A Paradigm for Quality-of-Service in Wireless Ad Hoc Networks Using Synchronous Signaling and Node States

John A. Stine, *Member, IEEE* and Gustavo de Veciana, *Senior Member, IEEE*

***Abstract*—Most limitations in mechanisms geared at achieving quality-of-service (QoS) in wireless ad hoc networking can be traced to solutions based on mapping wireless networks to a wireline paradigm of nodes and links. We contend that this paradigm is not appropriate since links are not physical entities and do not accurately represent the radio frequency (RF) media. Using the link abstraction makes arbitration of the use of the RF media cumbersome leaving only overprovisioning techniques to deliver QoS. In this paper, we argue that an appropriate paradigm should match the physics of the network. The critical resource is electromagnetic spectrum in a space; in turn, this results in a complex paradigm since the part of the spectrum-space that each node wants to use is unique to that node and its destination and will overlap with parts that other nodes may want to use creating interdependences among nodes. This paper describes protocol approaches for access and routing that seek solutions within this wireless paradigm. Access is arbitrated using synchronous signaling and topology is resolved through the dissemination of node states. This approach provides an intuitive framework that provides mechanisms that can be exploited to arbitrate RF media use and implement traffic engineering techniques to deliver QoS. Our proposed approach provides a novel way of tracking the state of the network that can serve as a unified state dissemination mechanism to simultaneously support routing, multicasting, and most QoS heuristics.**

***Index Terms*—Ad hoc network, channelization, collision resolution signaling (CRS), mobile ad hoc network (MANET), medium access control (MAC), multicast, node state routing (NSR), prioritization, propagation map, quality-of-service (QoS), resource reservation, synchronous collision resolution (SCR), synchronous collision resolution, traffic engineering, wormhole.**

I. INTRODUCTION

MOBILE ad hoc networks (MANETs) consist of wireless nodes that communicate with each other by cooperatively sharing a common wireless medium. These networks operate without infrastructure and self organize to create and maintain a topology. The most commonly cited applications are military, emergency relief and sensor networks, which are driven to ad hoc networking because of the unavailability of infrastructure. In most applications, and especially if ad hoc networks are to be commercially viable, quality-of-service (QoS) is necessary. Numerous mechanisms across the protocol stack affect QoS delivery. Queuing disciplines, admission control policies, resource provisioning, and physical layer adaptation can all affect the perceived QoS, but their effectiveness is very sensitive to the operational scenario (i.e., environment, node movement, and traffic loading) and their appropriateness is dependent on the application (i.e., the types of QoS differentiation required). Soft mechanisms that attempt to measure the state of the network and then admit and route traffic to where resources appear available, or hold resources in anticipation of their use, are often emphasized in lieu of hard mechanisms that arbitrate the use of resources in real time. However, soft techniques are the most sensitive to scenario and often fail on their own to provide sufficient service differentiation for applications. Meanwhile, proposed real-time resource arbitration mechanisms are ineffective in ad hoc networks on account of the medium access control schemes that are used.

It has been proposed that to achieve reliable QoS in ad hoc networks will require traffic engineering capabilities and to provide these capabilities will require the cooperation of three components: 1) a QoS capable medium access control (MAC) protocol; 2) a resource reservation scheme; and 3) a QoS routing protocol [1]. A significant impediment to creating these components and their cooperation is the current tendency in ad hoc network research to rely on concepts that were conceived with a wireline network in mind. The contribution of this paper is to provide an alternative set of concepts for access and routing based on a wireless perspective that enables these three components of QoS and their cooperation. The basic concepts we introduce would permit protocol designers to implement real-time resource arbitration that can differentiate any number of levels of service and provide an efficient, scalable, network state dissemination mechanism that supports not only routing, but also most other soft techniques for QoS. We provide simulation results that demonstrate that these mechanisms are very effective. Throughout this paper, we contrast our approach with current and ongoing work, making the case for reexamining some of the basic assumptions often taken for granted in designing MAC and routing protocols for ad hoc networking.

The protocols we present are suited for most ad hoc networking applications except those with long turnaround times for exchanges such as interplanetary networks or low duty cycles such as some terrestrial sensor networks. The primary applications are military and emergency relief networks which require stream-based services together with prioritization and preemption capabilities. The access protocol assumes synchronization to a common clock with at least the synchronization

Manuscript received October 15, 2003; revised March 15, 2004.

J. A. Stine is with The MITRE Corporation, McLean, VA 20111 USA (e-mail: jstine@mitre.org).

G. de Veciana is with the Electrical and Computer Engineering Department, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: gustavo@ece.utexas.edu).

Digital Object Identifier 10.1109/JSAC.2004.829347

found in TDMA type protocols, however, better synchronization yields better performance. The routing protocol assumes location awareness and that the radios have the ability to measure the strength and quality of a signal. Although, not the subject of this paper, we believe that the implementation of both our access and routing concepts can be exploited to yield location awareness and synchronization, a requirement for most proposed military systems.

Our presentation of this material begins with two introductory sections. Section II, contrasts the wireline networking paradigm to what we will call the "wireless networking paradigm." Section III reviews the state of the art in QoS mechanisms in ad hoc networks and explains their limitations. Then, in Sections IV and V, we present two concepts that we believe are key to understanding and designing ad hoc networks according to the wireless networking paradigm. We propose that synchronous collision resolution (SCR)[1] be used for medium access control and that node state routing (NSR) be used for routing. Both concepts yield a family of possible protocol designs. In these sections, we demonstrate how these concepts provide easy to understand mechanisms to enable the three QoS components and their cooperation. In Section VI, we describe how networks built using these concepts can be integrated with wireline networks.

II. COMPARISON OF THE WIRELINE AND WIRELESS NETWORKING PARADIGMS

Wireline networks are built using the paradigm of routers and links. The critical resources of a router are its buffers and CPU time and the critical resource of a link is its capacity. The router contains the logic to control its buffers and the use of capacity on any of its outgoing links and provides a capability to recover from failed links. This logic seeks the maximum use of resources without compromising performance. In the design of this logic, signaling, and information exchange is leveraged to reduce computational complexity since link capacity is the cheaper of the resources. When applied to a wireless network, the wireline paradigm is slightly modified to acknowledge that links may be time varying, i.e., they come and go based on the movement of nodes. Thus, the logic in the routers is expanded to include more advanced protocols that can respond to a changing topology and, thus, more signaling and information exchange.

The wireline networking paradigm above fails to adequately describe wireless networks on two accounts. First, in wireless networks, media capacity is constrained by the radio frequency (RF) spectrum that can be made available and this capacity is significantly less than that available for wireline networks, and so the chatty protocols of wireline networks are undesirable. Increased computational complexity of protocols is acceptable if it can reduce the loading on the RF media. Second, the wireline paradigm oversimplifies the nature of wireless links. Its abstraction of wireless links as having two states, available and not available, and being independent of each other is too crude. In wireless networks, links are not physical entities but states. The state of a link between two nodes cannot exist without the cooperation of their neighbors. All nodes that could interfere with an exchange between a pair of nodes must defer from transmitting during that exchange rendering all of the links involving the deferring nodes unavailable. We see that the resource of interest is not a link but a wireless channel in a geographic space.

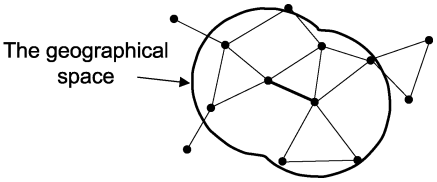


Fig. 1. Example of a channel space required for a connection between two nodes. The extent of the channel space determines the membership of the distributed queue, in this case all nodes.

Several observations form our wireless paradigm.

1) The critical resource is electromagnetic spectrum in a geographical space.
2) No single node controls the resource.
3) There is a distributed queue for each region in space formed by the queues of those nodes that share the electromagnetic spectrum in that region.
4) The geographical spaces overlap.
5) Nodes participate in several spatially distributed queues.

Fig. 1 provides an example scenario that illustrates these observations. The figure illustrates a connection between two nodes (thick line) and the geographic space that is consumed by an exchange between the two. This space is uniquely associated with this pair. The figure also illustrates the potential connectivity between all pairs of nodes (thin lines) in the network. All nodes in this example participate in the distributed queue of the connected nodes due to either being within the channel space or having destinations that are within the channel space. This scenario is made more complex when one considers the role of the physical layer. Physical layer characteristics can be used to enable more than this single link to exist in this example.[2]

Paradigm has a profound influence on research and development. The use of the wireline paradigm for ad hoc networks, with its discrete view of the nature of links, has bifurcated research. Research and development at the link and physical layers tend to perpetuate the abstraction of discrete links. Routing and other higher layer protocols use this abstraction to handle variation of topology and problems of QoS. The deficiency of this paradigm is that the true nature of the wireless environment is hidden from higher layer protocol mechanisms and that it perpetuates the need for high overhead protocols. Higher layer protocols usually assume they can control how traffic is offered to the communications medium. However, resource contention and the volatile nature of the wireless medium make it necessary to use buffers at the link layer. The higher layer protocols can only control how traffic is offered to these buffers. A suggested improvement is to allow cross-layer communications but this often falls short since access to the medium is distributed. To enable QoS requires MAC and physical layer mechanisms not only to isolate connections but also to arbitrate access based on the contents of the collective buffers. This requirement puts the

---

[1] Patent pending.

[2] An example would be channelization, where spreading code or frequency can be used to distinguish transmissions.

onus on the physical and link layers to solve the problems of ad hoc networking. Without this capability, higher layer protocols can only regulate the admission and routing of traffic and QoS becomes dependent on those protocols maintaining a condition, where media capacity is overprovisioned so that the adverse effects of congestion and access contention can be avoided.

Understanding and developing an intuition for the interactions involved in ad hoc networking using this wireless paradigm is difficult. Fortunately, the structure and mechanisms of the SCR MAC and Node State Routing protocols we discuss in this paper provide an intuitive framework to understand and design wireless networks.

III. QoS IN AD HOC NETWORKS

QoS is a broadly used term referring to network capabilities that result in user satisfaction. As broad as this definition is, so too are the proposals on how to deliver QoS in ad hoc networks. These methods include call admission protocols that first assess whether a flow should be permitted into the network [2]–[4], data rate adaptive protocols that adjust source coding to achieve a data rate that can be supported by the network [5], [6], routing protocols that find routes efficiently [7]–[12], routing protocols that search for long lived routes [13], [14], routing schemes that attempt to control the flow of traffic through regions in the network that can best support it (a.k.a. load balancing) [15]–[19], multipath routing to decrease the average delay of multi-packet messages and of streams [20], reservation schemes that attempt to reserve the transmission time at routers [21], [22], queueing disciplines implemented at nodes [23], access schemes that attempt to prioritize access to and reserve the RF media [24]–[28], access rate control to bound access delay [4], and physical layer adaptation protocols that adjust forward error correction rates [6], data transmission rates [29], [30] and/or transmission power [2], [30] to make links more reliable. A large list to be sure, but without access schemes to prioritize and reserve the RF media and to arbitrate its use among neighbors, these schemes are soft. That is, QoS is delivered by keeping the RF media overprovisioned. In our review of state of the art in QoS delivery mechanisms, we emphasize the MAC protocols that prioritize and reserve the use of the RF media. We follow with an overview of soft methods with the goal of identifying the common components and how they cooperate with MAC protocols.

A. Priority Access (PA)

PA, also referred to as differential service, is the simplest QoS mechanism. Queued packets at nodes are sent in order of their priority. Whereas differentiating service at a single node for a single resource is quite simple, arbitrating such access amongst a set of nodes vying for the same resource is not. The challenge in ad hoc networking is to make an access mechanism that causes contending nodes to defer to other contending nodes with higher priority (HP) packets. Most PA mechanisms fail in ad hoc networks because of hidden node effects or because they lack a common timing reference.

Three approaches have been proposed for differentiating service within contention access protocols: using different backoff or persistence parameters [24], sharing information [25], and using signaling [26], [27], [31]. Using different backoff or persistence parameters is a soft mechanism that gives nodes with HP packets more opportunities to gain access than nodes with lower priority (LP) packets; however, it does not cause nodes to defer to each other. In the second approach, nodes exchange information about the packets they have queued so they can learn when they should defer to neighboring nodes that have HP packets. The reliability of this mechanism depends on whether nodes can keep track of state information in volatile environments. In the third approach, a signaling tournament is used to arbitrate access. The basic idea is that, starting from a timing reference, nodes with the highest priority packets can signal first. These signals preempt nodes with LP packets such that only nodes with the same priority packet contend with each other for final access. Signaling mechanisms are challenged in multihop asynchronous access environments because of this requirement for a common timing reference.

We shall not discuss schemes that have been proposed for PA in wireless local area networks (LANs) that are inappropriate for ad hoc networks, e.g., schemes that rely on the availability of a central base station. Examples are polling schemes used by the point coordination function of the IEEE 802.11 MAC [32], [33] and various schemes used by satellites. Mechanisms built upon carrier sense multiple-access (CSMA) protocols that require all nodes to hear each other in order to be effective (i.e., have no protection against hidden terminal effects) are also not appropriate. Examples are the blackburst scheme[3] [31] proposed for use with the IEEE 802.11 distributed coordination function (DCF), the deterministic adaptive priority network access delay scheme [34] used by the Single Channel Ground Airborne Radio System, and the elimination yield nonpreemptive multiple-access (EYNPMA) scheme of the ETSI HIPERLAN I standard [27].

Some proposed modifications to the DCF of the IEEE 802.11 MAC [32] are more suitable for ad hoc networks and are examples of the three approaches listed above. The IEEE 802.11e protocol [24] is an example of using different backoff parameters. Packets are queued at nodes according to their priority with each queue having its own interframe space (IFS) for clear channel assessment and its own backoff contention window. Interframe spaces and backoff windows are longer for LP queues. The result is that differential service is created by disadvantaging LP packets (i.e., the poorer performance persists even when there are no HP packets). This mechanism does not guarantee that a HP packet will be sent before a LP packet, even at the same node, only that it will spend less time on average backing off.

The distributed wireless ordering protocol (DWOP) proposed in [25] piggybacks priority information on queued packets onto existing handshake messages of the 802.11 DCF. Each contending node uses request to send (RTS) and protocol data unit (PDU) packets to announce its head-of-line queued packet. Destinations echo these announcements in clear-to-send (CTS) and acknowledgment (ACK) packets. Neighbors monitor these transmissions and keep a table of these times. A node defers

---

[3] Although the DCF provides hidden terminal protections the blackburst is only heard local to the source and does not insure precedence at the destination.

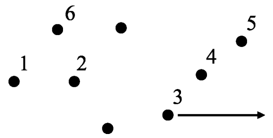


Fig. 2. Example failure scenario of WNOP. Node 3 has an entry for Node 1's head-of-line packet that it learned through one of Node 2's handshake responses. Deadlock may occur if Node 1 sends that packet without Node 3 learning the new state. This can occur in several ways. Node 1 may send the packet to another destination, say Node 6, where neither end is in range of Node 3. A second source, say Node 4, may transmit when the update is sent interfering with Node 3's reception of the changed state. Finally, Node 3 may move out of range before it learns the new state.

Fig. 3. Example failure scenario of BTPS. Node 1 has HP packets intended for Node 2. Node 4 has LP packets intended for Node 3. The four nodes can only hear their adjacent neighbors' transmissions. If Node 4 gains access first then it can suppress the HP stream from Node 1. As part of Node 4's contention, Node 3 sends a CTS that mutes Node 2. When Node 1 contends, Node 2 cannot respond. Node 1 considers the contention a failure, increases its contention window and contends again. The busy tone scheme cannot affect Node 4 and so Node 1 can only seize the channel if by chance its contention window expires between Node 4's transmissions.

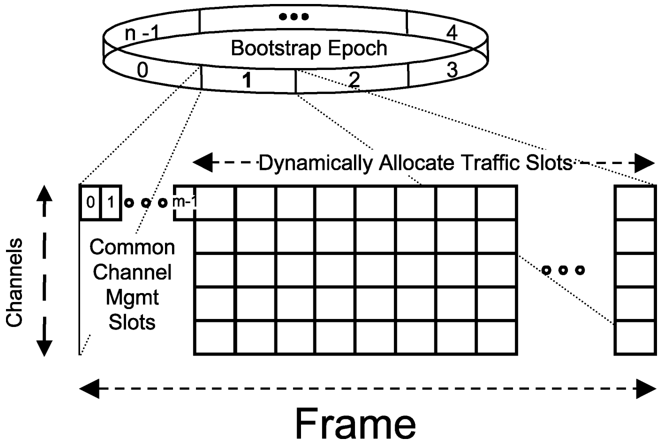


Fig. 4. Frame and time slot organization of the USAP.

from contention as long as a time on its table precedes the arrival time of its own head-of-line queued packet. This protocol supports ordered transmission as long as most nodes have a backlog of packets and nodes in the network can keep track of the network state. The latter requirement is most critical. Losing state information can result in deadlock, where nodes continuously think another has a HP packet. Fig. 2 illustrates and describes such a scenario.

The busy tone priority scheduling (BTPS) [26] protocol adds a signaling mechanism and an additional interframe space to the 802.11 DCF to create two levels of PA. Nodes with LP packets use the longer IFS. This space is sized such that nodes with HP packets can use signals to continuously reset the IFS used by the nodes contending to send LP packets. To prevent hidden terminal problems, two separate busy tones are used for signaling, one by the contending station, busy tone 1 (BT1), and a second for echoes, busy tone 2 (BT2). A HP contender sends a BT1 periodically during its backoff. Nodes that hear a BT1 signal echo a BT2 signal. Thus, a HP contender can interact with contenders up to two hops away and in most cases will have precedence over LP contenders. Unfortunately, the asynchronous nature of the DCF and virtual sensing prevent BTPS from guaranteeing precedence to HP contenders in ad hoc environments. In order for the different IFS's to work, all contenders must have a common perception of the start of the IFS. Since an LP contender could use a different end-of-transmission to trigger the start of its IFS, the LP contender could backoff in time to beat an HP contender. A more severe problem may occur as a result of virtual sensing. Fig. 3 illustrates an example of this type of scenario.

We have presented three asynchronous MAC protocols that have been proposed for PA and have described how they might fail. One cause of failure can be traced to the lack of a common timing reference in the execution of access arbitration. An alternative that makes a common timing reference more feasible is to use a synchronous access protocol. Other than our own work, the synchronous access mechanisms that we are aware revolve around time-division multiple-access/frequency-division multiple-access (TDMA/FDMA) and consist of scheduling slots and channels for pairs of nodes to use. This is a resource reservation strategy as opposed to a PA contention strategy and we will discuss their effectiveness in the next section.

From this review, we see that asynchronous PA access schemes in ad hoc networks are vulnerable to scenario effects that can render them ineffective. They also provide a limited number of priority distinctions. In contrast, the contribution of our proposal is a synchronous contention access mechanism that can be designed to provide any number of priority levels. The reliability of this PA is not affected by the scenario.

B. *Resource Reservation*

Resource reservation is a critical component of any mechanism that delivers QoS in environments where resources cannot be overprovisioned. The typical view of resource reservation in wireline networks is to reserve capacity along a route. QoS protocols such as the resource reservation protocol (RSVP) [35] and integrated service (IntServ) [36] protocol lie above the Internet protocol (IP) in the protocol stack and are predicated on the assumption that a router can control all of the capacity on an outgoing link. Their objective is to reserve capacity across multiple links. As described in Section II, this underlying assumption is not appropriate for wireless environments since the RF media capacity available to any relay node is contingent on that used by its neighbors and its use must be arbitrated.

The problem of link reservation is sufficiently daunting that designers have compromised on efficient use of capacity and have proposed solutions that attempt to mimic the wireline paradigm by either scheduling or assigning channels to pairs of nodes to create isolated links. The unified slot assignment protocol (USAP) [37] is an example of a TDMA scheduling protocol for ad hoc networks. It has been implemented in some experimental systems [38] and is currently the basis of access protocols proposed for future military communications systems.

Fig. 4 illustrates the structure of USAP as it is used to manage multiple channels. Note that there are m bootstrap slots per frame and n frames per epoch which allow $m \times n$ nodes within two hops of each other to broadcast schedules. Broadcasted schedules use bitmaps, where each position corresponds to one of the bootstrap slots. Nodes become identified with the bootstrap slot they use to transmit their schedules. A node joins the network by listening to the bootstrap packets and then selecting and transmitting its own schedule in a bootstrap slot it perceives is not being used. The schedules consist of three components: the slots, channels, and destinations of the node's transmissions; the slots and channels wherein the node receives packets; and, the slots and channels in which the node's one-hop neighbors transmit. Recipients use this information to identify the channels and slots they may use. However, a node's proposal to use a particular channel and slot may be rejected if a hidden node proposes the same reservation. In a static network, a slot remains reserved so long as the reserving node claims it. In a dynamic network, movement of nodes can require nodes to change their broadcast slot alignment and their schedules. As an example of the complications, consider what happens when a node with a schedule moves into an advantaged position, where it is in range of more nodes. It is possible more than one of its neighbors will be using the same bootstrap slot so there will be ambiguity as to who owns the bitmap positions. Also, reservations of its new neighbors may be in conflict with those of its old neighbors. This will affect every one of the advantaged node's neighbors, and then their neighbors, as scheduling attempts to achieve a two-hop separation of channel-slot reservations. These problems may be especially acute if some nodes in the network are airborne and relatively fast moving.

Scheduling approaches make PA a higher layer protocol task. Given that channel-slots are reserved between two nodes, the higher layer protocol arbitrates which packets are sent using them. What this does not do is arbitrate the media use based on priority. The prioritization of media use is determined by how channels and time slots are allocated among pairs of nodes. Preferably, nodes with more HP traffic should be able to reserve more channel-slots. As described, USAP uses a greedy algorithm to assign channels and slots and there is no arbitration on which nodes should have precedence. Responsive service will depend on whether channel-slots are reserved *a priori* or the scheduling algorithm can react quickly. There is no mechanism to preempt channel-slot reservations used by nodes to send LP traffic by neighboring nodes wanting to send queued HP traffic. With the exception of the ability to create a reservation to support streams, QoS delivery remains soft and, thus, dependent on the overprovisioning of the RF media.

As described, the effectiveness of reservation schemes, as epitomized by USAP, are highly dependent on the operational scenario and they do not arbitrate the prioritization of RF media use. In contrast, the contribution of our proposal is a scheme to reserve resources based on prioritizing RF media access. This allows protocol designers to create multiple levels of reservation and to enable preemption. The mechanism is less sensitive to the operational scenario since the reservation schedule is implicit and does not require all nodes in a two-hop region to agree.

C. QoS Routing and Other Higher Layer Protocols

QoS protocols at higher layers are soft techniques that generally consist of three components: a measure of network state and a way to observe it, a method to collect or distribute the state information, and finally, a heuristic that uses the information to deliver a QoS objective. Table I lists several QoS protocols and describes these components. This is an incomplete list as this is a well researched topic. Its purpose is to show that these three components are fairly universal and that there is a large variety of ways QoS can be enhanced. Rather than review and compare the various approaches our focus will be on how state information is measured and disseminated. Given the necessary state information many approaches are possible; however, in ad hoc networks, the measurement and collection of state information may itself have significant impact on QoS delivery. There is a delicate balance between the age of state information used by network protocols and the quantity of network capacity consumed in measuring and disseminating it. Finding and achieving this balance is the focus of much research. For example, the argument that separates the relative merits of using a reactive versus a proactive routing protocol is whether less overhead is more important than faster connectivity, a question that can only be answered in the context of the operational scenario and network application.

Although it is the dynamic nature of ad hoc networks that is most to blame for the overhead required to track network state, artifacts of the wireline paradigm also bear much blame. Two are most significant, the abstraction of topology and state information into link abstractions and the separation of protocols at the higher layer which cause multiple independent processes to measure and disseminate state information, sometimes redundantly. In ad hoc networks, the link abstraction confounds information about the environmental state (e.g., obstructions and local interference), the network state (e.g., node and traffic congestion), and the end nodes' states (e.g., proximity, mobility, energy reserves, and radio and antenna capabilities) into a pairwise state. Any of these factors can affect the state of the link and in proactive routing protocols frequent assessment and in reactive routing protocols recent assessment are necessary. Additionally, the loss of a link is ambiguous as the failure to observe a link does not identify its cause. The loss can be as much the effect of RF media unavailability as it is the physical inability of two nodes to talk to each other now or in the future. From Table I, we see that the goal in many MANET QoS protocols is to collect and to disseminate this same state information but from a nodal view. This seems natural as the equivalent protocols in wireline networks do the same thing. Unfortunately, the state information is not easily extracted from link state metrics unless they have been specifically designed with this in mind. This results in a plethora of ideas of how to measure and disseminate state information for QoS purposes. Either the network designer uses a routing protocol that commits to a set of QoS objectives or multiple protocols are applied that each implement state collection and dissemination mechanisms. Call admission, routing, multicasting, resource reservation, and network management are all examples. Returning to one of our original points that capacity in ad hoc networks is constrained and overhead should be avoided, the wireline paradigm that drives this redundancy can make QoS implementation counterproductive.

Our paper's second contribution is to provide a novel routing protocol that provides a unified mechanism to capture and

TABLE I
SOFT-STATE PROTOCOLS FOR QOS

Protocol	**Function**	**State Measure**	**Distribution Mechanism**	**Heuristic**
SWAN [4]	Rate and admission control	For rate control nodes measure packet access delay at the MAC layer. For call admission nodes measure the use of the channel.	Local measurements at the node are used to regulate the sending of packets locally Probing packets are sent to the destination prior to call admission to assess bandwidth availability. Fields in the header of packets is used for dynamic regulation	The source sends a probing packet toward the destination. Intermediate nodes reset a bottleneck bandwidth field if they cannot provide the bandwidth specified in the field. The destination sends the bottleneck bandwidth information to the source which admits the session if this bandwidth is sufficient. Regulation occurs through continued assessment of bandwidth. Intermediate nodes reset bits in the headers if they cannot support the flow. Destination nodes signal the source when this occurs and session setup starts over.
dRSVP [51]	Resource reservation	Individual nodes assess the quantity of bandwidth they have available. No details on how.	Mimics RSVP except bandwidth requests are expressed in ranges.	Uses three sets of messages: a path message forwarded from the source to the destination specifying the acceptable range of bandwidth, a Resv from the destination back to the source specifying the max capacity of the downstream bottleneck, and then a Resv/Notify message from the source to the destinations specifying the max capacity of the upstream bottleneck. Reservations are made if the bottleneck capacity falls within the requested range. Intermediate nodes reserve bandwidth based on the bottleneck capacities, both upstream and downstream.
INSIGNIA [22]	Resource reservation	Intermediate nodes make an assessment of the available bandwidth.	In-band signaling built into an IP header.	Sends packets toward the destination and each node along the way assesses if it can reserve the bandwidth and forwards the packet and its reservation status toward the destination. Destination reports back to the source on the status of the reservation. Intermediate nodes release resources when they stop receiving packets for the reservation.
BGCA [18]	Routing	Intermediate nodes make an assessment of the available bandwidth.	On-demand route discovery using broadcast route request (RREQ) packets	Intermediate nodes process and forward RREQ only if they have the bandwidth to support the request. The destination takes all received route request packets and selects the shortest route.
Modified AODV [19]	Routing	Assumes a TDMA access mechanism so nodes measure bandwidth in TDMA slots available and assures adjacent nodes do not commit the same slots.	On-demand route discovery using broadcast route request (RREQ) packets	Intermediate nodes process and forward RREQ only if they have the bandwidth to support the request. The destination selects the route of the first RREQ received.
LRR [30], [17]	Routing	Each node assigns a link resistance measure to links that accounts for energy to close the link, the available capacity, and past reliability. Different measures are associated with different traffic types.	Uses a distance vector routing algorithm	Seeks the least resistance path. Protocol assumes that sessions per link can be moved to an isolated channel without interference from neighboring nodes. Prioritizes access for packets at nodes based on service type.

SWAN – Stateless Wireless Ad Hoc Networks, dRSVP – dynamic RSVP, INSIGNIA – a name not an acronym, BGCA – Bandwidth Guarded Channel Adaptive Routing, AODV – Ad Hoc On Demand Vector Routing, LRR – Least Resistance Routing

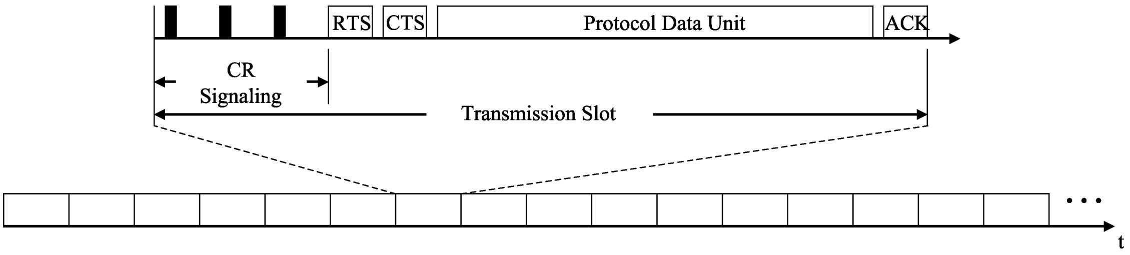


Fig. 5. Basic implementation of the SCR MAC protocol.

disseminate state information that can support not only basic routing but also most QoS heuristics.

D. Component Cooperation

Providing a complete QoS solution for ad hoc networks requires the cooperative interaction of PA, resource reservation, and QoS routing components [1]. The motivation for wanting this cooperation is that it enables the use of traffic engineering techniques to deliver QoS rather than just the use of resource overprovisioning. Creating cooperation requires that these protocols be able to create a view of the state of the network which captures the use of all resources and that these protocols provide mechanisms that can provision those resources. The contribution of our proposal is a paradigm for protocol development that does just this.

IV. SCR MAC PROTOCOL

SCR is a broad MAC definition and is best viewed as an access framework in which there are many possible designs. SCR is illustrated in Fig. 5 and has four key characteristics.

1) The wireless channel is slotted.
2) All nodes with packets to transmit attempt to gain access every transmission slot.
3) Contending nodes use signaling to arbitrate their access.
4) All packet transmissions that occur during a transmission slot are sent simultaneously.

Design choices that determine capabilities of SCR are the size and framing of transmission slots, the use of handshake packets, and the specific details of signaling.

SCR's characteristics make it well suited for multihop wireless environments. The synchronizing of access attempts and

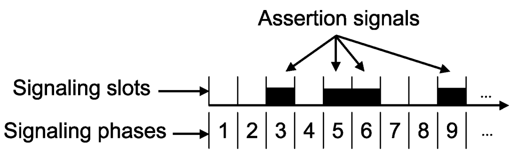


Fig. 6. CRS using single-slot phases.

the use of an interactive contention arbitration mechanism, collision resolution signaling (CRS), enable SCR to seek out a "preferred" collection of nodes to exchange traffic at the beginning of each transmission slot. At the conclusion of the signaling, the set of exchanging nodes is frozen without risk of mid-transmission collisions (e.g., hidden terminal collisions.) Just as the pairwise channel spaces overlap each other based on the propagation range of transmissions, so do the effects of signals, a fact that allows CRS to resolve a set of surviving source-destination (SD) pairs whose channel spaces can coexist. The definition of the "preferred" collection of nodes that CRS arbitrates is dependent on how the signaling mechanism is designed. At a minimum, CRS arrives at a relatively dense set of nodes that can exchange traffic simultaneously. Better, CRS may be designed to arbitrate access giving preference to nodes with high priority packets, thus enabling a queueing discipline across the distributed queues of an ad hoc network.

We provide a brief overview of how SCR resolves contention locally and spatially. A more detailed discussion of this topic can be found in [39]. We then introduce the four optional modifications that may be made to CRS that support QoS. These include a source directed echoing feature that can extend the range of CRS's effect and mechanisms that can be used to provide prioritized access, resource reservation, and channelization. We conclude Section IV with a discussion of issues and options the protocol designer should consider in tailoring CRS for a network's application.

A. Access Arbitration

Access arbitration consists of CRS and, optionally, an RTS-CTS handshake. CRS selects a subset of contenders that are good candidates for sending packets at the same time. The RTS-CTS handshake reduces the SD pairs to those that can exchange packets simultaneously.

Collision resolution signaling consists of a series of signaling slots organized into groups of slots called phases in which contending nodes may send very short signals. These signaling slots should not be confused with the longer transmission slots of Fig. 5. Rather, they occur within a transmission slot during a short period at the very beginning. There are numerous ways to design signaling. The simplest and generally most effective at arbitrating contention is illustrated in Fig. 6, and consists of one signaling slot per phase. In this design, a probability is assigned to each signaling slot and a contending node will signal in that slot with that probability. There are two assumptions that apply to signals and signaling slots.

1) Signals superimpose such that a receiver that hears multiple signals will still detect a signal.
2) Signaling slots and signals are sized to account for synchronization accuracy, propagation delay for the maximum range, detection time, and receive-to-transmit transition time such that the slot in which a transmitter sends and a receiver detects the signal is unambiguous. (See [39].)

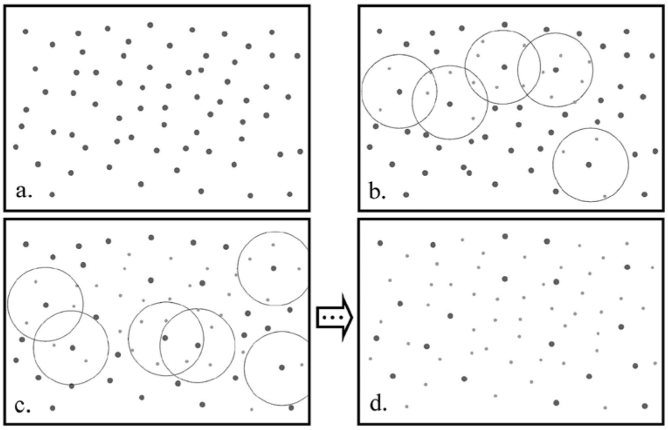


Fig. 7. Example of CRS: All nodes start off as contenders in (a). Then, through a series of signals, two sets of which are illustrated in (b) and (c), a final subset of contenders is selected in (d). The large dots are nodes that view themselves as contenders, the small dots are nodes that view themselves as having lost the contention, and the large circles represent the range of the signals. Contenders defer when they hear the signal of another contender.

The rules of signaling in this design are as follows.

1) At the beginning of each signaling phase a contending node determines if it will signal. It will signal with the probability assigned to the slot of that phase.
2) A contender survives a phase by signaling in a slot or by not signaling and not hearing another contender's signal. A contender that does not signal and hears another contender's signal loses the contention and defers from contending any further in that transmission slot.
3) Nodes that survive all phases win the contention.

Appendix A describes a design algorithm for selecting the probabilities for signaling slots in this type of mechanism.

The performance of CRS can be measured in two ways, how well does it resolve contentions locally and how well does it separate survivors spatially. CRS's ability to resolve contentions locally depends on the number of signaling phases used and the assignment of probabilities to the signaling slots of those phases. Using the design algorithm presented in Appendix A, nine-phases of signaling can be made $>99\%$ effective at resolving contention to just one survivor with as many as 450 nodes contending for access in range of each other. In multihop environments, the synchronized implementation of CRS (i.e., SCR) spatially separates survivors such that the probability that a survivor is in range of another is equivalent to the signaling design's contention resolution probability. Fig. 7 uses a series of panels to illustrate how SCR creates spatial separation among survivors.

At the conclusion of signaling, surviving contenders are separated but this is not necessarily true for their destinations where the concern about interference will be. The RTS-CTS exchange mitigates this concern. Fig. 8 illustrates the RTS-CTS exchange. As demonstrated, the role of the RTS-CTS exchange is neither to

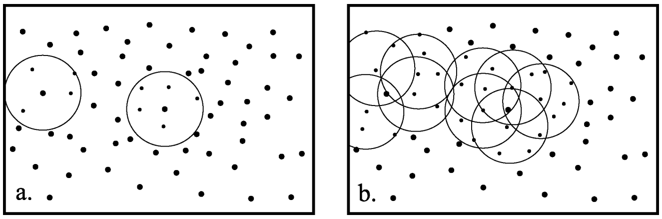


Fig. 9. Illustration of echoing: in (a), two nodes send assertion signals, and then in (b) those nodes that hear an assertion signal echo. All contenders within range of the assertion signal and the echoes that did not transmit an assertion signal of their own defer from contending.

limit collisions to smaller packets nor to extend channel use detection two hops for hidden terminal protection as is its purpose in the 802.11 MAC protocol. Rather, the RTS-CTS exchange verifies that SD pairs can "close" a connection and provides a feedback mechanism to support link adaptation. The following observations highlight the role the RTS-CTS exchange plays.

1) The RTS and CTS packet transmissions occur in the worst-case mutual interference. Their successful exchange is a good indication that they will also be able to exchange the subsequent PDU and ACK. Contenders that do not receive a CTS would be unlikely to exchange a PDU successfully and defer from further transmissions improving the mutual interference conditions for the remaining nodes.
2) A destination can assess the quality of an RTS packet and then use its CTS packet to signal the source to adjust one or more of its transmission parameters (e.g., data rate, FEC rate, transmission power) and, similarly, a source can assess the quality of a CTS packet and signal the destination in the PDU packet. There is one constraint on these changes, they must not increase interference at other receivers (e.g., power may be decreased not increased).
3) RTS-CTS exchanges enable energy conservation. At the conclusion of these exchanges, nodes not participating can enter a low-energy state for the remainder of the transmission slot and those that do participate can reduce their transmission powers to a level deemed feasible from the measurement of the RTS and CTS quality [40], [41].

It is possible that two nodes that are out of range of each other may contend to send a packet to the same node. In cases where neither capture nor other contenders can resolve this conflict, a blocking condition may occur. Blocking can be resolved by a simple signaling technique called echoing. Signaling phases are designed with two slots. Contenders signal in the first as described earlier but then neighbors who hear the signals echo them in the second slot. Fig. 9 illustrates the process. When echoing is used, a node survives a signaling phase by either signaling in the phase or by hearing neither a signal nor an echo. Signaling can be designed to use echoing all the time, to use echoing in a subset of the signaling phases, or to conditionally execute echoing. See Section IV-C for a discussion of the tradeoffs. Here, we propose a design that supports conditional echoing. In Fig. 10, we illustrate that we add an additional signaling slot to a nine slot signaling design that we call the echo invoke (EI) slot. The signaling proceeds as a normal nine-phase design until a node sees a need to invoke echoing (e.g., a node repeatedly survives CRS but cannot "close" an SD connection.) At that time, the contending node signals in the EI slot converting CRS for those nodes that hear the EI signal to a five-phase design with two slots in each phase. We selected the signaling probabilities for the four contention phases using the design algorithm in Appendix A with a design density k_t of four nodes. The resulting design can resolve blocks amongst 2–4 nodes with better than 90% probability in one contention and better than 99% probability after two contentions.

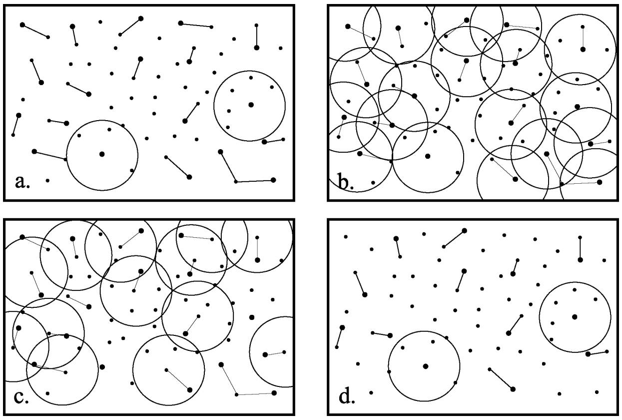


Fig. 8. Example of the RTS-CTS handshake finalizing the set of nodes to exchange packets: (a) illustrates the set of contenders that survived signaling and their intended destinations. The lines indicate SD pairs and large circles indicate nodes that are broadcasting. Contenders send RTS packets simultaneously. The large circles in (b) illustrate the ranges of these RTS transmissions. If a destination receives an RTS packet, it responds with a CTS packet, as in (c). These CTS packets are also sent simultaneously. Recipients of the RTS packets for broadcasts do not respond. In (d), all broadcasting nodes and those nodes that have received a CTS from their destination transmit PDUs.

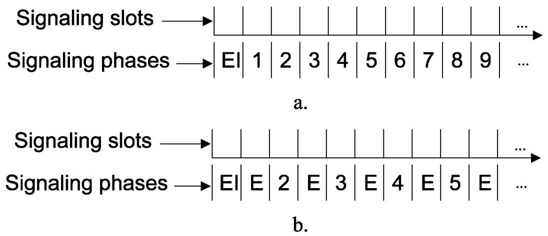


Fig. 10. Signaling design to selectively use echoing: in most contentions, nodes use the signaling design shown in a. If the source detects a blocking condition, knows the source to be an exposed node, or wants to broadcast a packet, it may invoke echoing. If a node signals in the EI slot then that node and all of its neighbors use the echoing design of (b). (a) Normal signaling. (b) Signaling after echoing is invoked.

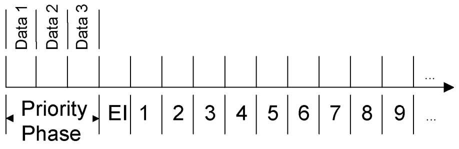


Fig. 11. Modified CRS for providing PA. This design provides four levels of priority, three levels, one associated with each slot in the priority phase and one level associated with not signaling in the priority phase.

B. Priority Access (PA)

PA is easily added to the CRS mechanism using a technique similar to that used in the EYNPMA protocol. [27] In Fig. 11,

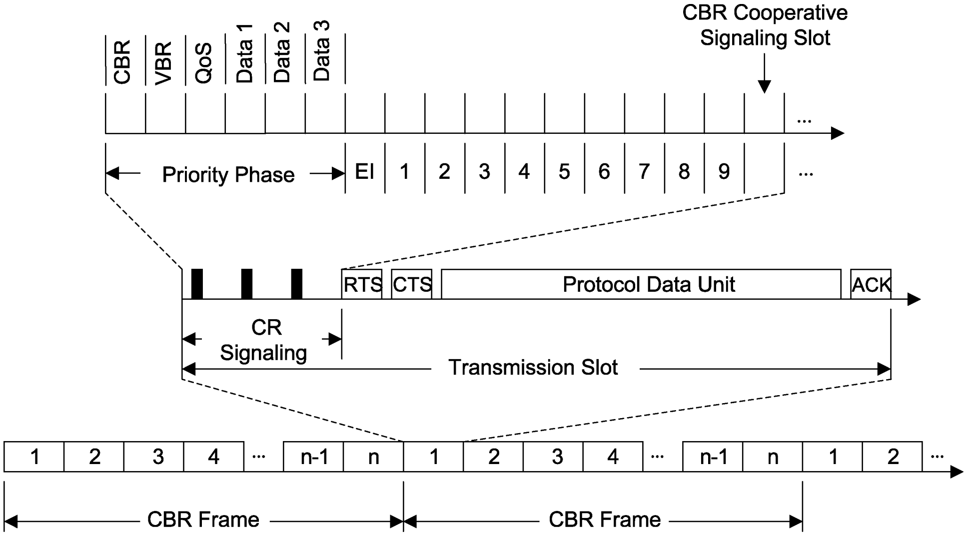


Fig. 12. Modified SCR for providing resource reservation. Reservations are a two step process. A node obtains a reservation by successfully contending using the QoS signaling slot. Then that node may use the CBR signaling slot in same ordinal slot of the subsequent CBR frames. The reservation lasts as long as the node uses the CBR slot. Destinations of reservations use the cooperative signaling slot to extend the reservation to the range of its signal. The VBR signaling slot is a best-effort priority slot to support bursty streams.

we add a multislot priority signaling phase to the front end of the CRS. Each slot in the phase is mapped to a different priority with highest priority first. Contenders use the slot that corresponds to the priority of the packet they are contending to send. If a node has a HP packet than its neighbors, it will signal first causing those neighbors to defer from contending. The remainder of CRS resolves the contention amongst nodes using the same priority. The priority phase of Fig. 11 provides four priorities for best-effort data packets. The first three are mapped to the Data 1–3 signaling slots and the fourth is not signaling at all.

C. *Resource Reservation*

Fig. 12 illustrates three modifications to CRS that enable resource reservation.

1) Three slots are added to the priority phase: a QoS slot for initiating a reservation, a constant bit rate (CBR) slot for holding a reservation, and a variable bit rate (VBR) slot for handling bursty streams.
2) The first signaling slot of a signaling phase is set aside for cooperative signaling.[4] This slot is used by the destination end of a reservation to preempt contenders within their range.[5]
3) The transmission slots are organized into frames. The size of the frame defines the period between slots of a reservation.

A node desiring to make a reservation contends using the QoS slot of the priority phase. If it wins the contention and successfully exchanges a packet, it may then use the CBR priority in the same ordinal slot of the subsequent frame and may continue to do so, so long as it used the CBR priority in the same ordinal slot of the previous frame. This two step method to earning the right to use the CBR priority for access prevents neighbors from gaining access and so effectively corresponds to a CBR reservation.

Destinations cooperate in these reservations. It is necessary to assure that no nodes within range of the destination will interfere with the destination's reception of the CBR packet. When a destination hears a CBR PA attempt it knows if it was the destination in the same ordinal slot of the previous frame. If so, it signals in the cooperative signaling slot.

CBR frames and transmission slots are sized to enable one packet per frame to support the minimum CBR data rate. If greater bandwidth is required, nodes may reserve multiple slots per CBR frame. Since some streamed traffic is bursty, we also enable a VBR priority. Nodes that are using the CBR priority to support a stream may use the VBR priority in a best-effort way to send additional packets of the same CBR stream. The VBR slot is optional and this same response to bursts on a stream can be handled by mapping the additional transmissions to one of the data priorities.

Reservations created with this mechanism are of the use-it or lose-it type. This type of reservation makes sense for CBR traffic which is expected to be regular. Since nodes with packets contend every transmission slot, the reserved slot is immediately available to another user when the node that held the reservation stops using it. The policy that is used to decide when to stop holding a reservation is a protocol design issue that will depend on the network application.

D. *Channelization*

Channelization is used to reduce mutual interference; however, in ad hoc networks, broadcasting on a common channel is necessary for the discovery of network topology. Any channelization scheme must enable nodes to broadcast on a common channel, but then to pull peer-to-peer transmissions to separate, preferably orthogonal, channels. This is challenging since destinations normally do not know on which channel to listen.

---

[4] Cooperative signaling may also be implemented by placing the cooperative signaling slot immediately after the CBR priority slot.

[5] Cooperative signaling is unnecessary when CRS is designed to use echoing all the time.

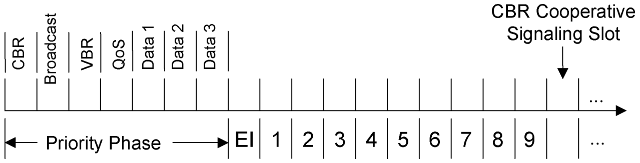


Fig. 13. Modified CRS to enable channelization. Nodes listen on broadcast channel when the broadcast slot is used to gain access, otherwise on the peer-to-peer receiver directed channel.

SCR uses receiver directed channelization. This means, in addition to a shared broadcast channel, all nodes will have a separate channel that they will use to receive peer-to-peer packets. Nodes broadcasting packets use the broadcast channel and nodes sending peer-to-peer packets use their destination's receive channel. We enable destinations to determine the channel to listen to through the addition of a broadcast signaling slot to the priority phase as illustrated in Fig. 13. This slot is used by nodes wanting to broadcast a packet. Not only does it provide a HP to broadcast packets over other best-effort packets it also serves to indicate to which channel a destination should listen. All nodes will know which priority was used to gain access at the conclusion of the CRS. Nodes that do not survive CRS listen to the broadcast channel if they hear the broadcast priority used, otherwise, they listen to their own peer-to-peer channel. Support for the selection and dissemination of receiver channels is provided by our NSR mechanism.

E. SCR Performance

We conducted several simulation experiments to characterize the performance of SCR's PA mechanism. We built a very accurate model of the SCR MAC and the physical layer using OPNET. See [41] for more details. We then executed two sets of experiments, one where all nodes are within range of each other and one in an ad hoc network configuration. The one-hop network consisted of forty nodes. Traffic arrivals at each of these nodes were Poisson, had the same rate, and were randomly and evenly distributed amongst four priority levels. Fig. 14 shows the performance as a function of total load. All priority levels have similar throughput and delay until the capacity of the channel is reached at which time the LP packets defer to the HP packets. The data rate (1 Mb/s) and sizing of packets (512 bytes) allowed 163 transmission slots per second. SCR successfully used 99% of these without congestion collapse. We repeated the experiment in an ad hoc configuration. We randomly placed 156 nodes on a square simulation surface seven transmission ranges on a side to achieve an average node degree of 10. We toroidally wrapped the surface[6] and implemented minimum hop routing.[7] A link was assumed possible if after path loss the received power would be 10 dB or more above the background noise without including interference from other contenders. We assume the physical layer characteristics of the 802.11 1 Mb/s DSSS waveform to include overheads, transition times, and the 10-dB processing gain. We loaded the network just as we did for the one-hop network except that we randomly selected a destination from amongst the 155 possible nodes for each packet arrival. We repeated the experiments for each or ten different randomly generated node placements. Fig. 15 exhibits the average performance of the ad hoc networks as a function of load. We see that the performance is similar to that of the one-hop network. The results demonstrate that at saturation the channel was reused about 15.5 times on average (i.e., 15.5 packets exchanged per transmission slot or in this scenario about 1 packet exchange per transmission slot per area covered by a transmission), which supported an end-to-end throughput of about 600 packets/s. Again, LP packets defer to HP packets but their delivery rate and latency are similar up until the offered load saturates the network. SCR provides effective use of channel capacity and nearly ideal access prioritization. SCR provides fair access in the sense that a node will have at least

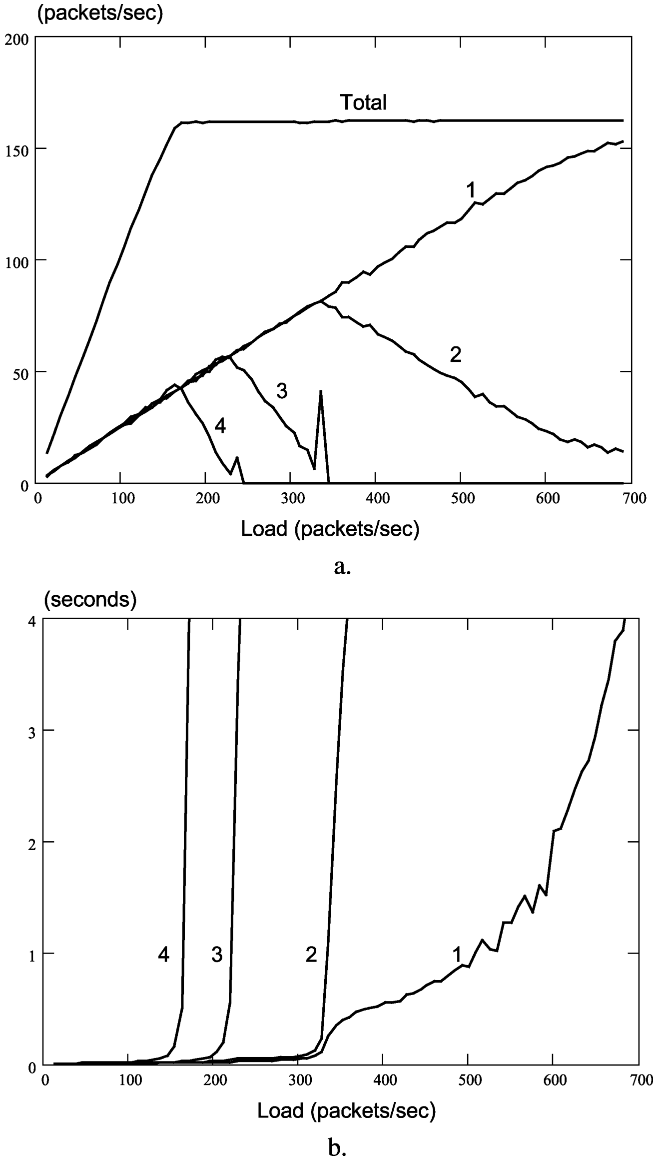


Fig. 14. Performance of the SCR PA mechanism in a single-hop environment. Demonstrates that SCR makes >99% of transmission slots usable despite amount of congestion and that HP packets do not suffer long delays as long as they do not single-handidly cause congestion. (a) Average throughput. (b) Average delay.

---

[6] The purpose of toroidally wrapping the surface is to eliminate edge effects.

[7] The router had perfect information about the topology and offered no traffic to the network.

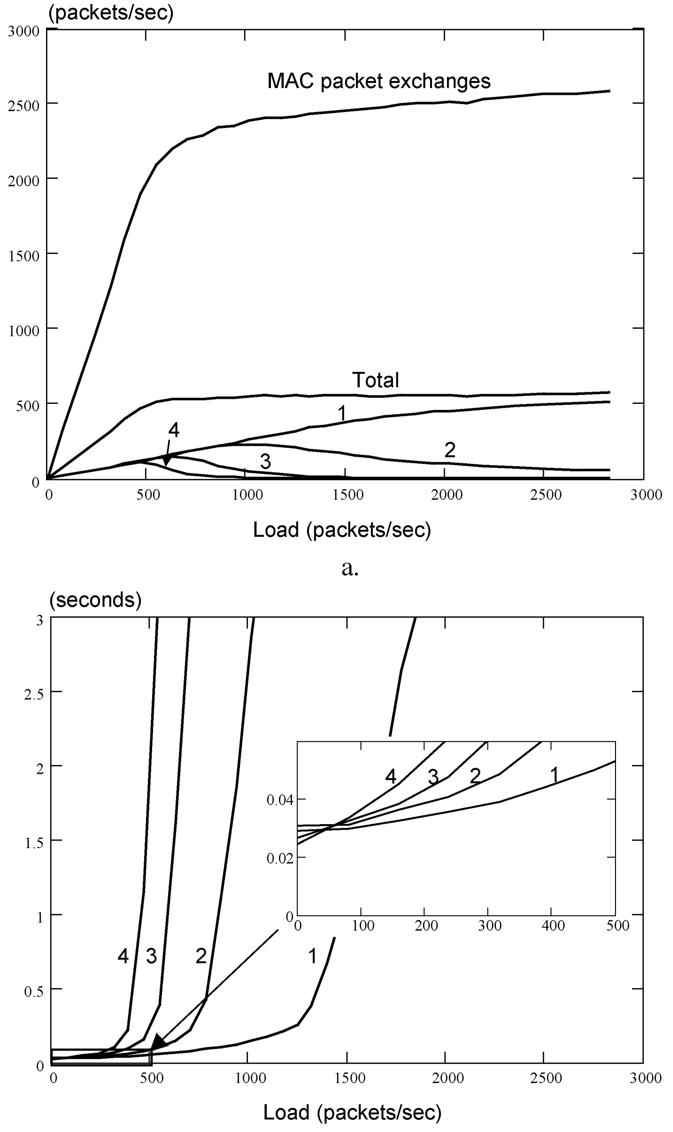
an equal opportunity to gain access as its contending neighbors if none of its neighbors are contending with a HP.

F. Design Considerations

The SCR design in Fig. 12 is meant to be a sample implementation. The preferred design of SCR will depend on physical layer capabilities and the network application. Physical layer capabilities will determine the feasible number of slots to use in signaling, the relative merits of using an RTS-CTS exchange in lieu of echoing or of using both, and whether channelization should be a goal. The network application will determine the number of priority and reservation levels, the policies used to map application traffic to these levels, and the policies used to set-up and to hold reservations. Table II lists physical layer capabilities and application issues that influence SCR design and describes the design choices.

A significant choice in SCR design is choosing the size of the transmission slot. This choice determines the maximum size of the PDU and the quantity of wasted capacity when packets are smaller than the PDU size. It determines the percentage of capacity that is consumed in signaling where the percentage decreases as the transmission slot gets larger. It determines the size and repetition rate of frames and the packetization delays for streamed data. Also, after the transmission slot size is selected, it is possible that packets may be too large to fit into a PDU and need to be fragmented. In our experimental implementations of SCR, we implemented a fragmentation and assembly mechanism that fragments packets at the source and reassembles them at their final destination. Packets may also be much smaller than the PDU. For this situation, we implemented a packet padding capability which our routing protocol uses to diffuse node states in the unused PDU space. There may be other strategies for fragmenting packets and using unused bandwidth.

A second set of design choices are the policies that govern when to invoke echoing and when to drop packets. Echoing should be invoked when blocking occurs, but detection of a block is ambiguous since the only symptom is the node successfully contends but cannot complete the RTS-CTS handshake. This is the same symptom of a routine collision or a temporary fade, which may repair before the next successful contention. Invoking echoing too quickly may reduce capacity unnecessarily. Packets should be dropped when the destination is no longer reachable. Again, this is detected by the repeated failure of the RTS-CTS handshake or not receiving an ACK in pure echoing designs. In our experimental implementation, we invoked echoing after three failed handshakes for the same packet and dropped that packet after three more failed handshakes after having invoked echoing.



a.

b.

Fig. 15. Performance of the SCR PA mechanism in a multihop environment. At capacity, MAC packet exchanges occur at a rate of ~1 exchange per transmission slot per transmission area. Average separation distance between end nodes is five to six hops, so end-to-end throughput is less. HP packets delivery is not affected by LP traffic. End-to-end delays do not vary significantly between traffic priorities until the channel capacity is reached. (a) Average end-to-end throughput. (b) Average end-to-end delay.

V. NODE STATE ROUTING (NSR)

Few research areas have a larger set of proposed solutions than routing in ad hoc networks. Classical taxonomies of routing divide protocols either into link state, distance vector or flood search algorithms. Taxonomies of routing in ad hoc networks further divide protocols into schemes that use either reactive or proactive topology discovery [7]. Due to the variability of topology in ad hoc networks the emphasis shifted from ways to calculate routes to developing innovative ways to discover and disseminate topology efficiently. Still, almost all proposals are based on the wireline networking paradigm. Connectivity in the network is understood based on the discrete pairwise link abstraction. In some way, pairs of nodes identify their ability to communicate with each other and then announce that a link exists between them. The routing protocols use these observations to find paths. We will refer to these types of protocols as link driven protocols meaning that the understanding of topology originates from the observation and dissemination of link states between nodes.[8] Our alternative to link driven protocols is the

---

[8] We use the word "driven" to avoid confusion with the well understood phrase "link state protocol" which refers to protocols that collect link states so that Dijkstra's algorithm may be used to calculate routes.

TABLE II
PHYSICAL AND SCENARIO FACTORS THAT AFFECT SCR SIGNALING DESIGN

Factor	Issue	Design Choice
Signaling power and signal detection threshold	These factors affect the range at which survivors are separated.	Larger separations may reduce interference at receivers but also reduces the number of survivors. The design goal is to choose these parameters to optimize the number of successful exchanges. [39]
Packet transmission power	Determines network degree. Network degree determines the potential for network partitions and the average number of hops between nodes, and affects the number of nodes that can transmit simultaneously in the network. Survivor density increases with network degree. [39]	Goal is to choose a transmission range that keeps the network connected but also attempts to maximize the total capacity of the network. In random networks, good connectivity occurs when network degree is >10. [48] This is also where CRS survivor density improvements begin to level off. [39]
Synchronization, transition times, signal detection time	These factors determine signal slot and interframe space sizing. The size of the signaling slot and interframe spaces determines the overhead of signaling. The cost of signaling is dependent on its relative size compared to that of the transmission slot.	Must weigh the impact of lost capacity with the desirability of CRS performance and the number of services provided.
Radios that can adapt transmission parameters, can point antennas, or can change channels	Benefiting from these capabilities requires a tighter compaction of survivors.	Echoing keeps neighboring survivors out of range of receivers and so at a range that does not benefit from these capabilities. RTS-CTS designs support tighter compaction and adaptation.
Radios that cannot adapt and cannot change channels	Tight compactions can result in unacceptable interference	Echoing mitigates the interference problem and may be preferred to RTS-CTS designs.
Some or all radios have a multiple channel output capability.	Requires the reception of multiple CTSs and ACKs	May revise CTS and ACK process to support multiple transmissions. May use echoing to further separate survivors to reduce interference in these receptions.
Some or all radios have a multiple input capability through the use of space-time array process techniques	Benefit comes from receivers being able to receive and distinguish multiple transmissions because they come from diverse directions. SCR already ensures that transmissions come from diverse directions.	Want to increase the potential that multiple transmissions will be sent to the same node. An RTS-CTS design is necessary. May also want to adjust transmission power to reduce network degree.
Radios have directional antennas	Echoing can be exploited to coordinate the pointing of antennas.	Echoing design can support the pointing of antennas in a way to increase the density of survivors. [49] RTS-CTS designs can exploit directional antennas also. The relative merit of one approach over the other depends on node degree and relative cost of signaling.
Maximum range of radio creates a low degree network.	There is greater potential for blocking.	Pure echoing designs may be preferred.
Average packet size is small	Transmission slot sizing for the small packet size increases the relative cost of signaling.	Less reliable CRS designs may create better performance because of reduced overhead.
Applications require specific service distinctions that may require preemption of real-time service.	Reservation priorities may be superseded.	Allowing priorities that can supersede reservations require policies for their use to prevent indiscriminate disruption of streams. It requires policies for the retention of reservations so that temporary disruption does not require re-instantiating the reservation.
Need reliable local broadcast.	Broadcasts are not acknowledged. To increase reliability requires reducing interference from other contenders.	Policy may be put in place to invoke echoing for broadcasts.

node state protocol. The distinction is that connectivity is not explicitly disseminated but is later inferred from the pairwise use of nodes states.

NSR requires two capabilities: location awareness and the ability to measure signal strength. With this information, a node creates a path loss map. Location and the path loss maps of all nodes provide sufficient information to determine connectivity and then topology. With this information, NSR provides the capability to predict connectivity at least as well as link driven protocols.

Using node states as the foundation of a routing protocol has advantages over link driven protocols in six ways.

1) The number of node states in a network is independent of the degree of the network. There is one state per node. In a link driven protocol, there is one state per pair of nodes. To illustrate the advantage consider a degree 10 network. For each ten node states there could be 45 bidirectional links, whereas only ten nodes states. The movement of one node changes one node state but possibly tens of link states. One node needs to advertise a changed node state, whereas tens of nodes may have to advertise changed link states. This advantage increases with network degree.
2) Node states provide more information about the network that is relevant to routing and QoS. In link-driven protocols metrics are derived by one of the end nodes or through the sharing of data between end nodes. In NSR, the state information at both end nodes and at all of their neighbors can be used to assign a value to a metric.
3) NSR allows the selection of the routing metric to be independent of the link discovery process. Pairwise coordination is not required to assign metric values and so the simultaneous use of different metrics does not entail any additional data gathering activities. In link driven protocols, metrics are assigned to links at the time of discovery. Using different metrics in reactive protocols requires nodes to use a separate discovery process for each metric and that cached information must be kept for each metric used. In link state protocols, nodes must coordinate the value of all metrics at discovery and then disseminate them. In distance vector protocols, in addition to discovery coordination, nodes must maintain a separate routing table and independently execute the distance vector algorithm for each metric.
4) NSR can be both predictive and adaptive. Rather than limiting routing to a previous understanding of topology, the use of node states enables the prediction of future topology. Although link driven protocols can implement mechanisms to estimate the longevity of links, they cannot estimate the creation of new links. New links must be discovered.
5) The quality of the node state can be leveraged to reduce the rate of its advertisement. To illustrate the advantage,

consider a network where one node does not move but multiple other nodes are moving. Most of the states of the stationary node do not change and so do not need to be disseminated as often. In a link driven protocol, the immobility of a node is irrelevant if multiple nodes are moving about it since its link states will still change.

6) Node states can support most QoS heuristics. NSR provides the unified mechanism to collect and to disseminate the required state information. As described in Section III-C, QoS protocols built on top of networks using link driven routing protocols must implement their own mechanisms to gather network state information.

A. Node State Information

There are two different routing constructs used in NSR, a node and a wormhole. The node construct is modeled as a point in space and is assumed to have connectivity with other nodes through the use of wireless connections. In many cases, nodes may be connected using a dedicated link such as a cable. To use these links within the node state routing protocol we define a second routing construct called a wormhole. We define our wormhole construct as a directed path between two points in the network. The basic algorithm used to select which routing constructs to use in a route considers the cost of sending a packet to a construct and the cost of using the construct. These costs are derived from the states of the nodes and the wormholes. Table III lists some proposed states that could be disseminated for each construct.

TABLE III
PROPOSED NODE STATES THAT ARE USEFUL FOR ROUTING AND QoS

STATE	DESCRIPTION
Address	MAC address of the node or the wormhole. In the case of the wormhole, the address is associated with the node at the front end.
1-meter Path loss	Pathloss of the first meter of propagation used with the log distance path loss model.
Propagation map	Propagation conditions can vary based on the location of nodes and the direction of propagation. To accommodate this concern we propose nodes measure and estimate a path loss exponent for the path loss model. We require each node that broadcasts a packet to announce the power level it is using. We assume that each destination node that hears a broadcast can determine the power level of the received signal and can then estimate a path loss exponent using the attenuation of the signal and the separation distance from the source. When propagation characteristics vary to different destinations, these states can be broken up into different sectors that account for these differences.
Cost	A cost that is assigned to using a node or a wormhole that is considered when assigning a metric to a connection.
Channel	The channel the node uses to receive a peer-to-peer packet. This state complements the channelization capability of SCR.
IP Addresses	IP addresses that are used by the node. It includes multicast addresses.
Direction	Current direction of movement of the node. Used to predict future topology
Dozing Offset	Used for node's that are using coordinated dozing. It is the offset in number of transmission slots from the time stamp that the dozing node will next wake up. See [40].
Dozing Period	Used for node's that are using coordinated dozing. It is the period at which the dozing node wakes. It is measured as an integer number of transmission slots. See [40].
Dozing State	This field identifies in which of the three dozing states the node is participating, the default state, the periodic waking state, or the coordinated dozing state. See [40].
Energy State	This is the state of the power supply being used by the node. It is proportional to the number of packets that the node can transmit using the contention power level. This assumes a maximum transmission power level is specified and that it is always used during the contention. Methods for estimating energy reserves are suggested in [50]. In the case of unconstrained energy nodes this level is set to the maximum value.
Fixed	This single bit identifies whether the node is stationary or it has the ability to move.
Infrastructure	This field identifies if the node is an infrastructure node and what special functions it performs. An infrastructure node may be an access point to the Internet or a real time service. It may also store node state information.
Location	The location defines where the node or where the wormholeís endpoints physically exist in the network. Node state routing requires location awareness.
CBR Bitmap	Bit map with one bit per slot in the CBR frame indicating in which slots this node perceives there to be a CBR reservation, either its own or a neighbor's.
Queue Size	The number of slot sized packets queued at the node. Used to identify congestion.
Receive Fraction	This is the fraction of the contention transmission power used by the node to receive a packet.
Time Stamp	This is the time that the reported state was measured. We assume time is absolute and synchronized throughout the network.
Velocity	Current velocity as measured by the node. Used to predict future topology.

B. Routing Using Node States

There are several problems involved in implementing node state routing with the foremost being: 1) how are node states disseminated; 2) how is connectivity inferred; and 3) what metric is used to build routes? There are numerous options. We present an implementation that attempts to be highly flexible and scalable.

1) Node State Dissemination: Node state routing is not immune from the concerns of overhead and protocol scaling. In our implementation of node state dissemination, we controlled overhead in two ways: 1) diffuse a node's state information at a rate that decreases with distance from the node and 2) force scaling by limiting the rate at which nodes may transmit node state information.

On a periodic basis, a node will broadcast a node state packet (NSP), which will include its own state and other states in its list restricted in number by the size of the transmission slot. The states that are included in these updates are selected by two criteria, a threshold that indicates whether an update is needed and a prioritization criterion to enable selection amongst several states that meet the update threshold. In the diffusion process, the update threshold depends on the distance between the node that owns the state and the node doing the rebroadcast. As an example, consider the following threshold based on time between updates.

$$\Delta T(d_{ij}) = \begin{cases} \Delta T_1, & d_{ij} \le 1 \\ c \cdot d_{ij} \cdot \Delta T_1, & d_{ij} > 1 \end{cases}. \tag{1}$$

ΔT_1 is the period in transmission slots at which one-hop neighbors are advertised, d_{ij} is the separation distance between the node transmitting the state, node i, and the node whose state is being disseminated, node j, and $\Delta T(d_{ij})$ is the period at which node i advertises the state of node j. The factor c is a constant that adjusts the rate at which states are updated. Different values for c can be used based on other states of the node. For example, the value of c would be greater for stationary nodes as opposed to mobile nodes. We prioritize updates by time difference with this criterion. State changes, such as changes in location, may also trigger the dissemination of node states, but this does not remove the requirement to disseminate states periodically since periodic dissemination is necessary to ascertain the health of nodes.

Scaling is forced using a minimum interval between NSP updates, i.e., a node may send one NSP per interval. However, NSP updates are accelerated when routing failures are observed. Loops do not occur in link state routing protocols if all nodes have the same link states. In NSR, nodes may have different node state information and loops may occur. We use the observation of a loop to trigger the accelerated updates. The goal of these updates is to synchronize the node state tables of all the nodes that form the loop so it can be broken. After identifying a looping condition, a node in the loop broadcasts its complete node state table or a subset that covers the region of interest, recalculates its routing tables and then forward the packet that was looping. This process is repeated so long as the packet remains in the loop. Ultimately, all nodes in the loop will have a common picture of the network and the packet will progress.

Loop identification mechanisms are easy to implement. There are node-based and packet-based approaches. In node-based approaches, unique information in the packet header is used to track the packets a node has handled. Incoming packets are compared with a list of handled packets and matches indicate a loop. The list is purged of old entries on a periodic basis that is much shorter than the time it takes to recycle the header information. Packet-based approaches either use hop count or a list of intermediate hops that have been traversed. Loops are detected in the first method if the hops exceed a threshold; and, in the second method, if a node sees its address in the list of nodes traversed. The hop count method is the easiest to implement but is inefficient in that updates continue at all hops past the threshold until the packet reaches its final destination rather than until the packet breaks free from the loop.

Through the diffusion and forced scaling mechanisms above, NSR aggressively employs a fisheye scope approach to routing [42], [43] that uses the distance effect [44] to mitigate the effect of stale states of distant nodes. The fisheye scope refers to the effect that a node's view of the network's topology is most accurate close to the node and decreases in accuracy with distance. This is accepted in the interest of scaling. Then, because location is a part of the routing calculation, the distance effect mitigates routing errors. The distance effect refers to the effect that the further nodes are apart from each other, the less effect their relative movement has on the direction between the two nodes. The next hop in routing a packet between the two, even with the stale information for distant nodes, is likely to be correct. The use of loop detection and the subsequent accelerated distribution of node states correct the situation when the network is too volatile for the diffusion mechanism.

2) Inferring Connectivity: Inferring connectivity involves predicting path loss between two nodes and determining if it is below a threshold for connectivity. Four observations on signal propagation are relevant to understanding our approach to predicting path loss.

1) Path loss generally increases as a power law function of distance.
2) Path loss may vary dramatically over short distances due to multipath effects.
3) Unlike multipath, losses that occur on a line-of-sight (LOS) path as a result of signal absorption, scattering, or occlusion cannot be regained. Non-LOS signal components that reach an occluded destination would be of a lower strength than if normal LOS propagation had occurred.
4) Radios can receive and detect signals with strengths that vary over a wide dynamic range, over 10^6 times.

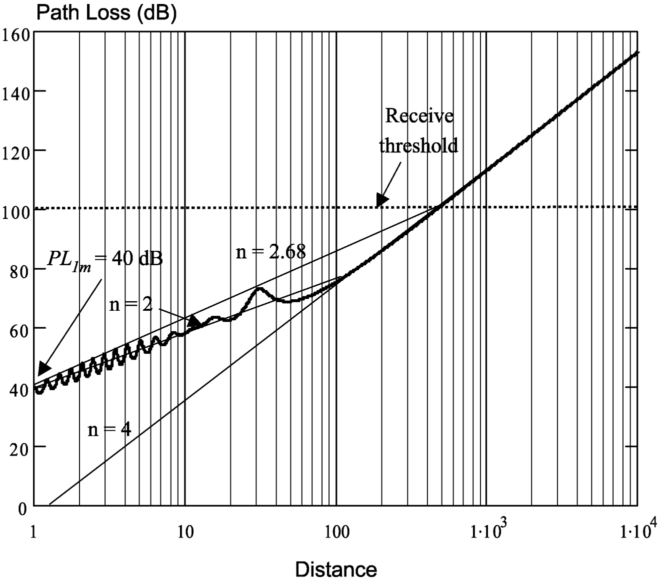


Fig. 16. Linear approximations of the two-ray propagation model.

The conclusions from these observations are that it is not practical to predict path loss with precision and that this is not necessary to infer connectivity. An approximation that is conservative in its estimate is suitable. The log-distance path loss model [45] can be used to provide such an estimate. It is a linear model when path loss (PL) and distance (d) are on a logarithmic scale, $\mathrm{PL(dB)} = \mathrm{PL(1\ m)} + 10n\log(d)$, and can be written as $\mathrm{PL} = \mathrm{PL}_{1\ m} d^n$ on a linear scale, where $\mathrm{PL}_{1\ m}$, the path loss of the first meter, and n, the path loss exponent, are the model's two parameters. We illustrate the suitability of this model in Fig. 16. It illustrates the path loss predicted by the two-ray propagation model[9] for a vertically polarized 2-GHz signal with antennas that are 1.5-m high, the piecewise linear approximation that goes with this prediction, and a third conservative linear model that would predict connectivity over the same range when the path loss threshold is 100 dB. If in a real environment there are deeper fades or greater losses, then a larger $\mathrm{PL}_{1\ m}$ or n can be used. There may be cases where a node cannot close a connection with a node at a closer range than another, which is further away. This case can usually be resolved using directional diversity to distinguish the two nodes. For this purpose, we propose a variable sized data structure that uses a series of words to specify path loss exponents on a directional basis. We use 8-bit words which allows us to specify 256 different pathloss exponents, in our case $n = 1.9$ to 7.0 in increments that provide equidistant changes in propagation

---

[9] The two-ray propagation model has been found to be representative of actual propagation in several studies [45]–[47].

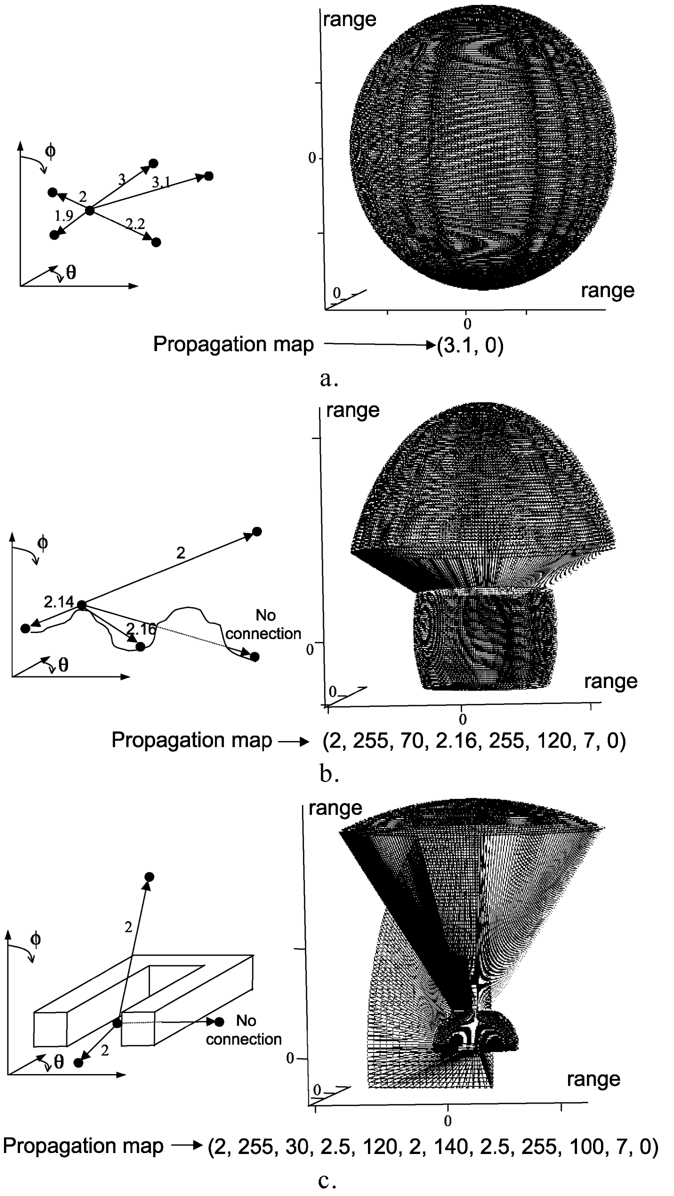


Fig. 17. Propagation scenarios with corresponding propagation maps and illustrations of the range of connectivity implied by these maps. (a) Example 1, since the larger exponent would predict connectivity for all the destinations, the propagation map consists of a single exponent. (b) Example 2, short data structure that predicts connectivity to all observed nodes but not to a known node that cannot be reached. (c) Example 3, data structure that captures propagation in an urban draw.

range and to divide a sphere into 256 longitudes (θ) and, by choice, 180 latitudes (ϕ), providing 46 080 sectors. Not all sectors need to be explicitly specified. The propagation map would have the form $(0, 0, n_{00}, \theta_{01}, n_{01}, \ldots \theta_{0x}, n_{0x}, 255,$ $\phi_1, \theta_{10}, n_{10}, \theta_{11}, n_{11}, \ldots, 255\,180)$. Since $\phi = 0$, $\theta = 0$, $\theta = 255$, and $\phi = 180$ occur predictably we can drop most from the structure. We still use $\theta = 255$ and 0 as delimiters in our abbreviated data structure. We illustrate several different propagation environments in Fig. 17 and show how they could be modeled using this abbreviated propagation map and how these propagation maps would predict maximum range by direction. It is possible to specify path loss exponents for seven different directions in the number of bits required for a single IPv6 address. Measuring pathloss exponents is tractable so long as sources can reliably specify the power they are using when transmitting and destinations can determine the range between nodes and the strength of the signals they receive. In cases where transmission characteristics vary by node, it may be necessary to provide another map structure for antenna gain, which can also be derived empirically over time. This latter map could be used to account for those effects that persist even with node movement such as a poor antenna connection, a damaged antenna, or a duffle bag placed next to an antenna on a vehicle. Methods to optimize propagation map size are beyond the scope of this paper.

Inferring connectivity is a two step process. The propagation maps provide the path loss of received signals. Since we seek bidirectional connectivity, in the first step, we determine the path loss using both end nodes' propagation maps, 1-m path loss values, and, if used, the combination of antenna gain maps, and then in the second step use the greater for determining connectivity.

3) Creating Routing Tables for QoS: After inferring connectivity between pairs of nodes, a node can easily calculate routes using Dijkstra's algorithm. Specific routing objectives can be incorporated using different link weights. The routing metric assigned to a link is built using the node states of the two end nodes of the link and possibly their neighbors. These metrics can be calculated to achieve many objectives. Table IV provides examples. These objectives may be combined to form additional metrics. An advantage of using node states is that a node can calculate multiple routing tables for different QoS objectives, and then route packets using the table that supports the QoS that the packet requires. For example, a table created using the energy conserving metric could be the default routing table but then when a stream reservation is being created nodes would use the "stream support" table or when a highly sensitive packet is being sent nodes would use the "trust" table.

C. Protocol Performance

We conducted a simple simulation experiment to test the viability of the node state routing. We placed 101 and then 202 mobile nodes on a square simulation area eight transmission ranges on a side to create networks with an average degree of 5 and 10, respectively. We saturated the network with payload traffic but gave NSPs priority in transmission. We used (1) to define the update rate with $\mathit{\Delta}T_1 = 3000$ transmission slots and $c = 3$. The minimum interval between updates was 50 transmission slots. We used a uniform propagation environment so propagation maps were single exponents. However, we limited the nodes to sending just eight node states in a node state update packet. The nodes moved continuously at the same speed using a random way point model with no pause time. All nodes moved at the same velocity, which was normalized to the range of the radio and the duration of the transmission slot. Table V translates these normalized velocities to units of kilometers per hour for various transmission bit rates and ranges.

Fig. 18 compares the routing tables at the nodes to those that would have been created using the same routing metric but with perfect information. Despite the mobility of the

TABLE IV
QoS ROUTING METRICS

Metric	Objective	Description	Relevant States
Energy Conserving	Conserve energy at energy constrained nodes to prolong battery life	The routing metric of a connection is made proportional to the amount of energy required to use the end nodes. This includes the amount of energy used by the source to transmit the packet and the amount of energy used by the connection destination to receive and process the packet. These costs are reduced for nodes that are not energy constrained and may be increased for nodes nearing the end of their battery life. Routing may also try to bypass nodes that are implementing a dozing schedule by penalize connections using these nodes. See [40] for more details.	1-meter path loss, propagation map, cost, dozing offset, dozing period, dozing state, energy state, receive fraction, location
Reliability	Use connections that are least likely to suffer interference	The routing metric is made proportional to path loss. This metric complements the SCR MAC. Since the SCR MAC separates contenders prior to packet transmission based on radio signal range, connections to destinations with low path loss are more likely to succeed on account of signal capture and the decreased probability that another contention survivor will also select it as a destination.	1-meter path loss, propagation map, location
Congestion Avoidance	Use connections through regions of the network that are least used	The routing metric is made proportional to the traffic queued at the source and destination ends of a connection and at their neighbors.	location, queue size
Stream Support	Use connections that are likely to persist for a long time and where transmission slots can be reserved.	The routing metric is made proportional to the expected longevity of the connections based on the understanding of movement of nodes. Preference would be given to connections between stationary nodes	1-meter path loss, propagation map, location, CBR bitmap, time stamp, velocity, direction
Trust	Use connections between nodes that are trusted	The routing metric is made proportional to the trustworthiness of the end nodes. Trustworthiness may be assigned manually to nodes, be assigned through an authentication mechanism, or be created by the context of the node's activities such as where they are and where they have been over time.	address, IP addresses, 1-meter path loss, propagation map, location, time stamp, velocity, direction
Long Distance Delivery	Use present and anticipated connections based on the expected progress of packets through a network	Connections are inferred based on anticipated location. The metric is then made proportional to the expected reliability of that prediction. Routing tables formed from this metric would be used for low priority packets that are expected to take a long time to move through the network to their final destination.	1-meter path loss, propagation map, location, time stamp, velocity, direction

TABLE V
EQUIVALENT VELOCITIES IN KILOMETERS PER HOUR FOR NORMALIZED VELOCITIES OF 0.000 01 AND 0.000 05 TRANSMISSION RANGES PER TRANSMISSION SLOT

v (% range/transmission slot)	**0.00001**				**0.00005**			
Data rate (Mbps) → Range (meters) ↓	**1**	**2**	**5.5**	**11**	**1**	**2**	**5.5**	**11**
300	1.8	3.0	4.9	6	9.3	15	24	30
1000	6.1	9.8	16	19.5	30	49	80	98
2000	12	19	31	38	60	97	157	191
10000	56	88	134	157	284	438	668	785

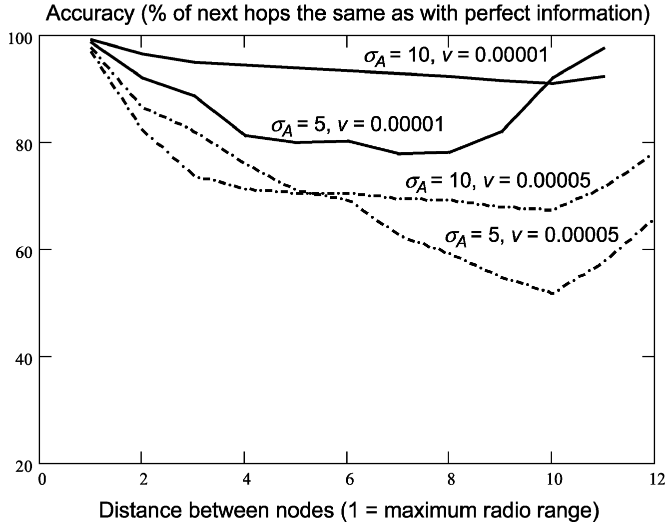


Fig. 18. Simulation results of the performance of the NSR diffusion update mechanism defined by (1) when $\Delta T_1 = 3000$ transmission slots and $c = 3$ for different node densities δ_A, and velocities v.

nodes in these simulations, the next-hop routing information remained surprisingly accurate even with the forced scaling. Route discrepancies are resolved as packets get closer to their destinations.

D. Path Reservation

NSR uses SCR to create multihop reservations for streams. To create a multihop connection, a node starts by reserving a transmission slot on the first hop of the route. If the node is successful in its contention then it sends a packet describing the connection required that includes the delay constraint and the destination ID.[10] The source node continues to send this packet in the same slot of each subsequent frame until a connection is established or it receives feedback that a connection cannot be made. In turn, the first-hop destination attempts to reserve a transmission slot along the next hop of the path in the same manner and then sends the same setup packet. This continues until a connection is established to the final destination. If the connection requires a bidirectional link the process repeats itself in the reverse direction. Nodes send their payload once they receive confirmation that the connected path has been established. The path is maintained as long as it is used.

---

[10] Node state information may be piggybacked onto these packets in order to fully use the transmission slot.

Delay constraints can be met by reserving slots carefully at each hop. Each node on a route is selective as to which transmission slots it attempts to reserve within a CBR frame. A node first estimates how many hops there are to the final destination. It uses this estimate and the total required delay to identify a suitable range of transmission slots within which to reserve a transmission slot. For example, say there are three hops to the final destination and the connection will allow 15 slots of delay. The first node reserves the first slot it can. Then, the second node attempts to reserve one of the next seven slots. If the node is unsuccessful at reserving one of these seven slots in the first frame it waits until the subsequent frame and attempts again. Say the node is successful at reserving the third slot following the arrival slot. Then, the next node would have a window of 11 slots in which to reserve the next-hop connection. Certain paths may not support the necessary end-to-end delay because of previous reservations. Variants of this path reservation protocol can use the CBR bitmap states and the queue size states associated with nodes to choose routes that can best support the delay requirements.

E. Exploiting Node States for Services

Node states provide many pieces of information that collectively provide a detailed view of network state. This information, together with the features of NSR and SCR can be exploited to provide additional services. We provide a brief description of proposed mechanisms to implement two interesting ones, multicasting, and traffic engineering.

Multicasting is supported by including multicast addresses in node states and providing a multicast packet format in SCR. Nodes subscribe to a multicast group by advertising the group address in their node state. Sources of traffic to multicast groups can identify the destinations and the best way to route packets to them using their node state tables. The multicast packet format provides space for a list of MAC addresses of destinations in the header. Sources consolidate the MAC address of multicast destinations that have the same next hop and will form as many packet replicas as there are next hops. Intermediate nodes that receive multicast packets attempt to route the packet to all the listed destinations. They also consolidate destination MAC addresses in packets with the same next hop, and transmit as many packets as there are next hops. Destination nodes that receive a multicast packet strip their address from the address list, and, if destinations remain, forward the packet to the remaining destinations as an intermediate node would.

Traffic engineering is implemented using the wormhole routing construct and the path reservation method. When it appears that there would be a benefit to routing traffic through specific regions of a network, a path can be reserved where desired and then advertised as a wormhole with a cost that encourages its use.

F. Design Considerations

NSR has a number of design options. The first option is the selection of states to collect. States determine what QoS mechanisms NSR can support. The states chosen will depend on what states can be measured and on the quantity of overhead that is acceptable in the network. The second option is the set of routing metrics to use and the methods to calculate their value. Table IV provides some ideas on the objectives for metrics and the relevant states. The third option is the set of parameters for node state updating and diffusion. The choice of parameters will depend on the operational scenario. Network volatility, network size (i.e., number of nodes), node density, and network dispersion can all affect the choice. The simulation results of Fig. 18 and the normalized velocities of Table V demonstrate that some tradeoffs are intuitive. Higher density networks mitigate the effect of less frequent updates since routes are more likely to be supported by nodes in the same geographic regions. Longer range radios also mitigate the effect of less frequent updates since they reduce the effect of node movement on network volatility. The last option is the policy for node state lifetime (i.e., the length of time a node state is considered suitable for routing). These policies will be dependent on network application and the operational scenario. If nodes are unlikely to drop out of a network and network topology changes slowly then long lifetimes are tolerable. If networks are highly volatile with changing membership, then short lifetimes may be necessary.

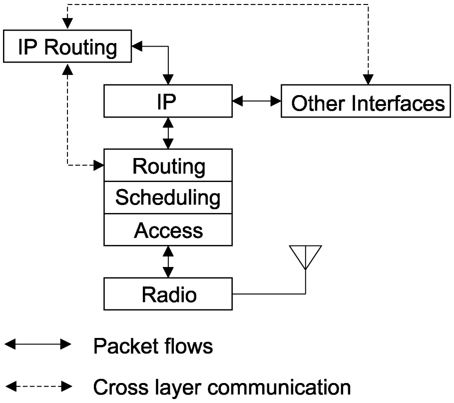


Fig. 19. Protocol stack for the ad hoc networking paradigm.

VI. HETEROGENEOUS NETWORKING

As presented, SCR and NSR are intended for a network with a homogeneous physical layer. Two questions follow: where should NSR fall in the protocol stack and how should this homogeneous network be integrated with other networks and network components? In answer to the first question, the best place to put a routing protocol that tries to conform to the wireless paradigm is in the link layer. As should be clear, the power of NSR comes from its tight integration with both the SCR MAC and the physical layer. Placing the routing function above IP complicates this integration. There are no advantages in routing performance that come from placing NSR above IP in a homogeneous ad hoc network. Unless there are multiple interfaces to IP, all packets are forwarded to the single wireless interface. NSR routes using MAC addresses. In Fig. 19, we illustrate a rich set of functions being performed by the link layer. An IP routing process (IPRP) is still shown but it contributes no traffic to the wireless network, rather it communicates to the link layer routing protocol and obtains its understanding of the membership and topology of the ad hoc network from the NSR node state and routing tables. The primary function of IPRP is to route to different interfaces and

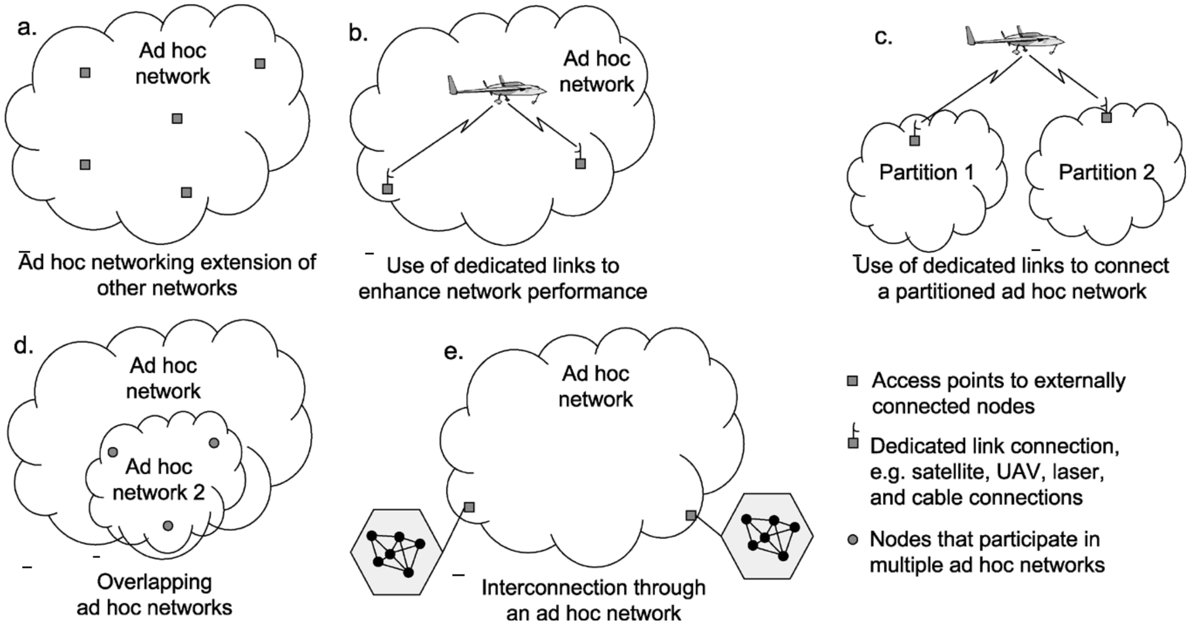


Fig. 20. Examples of heterogeneous network architectures.

beyond the ad hoc network. Thus, IPRP enables heterogeneous networking. It also serves a complementary function to NSR by identifying border routes[11] that can enhance the performance of the ad hoc network and by assisting in the dissemination of node states through these border routes to distant ends of an ad hoc network.

Fig. 20 illustrates several different heterogeneous configurations. Configuration (a) would be typical of a commercial or business ad hoc network, where the access points provide connectivity to a larger network. The NSR routing goal in this case would be to send as much traffic as possible through the fixed network. Nodes that are access points are identified by the interfaces node state. Explicit identification of wormholes between these access points is not required. Networks (b) and (c) show the use of dedicated connections to enhance ad hoc network performance and to prevent network partitioning. These connections would be represented as wormholes in the NSR protocol. In both of these cases but especially the second, IPRP would assist in advertising node states between the ends. Network (d) is the case where connections in either network can support routing in the other. The representation of this relationship may be either through a wormhole abstraction or as an interface's node state. The latter may be used in cases where the smaller ad hoc network is designed to use a longer range radio, where most connections are single hop as would be desirable for multicast groups.[12] Network (e) illustrates the case where special stub networks connect to the ad hoc network. NSR can be oblivious to the existence of these networks. IPRP has responsibility for learning and disseminating appropriate routing information to these networks. Again, consistent with our effort to minimize overhead on the ad hoc network, ad hoc network member nodes do not seek to discover this connectivity. Rather, IPRP at the access points advertise their connectivity by broadcasting this information across the ad hoc network as appropriate.

As has been described, NSR creates the abstraction for IPRP of a connected network. IPRP routes packets to the destination node in the ad hoc network not the next hop. For example, in cases where IPRP routes packets through the ad hoc network to an external network, it requests a route to the access point. The packet is handled by NSR and remains below IP on all hops until it reaches the access point, where it is forwarded up the protocol stack. This methodology also allows IP packets to be fragmented by SCR without requiring reassembly until they reach the end of their journey through the ad hoc network. Similarly, when NSR uses an external wormhole, the packet is tunneled through the IP network to the distant end, where it returns to being an SCR MAC packet. Finally, IPRP would arbitrate all higher layer QoS requests of the ad hoc network and translate these requests to appropriate SCR/NSR parameters. Admittedly, much is requested of IPRP. Due to its broad interaction with other protocols it would be best to design IPRP through a standardization process with a well defined interface to the link layer protocols. Modem designers would then design the link layer functions as would be optimum for the physical layer.

VII. CONCLUSION

In this paper, we argued that much of the previous work in ad hoc networking has been based on abstractions borrowed from wireline networks, e.g., the link, which is not well suited for ad hoc networks. We showed that SCR, a synchronous MAC based on a combination of CRS and an RTS-CTS handshake, is extremely versatile in terms of achieving spatial reuse, avoiding throughput collapse at network saturation, and enabling effective QoS functions. We described the CRS mechanisms that can

---

[11] A border route is a route that passes through an external network to another node in the ad hoc network. They are abstracted as wormholes within NSR.

[12] An example of where this type of connectivity would be useful is in a military organization, where all members of the multicast group belong to the same organization.

be used to prioritize access, channelize the network, and reserve resources. We demonstrated in a simulation that SCR provides near perfect prioritized access, even in multihop ad hoc networks, where we are aware of no other access protocol that can claim the same capability. At the same time, we proposed a complementary routing approach based on node states, which has many advantages over its link driven counterparts. It reduces the quantity of objects tracked, i.e., node states rather than pairwise link states, it supports the creation of multiple routing metrics without the need for pairwise coordination over the air, and allows multiple routing policies to coexist that can address different QoS objectives. We describe the critical data elements of node states and demonstrate how they are used to understand topology. We proposed a method to make this protocol scalable and demonstrated its effectiveness in a simulation. Finally, we discussed issues in integrating wireless ad hoc networks with IP networks. The protocol approaches we have proposed are open for many adaptations and improvements; however, they represent a major advance in achieving QoS in ad hoc networks from reliance on overprovisioning techniques to the use of traffic engineering techniques.

APPENDIX A
DESIGN OF COLLISION RESOLUTION SIGNALING USING SINGLE-SLOT PHASES

Let p^x be the probability that a contender will signal in phase x. Let $\mathbf{P}^x$ be the transition matrix of phase x and $\mathbf{Q}^n$ the transition matrix of the CRS design. The elements of $\mathbf{P}^x$ may be defined using

$$\mathbf{P}^x_{k,s} = \begin{cases} \binom{k}{s}(p^x)^s(1-p^x)^{k-s}, & 0 < s < k \\ (p^x)^k + (1-p^x)^k, & 0 < s = k \cdot \\ 0, & \text{otherwise} \end{cases}$$

where the entry $\mathbf{P}^x_{k,s}$ is the probability that s of k contenders survive the signaling phase. Note that s will never be 0 when $k > 1$ and will never be greater than k. The transition matrix of an n phase CRS design is $\mathbf{Q}^n = \prod_{x=1}^{n} \mathbf{P}^x$, and the probability that there will be just one surviving contender when there are k contenders at the beginning of signaling is $\mathbf{Q}^n_{k,1}$.

The objective of CRS design is to optimize the probability that just one node will survive the signaling by selecting the signaling probabilities, p^x. Designing CRS to maximize the probability that just one node survives when $k1$ nodes contend is relatively simple; however, a characteristic of CRS is that this maximum may result in a lower resolution probability when $k2$ nodes contend $k2 < k1$. Fig. 21 illustrates the effect. We define the design methodology as an optimization problem that will maximize the single node survivor probability for $k1$ without letting this probability for all $k2 < k1$ to be less than that at $k1$.

Let q^n be the set of p^x for an n phase CRS design, k_t be a target density of contending nodes, m be the total number of signaling slots allowed, and $S(q^n, k_t, m)$ be the probability that there will be only one surviving contender. Then, the optimization problem is

$$\begin{aligned} \max_{q^n} \quad & S(q^n, k_t, m) \\ \text{s.t.} \quad & S(q^n, k, m) \geq S(q^n, k_t, m) \quad \forall k, 0 < k < k_t. \end{aligned}$$

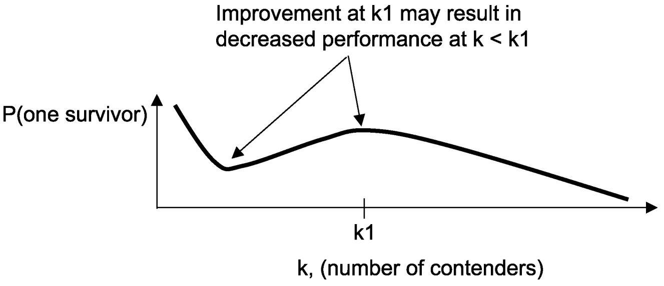


Fig. 21. Effect of optimizing a signaling design for a single contender density.

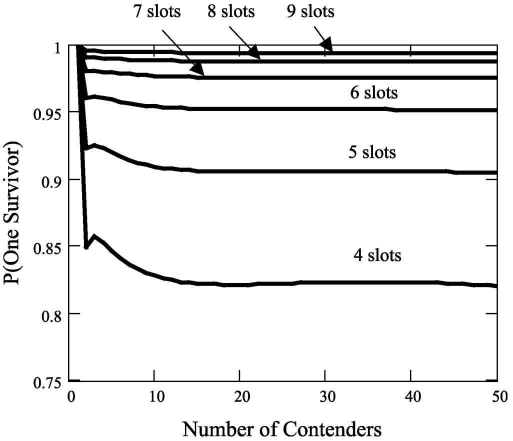


Fig. 22. 4–9 single-slot phase designs optimized for a 50 contender density.

The best solution for a finite set of signaling probability values can be found through an exhaustive search. The resulting performance of designs using four through nine phases and a design density of $k_t = 50$ are shown in Fig. 22. The nine-phase design has better than .99 probability that just one node will survive signaling for all practical densities of contenders.

ACKNOWLEDGMENT

J. A. Stine thanks R. Durst and K. Grace for their support of this research and for their reviews and suggestions on this paper. The authors thank the reviewers for their many helpful suggestions.

REFERENCES

[1] D. Perkins and H. Hughes, “A survey of quality-of-service support for mobile ad hoc networks,” *Wireless Commun. Mobile Comput.*, vol. 2, no. 5, pp. 503–513, Sept. 11, 2002.

[2] M. Chiang and G. Carlsson, “Admission control, power control, and QoS analyses for ad hoc networks,” in *Proc. IEEE Int. Conf. Communications*, 2001, pp. 245–249.

[3] Y. Dong, D. Makrakis, and T. Sullivan, “Effective admission control in multihop mobile ad hoc networks,” in *Proc. Int. Conf. Communications Technology*, 2003, pp. 1291–1294.

[4] G. Ahn, A. Campbell, A. Veres, and L. Sun, “Supporting service differentiation for real-time and best-effort traffic in stateless wireless ad hoc networks (SWAN),” *IEEE Trans. Mobile Comput.*, vol. 1, pp. 192–207, July–Sept. 2002.

[5] M. Kazantzidis, M. Gerla, and S. Lee, “Permissible throughput network feedback for adaptive multimedia in AODV MANETs,” in *Proc. IEEE Int. Conf. Communications*, 2001, pp. 1352–1356.

[6] K. Takahata, N. Uchida, and Y. Shibata, “QoS control for real-time video stream over hybrid network by wired and wireless LANs,” in *Proc. IEEE Int. Conf. Advance Information Networking Applications*, 2003, pp. 45–51.

[7] E. Royer and C.-K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Pers. Commun.*, pp. 46–55, Apr. 1999.

[8] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," IETF RFC 3561, July 2003.

[9] T. Clausen and P. Jacquet, "Optimized link state routing protocol (OLSR)," IETF RFC 3626, Oct. 2003.

[10] R. Ogier, F. Templin, and M. Lewis, "Topology dissemination based on reverse-path forwarding (TBRPF)," IETF RFC 3684, Feb. 2004.

[11] Quality of service for ad hoc on-demand distance vector routing, C. Perkins and E. Belding-Royer. [Online]. Available: http://people.nokia.net/~charliep/txt/aodvid/qos.txt

[12] M. Perlman and Z. Haas, "Determining the optimal configuration for the zone routing protocol," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1395–1414, Aug. 1999.

[13] R. Dube, C. Rais, K.-Y. Wang, and S. Tripathi, "Signal stability-based adaptive routing (SSA) for ad hoc mobile networks," *IEEE Pers. Commun.*, pp. 36–45, Feb. 1997.

[14] C.-K. Toh, "Associativity-based routing for ad-hoc mobile networks," *Wireless Pers. Commun.*, vol. 4, no. 2, pp. 1–36, Mar. 1997.

[15] S. Chen and K. Nahrstedt, "Distributed quality-of-service routing in ad hoc networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1488–1505, Aug. 1999.

[16] J. Liu, Q. Zhang, B. Li, W. Zhu, and J. Zhang, "A unified framework for resource discovery and QoS-aware provider selection in ad hoc networks," *Mobile Comput. Commun. Rev.*, vol. 6, no. 1, pp. 13–21, Jan. 2002.

[17] M. Pursley, H. Russell, and P. Staples, "Routing multimedia packets in a frequency-hop packet radio network," in *Proc. IEEE Military Communications Conf.*, 1996, pp. 220–224.

[18] X. Lin, Y. Kwok, and V. Lau, "BGCA: Bandwidth guarded channel adaptive routing for ad hoc networks," in *Proc. Wireless Communications Networking Conf.*, 2002, pp. 433–439.

[19] C. Zhu and M. S. Corson, "QoS routing for mobile ad hoc networks," in *Proc. Joint Conf. IEEE Computer Communications Societies*, 2002, pp. 958–967.

[20] S. Das, A. Mukherjee, S. Bandyopadhyay, K. Paul, and D. Saha, "Improving quality-of-service in ad hoc wireless networks with adaptive multi-path routing," in *Proc. IEEE Global Telecommunications Conf.*, 2000, pp. 261–265.

[21] C. D. Young, "USAP: A unifying dynamic distributed multichannel TDMA slot assignment protocol," in *Proc. IEEE Military Communications Conf.*, 1996, pp. 235–239.

[22] S. Lee, G. Ahn, X. Zhang, and A. Campbell, "INSIGNIA: An IP-based quality of service framework for mobile ad hoc networks," *J. Parallel Distrib. Comput.*, pp. 374–406, 2000.

[23] R. Kakaraparthi, R. Duggirala, and D. Agrawal, "Efficient message scheduling in ad hoc networks," in *Proc. IEEE Wireless Communications Networking Conf.*, 2000, pp. 1226–1231.

[24] D. Gu and J. Zhang, "QoS enhancement in IEEE 802.11 wireless local area networks," *IEEE Commun. Mag.*, pp. 120–124, June 2003.

[25] V. Kanodia, C. Li, A. Sabharwal, B. Sadegi, and E. Knightly, "Ordered packet scheduling in wireless ad hoc networks: Mechanisms and performance analysis," in *Proc. Mobile Ad Hoc Networking Computing*, 2002, pp. 58–70.

[26] X. Yang and N. Vaidya, "Priority scheduling in wireless ad hoc networks," in *Proc. Mobile Ad Hoc Networking Computing*, 2002, pp. 71–79.

[27] ETSI, "EN300 652 V 1.2.1," Broadband radio access networks (BRAN); High performance radio local area network (HIPERLAN) Type 1; Functional specification, July 1998.

[28] J. Sheu, C. Liu, S. Wu, and Y. Tseng, "A priority MAC protocol to support real-time traffic in ad hoc networks," *Wireless Networks*, vol. 10, pp. 61–69, Jan. 2004.

[29] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Proc. 7th Annu. Int. Conf. Mobile Computing Networking*, 2001, pp. 236–250.

[30] A. Banerjee, P. Chhabra, H. Russell, and E. Whitehill, "Performance of the reliable all-informed voice networking (RAVEN) system," in *Proc. IEEE Military Communications Conf.*, 2001, pp. 49–54.

[31] J. Sobrinho and A. Krishnakumar, "Quality-of-service in ad hoc carrier sense multiple access wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1353–1368, Aug. 1999.

[32] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, ANSI/IEEE Standard 802.11, 1999.

[33] J. Stine and G. de Veciana, "Tactical communications using the IEEE 802.11 MAC protocol," in *Proc. IEEE Military Communications Conf.*, 1998, pp. 578–582.

[34] *Department of Defense Interface Standard, Digital Message Transfer Device Subsystem*, MIL-STD-188-220C, May, 22 2002.

[35] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource reservation protocol (RSVP)-1 Version 1 functional specification," IETF RFC 2205, Sept. 1997.

[36] R. Braden, D. Clark, and S. Shenker, "Integrated services in the Internet Architecture—An overview," IETF RFC 1633, June 1994.

[37] C. D. Young, "USAP multiple access: Dynamic resource allocation for mobile multihop multichannel wireless networking," in *Proc. IEEE Military Communications Conf.*, 1999, pp. 271–275.

[38] R. Bittel, E. Caples, C. Young, and F. Loso, "Soldier phone: An innovative approach to wireless multimedia communications," in *Proc. IEEE Military Communications Conf.*, 1998, pp. 903–907.

[39] J. Stine, G. de Veciana, K. Grace, and R. Durst, "Orchestrating spatial reuse in wireless ad hoc networks using synchronous collision resolution," *J. Interconnection Networks*, vol. 3, no. 3/ 4, pp. 167–195, Sept./Dec. 2002.

[40] J. A. Stine and G. de Veciana, "A comprehensive energy conservation solution for mobile ad hoc networks," in *Proc. IEEE Int. Communication Conf.*, 2002, pp. 3341–3345.

[41] J. Stine, "Integrating the physical and link layers in modeling the wireless ad hoc networking MAC protocol synchronous collision resolution," in *Proc. OPNETWORK*, 2003, pp. 1–9.

[42] G. Pei, M. Gerla, and T.-W. Chen, "Fisheye scope routing: A routing scheme for wireless ad hoc networks," in *Proc. Int. Communications Conf.*, June 2000, pp. 70–74.

[43] K. Grace. Mobile mesh routing protocol. [Online]. Available: http:www.mitre.org/work/tech_transfer/mobilemesh/draft-grace-manet-mmrp-00.txt

[44] S. Basagni, I. Chlamatac, V. Syrotiuk, and B. Woodward, "A distance effect routing protocol with group mobility," in *Proc. 4th Annu. IEEE/ACM Conf. Mobile Computing Networking*, 1998, pp. 76–84.

[45] T. Rappaport, *Wireless Communications, Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[46] H. Xia, H. Bertoni, L. Baciel, A. Lindsay-Stewart, and R. Rowe, "Radio propagation characteristics for line-of-sight microcellular and personal communications," *IEEE Trans. Antennas Propagat.*, vol. 41, pp. 1493–1447, Oct. 1993.

[47] D. Green and M. Obaidat, "An accurate line of sight propagation performance model for ad-hoc 802.11 Wireless LAN (WLAN) devices," in *Proc. IEEE Int. Communication Conf.*, 2002, pp. 3424–3428.

[48] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," in *Proc. 3rd ACM Int. Symp. Mobile Ad Hoc Networking Computing*, 2002, pp. 80–91.

[49] K. Grace, J. Stine, and R. Durst, "An approach for modestly directional communications in mobile ad hoc networks," in *Proc. Int. Conf. Computer Communications and Networks*, 2003, pp. 377–384.

[50] S. Singh, M. Woo, and C. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *Proc. 4th Annu. ACM/IEEE Int. Conf. Mobile Computing Networking*, 1998, pp. 181–190.

John A. Stine (M'99) received the B.S. degree in general engineering from the United States Military Academy at West Point, West Point, NY, in 1981, and the M.S. degrees in manufacturing systems and electrical engineering and the Ph.D. degree in electrical engineering from the University of Texas at Austin in 1990 and 2001, respectively.

He served as an Engineer Officer in the U.S. Army for 20 years with relevant assignments as an Assistant Professor of Electrical Engineering at the United States Military Academy and as a Lead Analyst in the Army's Task Force XXI experiment, which was the Army's first attempt to network a brigade size maneuver force. He has been with The MITRE Corporation, McLean, VA, since 2001, where he does research in ad hoc networking and consults on projects concerning ad hoc networking, spectrum management, and modeling and simulation of command, control, communications, computer, intelligence, surveillance, and reconnaissance (C4ISR) systems.

Dr. Stine is a Registered Professional Engineer in the State of Virginia.

Gustavo de Veciana (S'88–M'94–SM'01) received the B.S., M.S, and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1987, 1990, and 1993, respectively.

He is currently a Professor in the Electrical and Computer Engineering Department, University of Texas at Austin. His research focuses on the design, analysis and control of telecommunication networks. His current interests include measurement, modeling and performance evaluation; wireless and sensor networks; architectures and algorithms to design reliable computing and network systems.

Dr. de Veciana has been an Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING. He is the recipient of a General Motors Foundation Centennial Fellowship in Electrical Engineering, a 1996 National Science Foundation CAREER Award, and corecipient of the IEEE Bill McCalla Best ICCAD Paper Award for 2000.