Wroclawski, "Specification of the Controlled-Load Network Element Service," *RFC 2211*, Sept. 1997.