

Performance, Optimization, and Cross-Layer Design of Media Access Protocols for Wireless Ad Hoc Networks

Stavros Toumpis and Andrea J. Goldsmith

Department of Electrical Engineering, Stanford University, Stanford, CA, 94305

Abstract—We introduce a methodology for studying wireless ad hoc networks in a multihop traffic environment. Our approach is to use theoretical upper bounds on network performance for evaluating the effects of various design choices: we focus on power control, the queuing discipline, the choice of routing and media access protocols, and their interactions.

Using this framework, we then concentrate on the problem of medium access for wireless multihop networks. We first study CSMA/CA, and find that its performance strongly depends on the choice of the accompanying routing protocol. We then introduce two protocols that outperform CSMA/CA, both in terms of energy efficiency and achievable throughput. The *Progressive Back Off Algorithm (PBOA)* performs medium access jointly with power control. The *Progressive Ramp Up Algorithm (PRUA)* sacrifices energy efficiency in favor of higher throughput. Both protocols slot time, and are integrated with queuing disciplines that are more relaxed than the First In First Out (FIFO) rule. They are totally distributed and the overhead they require does not increase with the size and node density of the network.

I. INTRODUCTION

Wireless ad hoc networks are collections of nodes communicating over a wireless channel. Since wireless signal power decays with distance and in the presence of obstructions, each node can communicate directly with only some of the other nodes that typically lie in its vicinity. On the other hand, the traffic requirements of the nodes are taken to be arbitrary, therefore it is necessary that nodes cooperate to forward packets to their final destinations. The lack of any wired infrastructure, the nature of the wireless channel, and the need for robustness and scalability create many challenging design problems in the link, network and higher layers of the OSI hierarchy [1], [2], [3].

Here we develop a framework for studying the effect of various design choices on the traffic carrying capability of wireless ad hoc networks. Specifically, we compare the theoretical upper bounds on their performance (that can be calculated with the methods presented in [4], [5]) with their performance after important design choices are made. Many interesting facts are highlighted from this approach: the throughput enhancement offered by power control is limited in multihop environments; FIFO queuing is suboptimal; all routing protocols that only use a single route per source-destination pair typically suffer a throughput penalty; CSMA/CA should not be used over weak links; distributed MAC protocols inherently cannot achieve the capacity.

We then concentrate on the Medium Access Control (MAC) sub-layer [6], [7], [8], [9], [10], [11]. We propose two contention based protocols, the **Progressive Back Off Algorithm (PBOA)**, and the **Progressive Ramp Up Algorithm (PRUA)**. PBOA performs jointly medium access and power control. PRUA sacrifices energy efficiency in favor of a tighter packing of transmissions. Both protocols are integrated with a queuing discipline that is more relaxed than the ‘First In First Out’ (FIFO) rule. Time is divided in frames, and each frame consists of a contention slot and a data slot. Because time is slotted,

our protocols can easily be used for transmission scheduling, by assigning more than one data slot to each contention slot.

The rest of the paper is organized as follows: In Section II we specify the system model. In Section III we briefly describe previous research on theoretical upper bounds on the performance of ad hoc networks. In Section IV we discuss CSMA/CA and its shortcomings. The PBOA and PRUA protocols are introduced and analyzed in Sections V and VI respectively. We conclude in Section VII. Throughout the paper, terms being defined are set in **boldface**.

II. WIRELESS AD HOC NETWORK MODEL

We think of a wireless ad hoc network as a collection of nodes communicating over a common wireless medium, by exchanging data packets. Each node is equipped with a transmitter, a receiver and a buffer used for storing data. We assume that a given node cannot transmit and receive at the same time. Regarding the traffic, we assume nodes are not interested in multicasting packets, so that each created packet has a single destination. Beside this assumption, the traffic requirements of the nodes are arbitrary.

A. Link Model

Each node i can transmit with any power P_i between 0 and a given maximum P_i^{max} . When node i transmits with power P_i , node j receives the signal with power $G_{ij}P_i$. We define the **channel gain matrix** to be the $n \times n$ matrix $G = \{G_{ij}\}$, with $G_{ii} = 0$, $i = 1, \dots, n$. We model thermal noise and background interference (i.e. interference from other networks, etc.) jointly, as a single source of noise with power η_i for node i .

Let \mathcal{T} be the set of transmitting nodes at a given time and assume that node $j \notin \mathcal{T}$ is receiving information from node $i \in \mathcal{T}$. Then the **signal to interference and noise ratio (SINR)** at node j will be

$$\gamma_{ij} = \frac{G_{ij}P_i}{\eta_j + \sum_{k \in \mathcal{T}, k \neq i} G_{kj}P_k}$$

We assume that nodes use for all transmissions a common transmission rate R , and that the receiving node j will be able to decode the signal from i , with a negligible probability of error, provided the signal to interference and noise ratio γ_{ij} is constantly above a given threshold γ_T : $\gamma_{ij} \geq \gamma_T$. Otherwise, the signal is lost. This is a somewhat pessimistic assumption, as typically there is a non-negligible probability that a packet that is below but close to the SINR threshold will be correctly decoded. However, we will use this assumption for clarity of exposition, as the design of our protocols will not be affected and the numerical results will not change significantly if a more accurate model is used.

The common transmission rate R is a function of γ_T . The specific dependence of R on γ_T is not important, however one

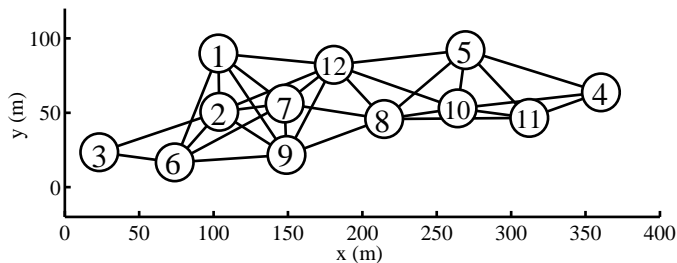


Fig. 1. A wireless ad hoc network of 12 nodes. A signal transmitted by a node reaches all others. However, signal strength decays with distance, and only nodes connected by a straight line can communicate directly with each other, in the absence of interference from competing transmissions.

choice could be Shannon’s capacity formula: $R = W \log_2(1 + \gamma_T)$ where W is the bandwidth of the channel. Note that the same formula, modified by multiplying γ_T by an appropriate factor, also gives the rate that can be achieved using uncoded or coded M-QAM, and a negligible bit error rate is tolerated [12].

As an example, let us consider the wireless ad hoc network of Fig. 1. The 12 nodes are placed in the box $\{0 \leq x \leq 400 \text{ m}, 0 \leq y \leq 100 \text{ m}\}$ randomly. For all nodes, $P_i^{max} = 0.3 \text{ W}$ and $\eta_i = 10^{-12} \text{ W} = -90 \text{ dBm}$. Also, the threshold SINR is $\gamma_T = 10 \text{ dB}$, and the common communication rate is $R = 1 \text{ Mbps}$. The link gain between nodes i and j is set to $G_{ij} = K(\frac{d_0}{d_{ij}})^\alpha$, with $K = 10^{-8}$, $d_0 = 1 \text{ m}$, and $\alpha = 4$, so that the power of the transmitted signal decays exponentially with distance. We choose not to model shadowing and fast fading for the sake of simplicity, although the extension would be straightforward. In the figure, nodes that can communicate with each other, in the absence of interference and with the transmitters using the maximum available power, are connected by straight lines. Nodes that can communicate directly but are not very close are more sensitive to competing transmissions.

B. When not to FIFO

Queuing disciplines other than the *First In First Out (FIFO)* rule have attracted significant research interest in the context of wired networks [13], [14]. However, to the best of the authors’ knowledge, they have not been considered in the context of wireless ad hoc networks, although they have the potential to increase spatial reuse.

As an example, consider the nodes of Fig. 1. Suppose that only node 7 is transmitting, to node 2, and node 8 has two packets in its queue, one for 7 and one for 10. If 8 follows the FIFO discipline and the packet at the head of its queue is intended for 7, 8 will have to wait for 7 to finish its transmission. However, if there is no FIFO restriction, 8 can go ahead and transmit the packet intended for 10 first without harming 7’s transmission, and then wait for 7 to become available (if it has not already become available), thus saving time and improving the channel utilization.

In the following we will *not* assume that nodes follow the FIFO rule. Rather, the queuing discipline will also be subject to optimization.

III. CAPACITY REGIONS OF WIRELESS AD HOC NETWORKS

A. Capacity Regions

Characterizing the transporting capability of wireless ad hoc networks is not a trivial task, for a number of reasons: First,

there can be multiple simultaneous transmissions. Also, some packets will travel through multihop routes, and will be transmitted multiple times. When calculating the capacity, such packets should only be counted once. Finally, each of the n nodes can have up to $(n - 1)$ different destinations and the number of packets that will be delivered to their final destination per unit of time clearly depends on the communicating pairs.

As a result, the transporting capability of the network must be described in terms of an $n(n - 1)$ -dimensional region, which we call the **capacity region** of the network. We refer to the points in the capacity region as **rate matrices**. Each rate matrix is a collection of $n(n - 1)$ rates with which the $n(n - 1)$ source-destination pairs can communicate, if the nodes follow some time division and multihop routing schedule.

Points on the boundary of the capacity region are achievable by an *optimal* time division and multihop routing schedule [5]. It is intuitively clear that this optimal schedule requires coordination among nodes on a global level. Specifically, arbitrarily distant nodes will need to coordinate their transmissions. Therefore, this time division and routing schedule must be computed in a central location, and distributed routing and media access algorithms in general cannot achieve the capacity. In light of Section II-B it should not be a surprise that this optimal schedule will typically involve non-FIFO queuing. In addition, packets belonging to a given source-destination will typically be routed through many different routes [5].

We define the **uniform capacity** of the network to be the maximum possible aggregate rate if all nodes send packets to all other nodes and the communication rates of all $n(n - 1)$ streams are equal. The uniform capacity corresponds to the rate matrix on the boundary of the capacity region whose coordinates are all equal and represents a simple figure of merit for the capacity of the network.

For more details on capacity regions we refer the interested reader to [4], [5], where we study the problem in depth, and develop the mathematical tools required to calculate them.

As an example, let us consider the capacity region of the network of Fig. 1. Since there are 12 nodes, its dimensionality will be $12 \times (12 - 1) = 132$, and it clearly cannot be plotted. Therefore, in Fig. 2 we plot a slice (line (a)) along which only the rate components $R_{3,12}$ and $R_{4,8}$ are non-zero. This slice describes the ability of the network to simultaneously forward packets from node 3 to node 12, and from node 4 to node 8, using multihop routing, time division and power control. The uniform capacity of the network is $C_u = 0.93 \text{ Mbps}$.

B. Effects of Power Control on Capacity

An important advantage of our formulation is that it allows us to change fundamental assumptions about the network, and find how these changes affect its inherent transport capabilities, in terms of its capacity region. As a simple application of this principle, let us assume that the nodes can no longer control their power, i.e., they either transmit with maximum power or not at all. In line (b) of Fig. 2 we plot the new slice of the capacity region. The uniform capacity now becomes $C_u^{NPC} = 0.86 \text{ Mbps}$. The 7.5% difference is the price we pay for not performing power control.

For the example network, and also for all other networks generated in a similar manner, the gains from allowing power control appear quite modest. This should be attributed to the multihop nature of the network. Here, it is not only important that a node transmits, but also that it transmits to a *distant* node, so that the data packet will move significantly closer to its final destination. On the other hand, as was also observed in [15], power control mainly favors the communication between nodes that are *close*, since now the transmitter can use much less

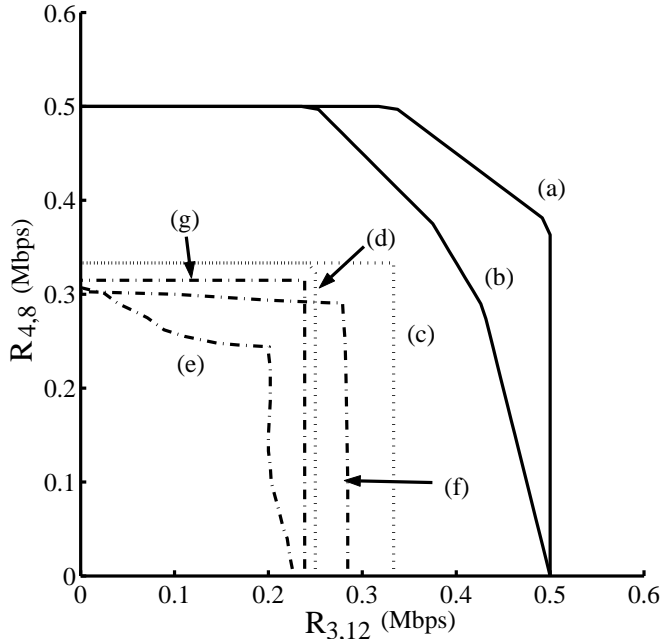


Fig. 2. Capacity region slices of the example ad hoc network along the plane $R_{ij} = 0$, $(i, j) \neq (3, 12), (4, 8)$, $i \neq j$: (a) Power control, optimal routing. (b) No power control, optimal routing. (c) Power control, RP-120. (d) No power control, RP-120. (e) RP-120, CSMA/CA. (f) RP-120, PBOA, with $m = 15$, $p = 0.8$. (g) RP-120, PRUA, with $m = 8$, $p = 0.3$, $P_T = 10^{-11}$ W.

power than the maximum, so that it creates a negligible amount of interference to other transmitters and still be able to transmit its packet to its destination.

Note that the gains from using power control are modest only when assuming that the network can achieve capacity. On the other hand, if the network is operating under a suite of protocols whose performance is significantly below the capacity, then in theory enhancing the suite of protocols with power control capabilities could improve the capacity significantly. In addition, power control also implies energy conservation: When the nodes of the example network achieve the uniform capacity, the energy dissipated for the end-to-end transmission of a single packet is 5.95 mJ without power control, and 3.25 mJ with power control.

C. Effects of the Routing Protocol on Capacity

Over the last few years the development of routing protocols for use in wireless ad hoc networks has attracted significant research interest and the related body of literature is extensive [16]. Most proposed routing protocols aim to discover a single route with the fewest possible hops between the two nodes that wish to communicate. However, there is no guarantee that this is the *best* route. Indeed, the best single route will depend on the traffic requirements of the network [17]. Furthermore, typically the optimal routing strategy does not consist of a single route. In [18] it was experimentally shown that allowing for alternate routes improves the performance of the network in terms of end-to-end throughput and delay. As expected, the *optimal* routing strategy, as determined with the methods of [4], [5], typically requires that a large number of nodes lying between the source and destination cooperate, the result being that a packet may travel along any of a large number of possible routes.

As an example, let us define the following three routing protocols for the example network of Fig. 1: Two nodes will use a

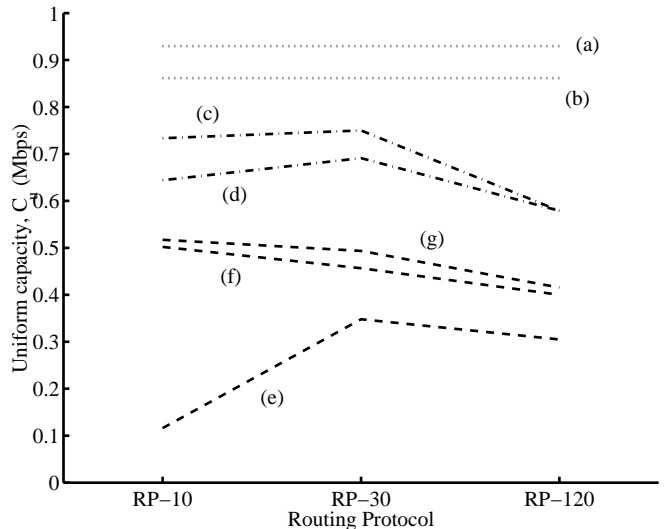


Fig. 3. Uniform capacities versus the routing protocol for the example network: (a) Power control, optimal routing. (b) No power control, optimal routing. (c) Power control, minimum-hop routing. (d) No power control, minimum-hop routing. (e) Minimum-hop routing, CSMA/CA. (f) Minimum-hop routing, PBOA, $m = 15$, $p = 0.8$. (g) Minimum-hop routing, PRUA, $m = 8$, $p = 0.3$, $P_T = 10^{-11}$ W.

single, minimum-hop route to communicate, with the requirement that they only use links whose SNR (in the absence of competing transmissions) is above a given threshold. For the three routing strategies, the thresholds are set to 10 dB (so that all links may be used for routing), 30 dB, and 120 dB (so that only the strongest links are used) respectively. We call these protocols RP-10, RP-30 and RP-120, respectively.

In Fig. 2 we plot capacity region slices of the example network, assuming the nodes control their power and operate under RP-120 (line (c)) and assuming no power control and RP-120 (line (d)). In Fig. 3 we plot the uniform capacities of the network for all three routing protocols, with (line (c)) and without (line (d)) power control. The uniform capacity is reduced by about 30% for all three routing protocols. In this discussion we have ignored the overhead associated with discovering new routes and using obsolete routing information. If we include this overhead, then depending on the level of node mobility, the actual throughput achieved by the network will further decrease.

IV. CARRIER SENSE MULTIPLE ACCESS WITH COLLISION AVOIDANCE (CSMA/CA)

Under CSMA/CA [7], [8], [9], a node that wants to send a data packet will first wait for the channel to become available and then transmit a short *RTS* (*Request To Send*) packet. The potential receiver, assuming it perceives an available channel, will immediately respond with a *CTS* (*Clear To Send*) packet that authorizes the initiating node to transmit, and also informs neighboring *hidden* nodes (i.e., nodes that are outside the communication range of the transmitter and might harm the ongoing transmission) that they will have to remain silent for the duration of the transmission. Nodes that overhear the RTS packet will refrain from transmitting until the CTS packet can be received, and those overhearing the CTS packet will refrain from transmitting until the whole data packet is sent. A node can only send an RTS packet if it perceives an idle channel and has not been silenced by another control packet. A node will only transmit a CTS packet if it has not been silenced by another control packet. The RTS/CTS handshake is coupled with a backoff

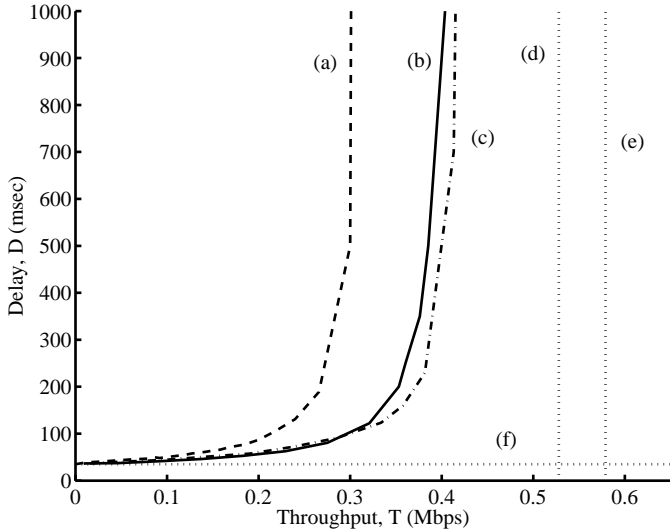


Fig. 4. Throughput-delay curves for the example network, under RP-120. (a) CSMA/CA. (b) PBOA with $m = 15$, $p = 0.8$. (c) PRUA with $m = 8$, $p = 0.3$, $P_T = 10^{-11} W$. (d) Uniform capacity with no power control. (e) Uniform capacity with perfect power control. (f) Lower bound on delay, caused by data packetization.

mechanism that ensures that nodes do not transmit immediately after the channel becomes available, but rather wait for a random period, to minimize the probability of a collision. In some incarnations [7], [9], after the data packet is correctly decoded, the receiver sends an *ACK* (*Acknowledgement*) packet to verify the reception on the link layer.

CSMA/CA is a substantial improvement over plain CSMA. However, it does not completely eliminate the hidden node problem, since, for example, it is still possible that CTS packets will collide. In addition, hidden nodes are not always within the communication range of the receiver. Secondly, since the potential transmitter must receive a control packet, it is necessary not only for nodes around the receiver to remain silent, but also for nodes around the transmitter. This artificial restriction leads to reduced spatial reuse. It is also well understood [19] that under heavy traffic, too much time (equivalently bandwidth) is wasted on collision avoidance and resolution, either because nodes back off excessively or because of frequent data and control packet collisions. Finally, there is no known simple way to integrate CSMA/CA with a power control algorithm.

As an example, let us study the performance of CSMA/CA, as it is implemented in the ad hoc portion of the IEEE 802.11 protocol (the Distributed Coordination Function) [9], for the example network of Fig. 1. We assume Poisson arrivals of packets of fixed size (set to 10 *Kbits*).

We refer to the set of all possible combinations of rates that can be achieved by the $n(n-1)$ source-destination pairs of the network when the nodes operate according to a MAC protocol as the **capacity region of the network under the given MAC protocol**. Contrary to the capacity regions defined previously, which we calculate analytically, this capacity region is obtained by simulation, as the accurate mathematical model for all MAC protocols we discuss is intractable. In addition, the shape of this capacity region will depend not only on the average arrival rates of the $n(n-1)$ streams, but also on the complete statistics of the arrival processes. In line (e) of Fig. 2 we plot the slice of the capacity region that corresponds to the two streams $3 \rightarrow 12$, $4 \rightarrow 8$ when nodes operate under CSMA/CA and RP-120.

We refer to the maximum aggregate communication rate

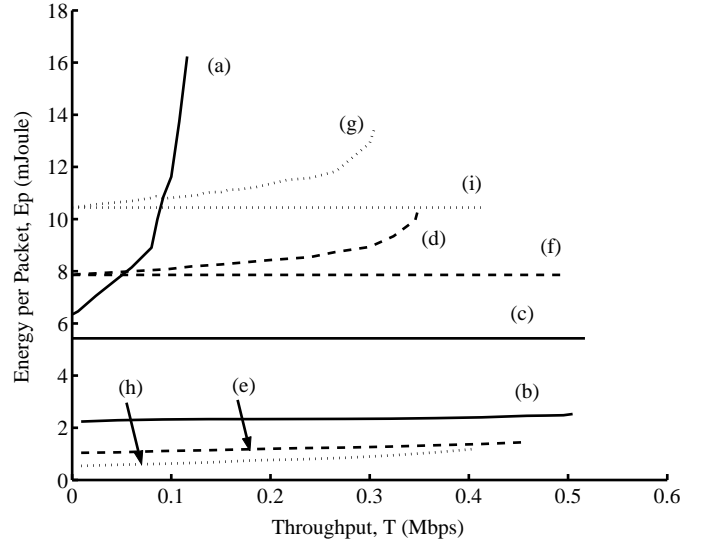


Fig. 5. Average energy consumed for the transportation of a packet to its final destination as a function of the achieved throughput, for the example network: (a) CSMA/CA, RP-10. (b) PBOA, RP-10. (c) PRUA, RP-10. (d) CSMA/CA, RP-30. (e) PBOA, RP-30. (f) PRUA, RP-30. (g) CSMA/CA, RP-120. (h) PBOA, RP-120. (i) PRUA, RP-120. For PBOA, $m = 15$, $p = 0.8$. For PRUA, $m = 8$, $p = 0.3$, $P_T = 10^{-9} W$.

achieved by the network, under uniform traffic conditions, when it is operating according to a MAC protocol as the **uniform capacity of the network under the given MAC protocol**. In Fig. 3 we plot the uniform capacities (line (e)) of the network under CSMA/CA and for the three routing protocols of Section III-C.

In Fig. 4 we draw the throughput-delay curve of CSMA/CA under RP-120 and with uniform traffic load (line (a)). Finally, in Fig. 5 we plot the average energy consumed for the transportation of a packet to its final destination with CSMA/CA, and under the three routing protocols RP-10, RP-30, RP-120 (lines (a), (d), (g) respectively).

From Figs. 2, 3 and 4 we see that, because of the problems of CSMA/CA that we have identified, the throughput it achieves is only 50% of the theoretical maximum, in the case of RP-30 and RP-120, i.e., those protocols that avoid weak links. More dramatically, CSMA/CA performs very poorly when coupled with the protocol that is using all links, RP-10, achieving less than 20% of the theoretical maximum. The reason for this very poor performance is that, under this routing strategy, for the transmissions over the weaker links there are many hidden nodes (i.e., nodes that will destroy the ongoing transmission if they transmit themselves) that lie outside the communication range of the receiver, so they never receive the CTS packet that instructs them to remain silent. Therefore, packets run a very high risk of being destroyed. Consequently, the average energy required for the successful relaying of a packet becomes very large when the network operates close to its capacity (see line (a) of Fig. 5).

The above result is not an artifact of the particular topology used, but rather highlights a general problem: CSMA/CA (and consequently IEEE 802.11) should not be used for communication over relatively weak links. Rather, it should be coupled with a routing protocol that actively avoids such links. Even if this is the case, using CSMA/CA is still problematic. For example, a strong link may simply not exist. Secondly, even if it does, the routing protocol may fail to find it because of collisions. We can summarize our findings in the following statement: CSMA/CA is not robust with respect to the network

topology and the choice of routing protocol.

In [11], [19] it was shown that the efficiency of the RTS/CTS handshake can be improved if time is slotted, and nodes start the contention at prearranged times. Then transmissions can be packed closer, and contention periods can be made shorter. In the next sections we present two protocols that are based on this idea.

V. PROGRESSIVE BACK OFF ALGORITHM (PBOA)

A. Frame Format

We start by specifying the frame format used by PBOA. As is shown in Fig. 6, time is partitioned in consecutive frames of fixed duration and each frame consists of a contention slot, followed by a data slot. In turn, the contention slot is divided into m pairs of minislots. The first minislot of each pair is reserved for the transmission of RTS packets, and the second slot for the transmission of CTS packets. The nodes contend for channel access during the minislot period and, depending on the outcome of the contention, some will transmit a data packet during the data slot. As a consequence, all data packets are required to have length equal to that of the data slot.

The time that must be devoted to a minislot will depend on the system parameters. As a typical case, let us assume that nodes transmit with a rate of 1 *Mbps*, the size of data packets is 10 *Kbits*, and the amount of time required by a transceiver to switch from receive mode to transmit mode is 10 μs . Also, let us assume a guard time of 3 μs , a maximum channel propagation delay of 2 μs , and that the time required for receiver synchronization is another 5 μs . As will be shown, each RTS packet will contain the MAC addresses of the source and the destination. Each CTS packet will also contain the source and destination MAC addresses, and in addition a scaling factor used for power control. Assuming that each of the addresses and the scaling factor are 8 *bits* long, the combined payload of an RTS-CTS pair is 40 *bits*. This brings the duration of a minislot pair to 80 μs , or 0.8% of the duration of the data packet. This calculation is only indicative and the actual overhead may be different if other system requirements are set. From the above discussion, it is clear that nodes must be equipped with clocks that allow global coordination on the level of a few microseconds. For a detailed discussion on minislot overheads and time synchronization we refer the reader to [10] and the references therein.

Finally, we assume that the mobility of nodes does not cause the channel gains to fluctuate significantly over the duration of a contention slot and the accompanying data slot. Using the parameters above, this duration is in the order of milliseconds, so our assumption is reasonable.

B. Protocol Operation

Before going into the details of how PBOA operates, we present the main idea: At the beginning of the contention period, all nodes that have a packet awaiting transmission will start contending for the channel. As the contention period progresses, nodes that are unsuccessful progressively back off, so that others may access the channel. At the same time, nodes that have succeeded in capturing the channel take advantage of the remaining time in the contention period to discover the minimum amount of power that is needed to transmit their data packet. The benefits are two-fold: (i) energy is conserved and (ii) other nodes may find that interference is reduced and so may also succeed in capturing the channel.

We now specify in detail the operation of the protocol during the contention period. At any time, a node will belong to one of three groups: the **contending** nodes, the **locked** nodes and

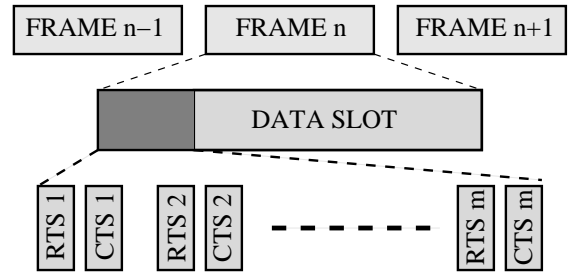


Fig. 6. Frame format of PBOA and PRUA.

the **silent** nodes. At the beginning of the contention period all nodes that have a packet to transmit are placed in the contending group. Each of them will initially try to transmit the packet at the head of its queue (to the node that represents the final destination of the packet or the next hop in its route). The rest of the nodes form the silent group.

At the beginning of the i -th RTS minislot: (I) Silent nodes listen to the channel. (II) Contending nodes transmit an RTS packet to their intended receiver with maximum power. (III) Locked nodes transmit an RTS packet to their intended receivers. The power they use for transmitting was specified by the intended receiver at a previous CTS minislot.

At the beginning of the i -th CTS minislot: (I) Silent nodes that received an RTS packet from a contending node successfully (i.e., the SINR γ_i was greater than or equal to the threshold SINR γ_T) will send, with maximum power, a CTS packet to notify that node of its success. Each of them will specify a factor equal to $g = \min((1 + \epsilon) \frac{\gamma_T}{\gamma_i}, 1)$ by which the potential transmitter must modify its power, where ϵ is a small design parameter that compensates for imperfect SINR estimation and changing channel conditions. Therefore, provided the potential transmitter decodes the CTS packet successfully, it will transmit in the next RTS slot with power $P_{i+1} = gP_i \leq P_i$ where P_i is the power used in the current RTS minislot. The success in the next RTS minislot is *guaranteed*, as no other node will be powering up, we have assumed that the channel gains remain fixed during the contention period, and transmissions with a SINR above the threshold are always successful. (II) Silent nodes that received an RTS packet (intended for them) from a locked node with an SINR $\gamma_i > (1 + \Delta)\gamma_T$ will transmit a CTS packet notifying the transmitter that it can power down to $P_{i+1} = P_i(1 + \epsilon) \frac{\gamma_T}{\gamma_i}$. We require that $\Delta > \epsilon$. (III) All other nodes remain silent.

At the end of the i -th CTS minislot: (I) Contending nodes that received a CTS packet (intended for them) become locked. (II) Contending nodes that did not receive a CTS packet will remain in the contending group with probability p , or will move in the silent group with probability $1 - p$. Any node that remains in the contending group tries to pick a new potential receiver. Specifically, it moves down its queue of buffered packets awaiting transmission (possibly returning to the head once the tail is reached) and searches for a packet with a different destination from the current. If such a packet does not exist, the node will contend for the same packet in the next minislot pair.

At the start of the data slot period, all locked nodes transmit their data packet, with the power level that was last specified to them by their intended receivers. The protocol guarantees that the data packets will be received successfully.

Note that the nodes do not use the FIFO queuing discipline. Rather, nodes that are not being successful try to pick a new potential receiver, for which the conditions may be more favorable (for example, maybe the first potential receiver is also contending, but the second is not). Also, a node that receives an RTS

packet from a locked node will only respond with a CTS if the SINR is above a threshold $(1 + \Delta)\gamma_T$, with $\Delta > \epsilon$. The reason for not *always* sending a CTS packet is for minimizing congestion during the CTS minislot.

Note that we have integrated the RTS/CTS mechanism with the power control algorithm of [20], [21]. We achieved this in a very straightforward manner, without having to resort to using a separate control channel [15] or a centralized scheduling algorithm [22]. Also, we have totally dispensed with the carrier sense mechanism.

The number of minislot pairs m and the probability of retransmission p are design variables. The optimal value for m depends on the overhead associated with each minislot, and also on the network topology and traffic pattern. The optimal value for p strongly depends on the choice of m . For more information on this issue, we refer the reader to [23].

C. Performance of PBOA

Line (f) of Fig. 2 is a slice of the capacity region of the network of Fig. 1, under PBOA and RP-120. We assume the system parameters of Section V-A and set $m = 15$, $p = 0.8$. In line (f) of Fig. 3 we plot the uniform capacity of PBOA for each of the three routing protocols RP-10, RP-30, RP-120. In line (b) of Fig. 4 we plot the throughput-delay curve of PBOA under RP-120, obtained by assuming that all of the 132 possible source-destination pairs communicate with a common rate. The maximum throughput is now $T_{max} = 0.403$ Mbps, as we see from the figure. All curves are determined by simulation.

As the figures show, the performance of PBOA in terms of throughput is clearly superior to that of CSMA/CA for all three routing protocols. This is expected since slotting time allows a closer packing of transmissions and a more efficient contention period. PBOA performs much better than CSMA/CA especially in the case of the routing protocol that uses weak links, RP-10. PBOA is therefore robust with respect to the topology and the routing protocol.

In addition, as Fig. 5 shows, PBOA is far superior to CSMA/CA in terms of energy efficiency. For example, under RP-10, PBOA uses 2.24 mJ per packet under low traffic conditions and 2.53 mJ under heavy traffic conditions. (Under low traffic conditions there are fewer transmissions going on and lower power levels can be used.) On the other hand CSMA/CA needs 6.35 mJ per packet under low traffic and 16.23 mJ under heavy traffic. (Under heavy traffic data packet collisions are frequent, so data packets need to be retransmitted, and more energy is lost.)

However, the throughput achieved by PBOA is still considerably below the theoretical maximum. There is no single source for this gap, however it is intuitively clear that one of the reasons is the following: During the contention period there are nodes that contend for the channel *in vain*. For example, a node may be sending an RTS to a node that has already sent an RTS and received a CTS in reply. The first node would have no possibility of success, but it would still be transmitting, thus hurting other nodes that still have a chance. If the nodes could make more informed decisions, the contention for the channel would be more efficient. This idea is the motivation for the **Progressive Ramp Up Algorithm (PRUA)**.

VI. PROGRESSIVE RAMP UP ALGORITHM (PRUA)

A. Protocol Operation

PRUA uses the same frame format as PBOA (see Fig. 6). However, the rules of contention are very different. Before we specify its operation in detail, we briefly discuss the main idea: During the contention period, a node that has a packet awaiting

transmission monitors the channel. Whenever it senses favorable conditions, it transmits an RTS with some small probability at the beginning of the next RTS minislot. If it is successful, it will persist sending RTS packets (so that it makes it known to competitors that it was successful). Otherwise, it backs off and may try again at a later RTS minislot. Nodes do not control their transmitter power.

We now specify in detail the operation of the protocol. During the i -th RTS minislot, a node A will transmit an RTS packet to a destination B if, in the previous minislot pair, it sent an RTS packet and received a CTS packet from B in reply. Alternatively, node A will send an RTS packet if all of the following conditions are met: **(I)** A has not sent a CTS packet in the previous minislot (i.e., it is not awaiting a packet in the coming data slot). **(II)** During the previous CTS slot, the combined power of all signals received at A did not exceed a predetermined threshold P_T (Otherwise, A may interfere with other transmissions). **(IIIa)** If A has not decoded an RTS packet in the previous RTS minislot, then A must have a non-empty queue. It will contend for the packet at the head of the queue. **(IIIb)** If A has correctly decoded an RTS packet in the previous RTS minislot from some node C, then there must be a packet in the queue, intended for some node B, such that B is able to decode the packet from A in the presence of interference from node C (specifically, B must be different from C). A will choose to contend for the packet that is closest to the queue head and satisfies this condition. **(IV)** A must perform a biased coin toss with probability of success p , and succeed.

During the i -th CTS minislot, all nodes that successfully decoded in the i -th RTS minislot an RTS packet intended for them reply with a CTS packet. In order to transmit a data packet, a node must receive a CTS packet in the last CTS minislot.

Nodes do not necessarily transmit the packet at the head of their queue, but rather the packet closest to the head of the queue for which the conditions appear more favorable. As in CSMA, nodes employ carrier sense. However, nodes now sense the CTS power from other receivers and not the power of ongoing transmissions, as in CSMA, so the decisions they make are based on much more relevant information.

The number of minislots m , the probability p and the carrier sense threshold P_T are design variables. As with PBOA, the optimal value for m depends on the overhead associated with each minislot, the network topology and the traffic pattern. The optimal value for p strongly depends on the choice of m . The optimal choice for P_T depends mostly on the channel model. For more information on this issue, we refer the reader to [23].

B. Performance of PRUA

Line (g) of Fig. 2 is a slice of the capacity region of the example network under PRUA and RP-120. We assume the system parameters of Section V-A and set $m = 8$, $p = 0.3$, $P_T = 10^{-11}$ W. In line (g) of Fig. 3 we plot the uniform capacities of the network for the three routing protocols RP-10, RP-30, RP-120, and under the same parameters. In line (c) of Fig. 4 we plot the throughput-delay curve of PRUA, assuming the same parameters, RP-120 and uniform traffic conditions. Finally, in Fig. 5 we show the average energy consumed per packet under PRUA and for all three routing protocols RP-10, RP-30, RP-120 (lines (c), (f), (i)). Since transmitters always transmit with their maximum power, and transmissions are always successful, the average energy does not depend on the throughput. All curves are determined by simulation.

The results show that PRUA is superior to PBOA in terms of throughput performance; allowing nodes to make more informed decisions helps marginally more than power control. On the other hand, PRUA is more wasteful in power: With RP-120,

PRUA requires 10.44 mJ per packet under any traffic intensity, whereas we have seen that PBOA requires 0.94 mJ per packet under low traffic, and 1.18 mJ under heavy traffic. On the other hand, because with PRUA there are no data packet collisions, PRUA does not require as much power as CSMA/CA does, especially under RP-10 (compare lines (a) and (c) of Fig. 5).

Although PRUA outperforms PBOA and CSMA/CA in terms of achievable throughput, the above figures show that there is clearly a large gap between the performance of PRUA and the theoretical maximum, in terms, for example, of uniform capacities. This is rather disappointing, given the relative advantages that PRUA has over CSMA/CA. The existence of this gap is due to the distributed nature of our protocols, in order to achieve capacity, it is necessary that *arbitrarily distant* node coordinate their transmissions. This behavior is not feasible by *distributed* MAC protocols.

VII. CONCLUSIONS

We study the performance of wireless ad hoc networks in a multihop traffic environment, with an emphasis on the medium access control sub-layer. Our approach is to evaluate the effects of various design choices by using theoretical upper bounds on the performance that are given in terms of capacity regions.

Applying this framework we arrive at a number of important observations: **(I)** Power control can be very helpful in terms of energy efficiency, but its gains are limited in multihop environments. **(II)** The First In First Out (FIFO) queuing discipline is suboptimal for use in wireless ad hoc networks. A more relaxed rule can lead to a tighter packing of transmissions. **(III)** The choice of routing protocol can significantly reduce the capacity of the network. Specifically, any routing strategy that assigns a single route to a node pair typically suffers a penalty in its performance. **(IV)** CSMA/CA is not well-suited for communication over weak links. Therefore, it should be coupled with routing protocols that avoid, if possible, such links. **(V)** Medium access protocols that are distributed inherently carry a penalty in their performance since, to achieve capacity, it is necessary that nodes that are separated by an arbitrarily large distance coordinate their transmissions. By definition, this is impossible in a distributed algorithm.

We also introduce two contention based protocols that perform medium access control in wireless ad hoc networks. The Progressive Back Off Algorithm (PBOA) performs medium access jointly with power control. The Progressive Ramp Up Algorithm (PRUA) sacrifices energy efficiency in favor of higher throughput. With both protocols time is divided in frames, each frame consisting of a number of RTS-CTS minislot pairs used for contention and a data slot. Both protocols use queuing disciplines that are more relaxed than the FIFO rule and promote dense spatial reuse. They are totally distributed and their control overhead does not scale with the size and node density of the network. Both are superior to CSMA/CA (and protocols based on CSMA/CA, such as IEEE 802.11), both in terms of achievable throughput and energy efficiency. In particular, they are robust with respect to the choice of routing protocol and network topology. Although we only present results for an example topology, we have also studied a large number of other topologies and have arrived at very similar results.

Our contention based protocols may be trivially converted to transmission scheduling protocols, by assigning more than one data slot to each contention slot, as in [10]. Alternatively, PRUA in particular can be easily modified to support reservations, with a mechanism similar to that of CATA [11].

Finally, PBOA and PRUA are not the only protocols that could follow the frame format of Fig. 6. Within this framework, other designs are certainly possible. These extensions are the subject of future work.

REFERENCES

- [1] *IEEE J. Select. Areas Commun.: Special Issue on Wireless Ad Hoc Networks*, vol. 17, no. 8, Aug. 1999.
- [2] C.-K. Toh, *Ad Hoc Mobile Wireless Networks: Protocols and Systems*, Englewood Cliffs, NJ: Prentice Hall, 1st edition, Dec. 2001.
- [3] *IEEE Personal Commun. Mag.: Special Issue on Advances in Mobile Ad Hoc Networks*, vol. 8, no. 1, Feb. 2001.
- [4] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," in *Proc. IEEE ICC*, New York, NY, Apr. 2002, vol. 5, pp. 3168–3173.
- [5] S. Toumpis and A. J. Goldsmith, "Capacity regions for ad hoc networks," to appear in *IEEE Trans. Wireless Commun.*
- [6] P. Karn, "MACA: A new channel access method for packet radio," in *Proc. CNC*, London, Ontario, Canada, Sep. 1990, vol. 1, pp. 134–140.
- [7] V. Bharghavan, A. Demers, S. Shenkar, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in *Proc. ACM SIGCOMM*, London, UK, Aug. 1994, vol. 1, pp. 212–225.
- [8] C. L. Fullmer and J. J. Garcia-Luna-Aceves, "Floor acquisition multiple access (FAMA) for packet-radio networks," in *Proc. ACM SIGCOMM*, Cambridge, MA, Aug. 1995, vol. 1, pp. 262–273.
- [9] *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, 1997.
- [10] C. Zhu and M.S. Corson, "A five phase reservation protocol (FPRP) for mobile ad hoc networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Apr. 1998, vol. 1, pp. 322–331.
- [11] Z. Tang and J. J. Garcia-Luna-Aceves, "A protocol for topology-dependent transmission scheduling in wireless networks," in *Proc. IEEE WCNC*, New Orleans, LA, Sep. 1999, vol. 3, pp. 1333–1337.
- [12] M. V. Eyuboğlu and Jr. G. D. Forney, "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 301–314, 1992.
- [13] D. P. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, NJ: Prentice Hall, 2nd edition, Dec. 1991.
- [14] M. Shreedhar and G. Varghese, "Efficient fair queuing using deficit round robin," *IEEE/ACM Trans. Networking*, vol. 4, no. 3, pp. 375–385, June 1996.
- [15] J. P. Monks, V. Bharghavan, and W. W. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proc. IEEE INFOCOM*, Anchorage, AL, Apr. 2001, vol. 1, pp. 1–11.
- [16] C. Perkins, Ed., *Ad Hoc Networking*, Boston, MA: Addison-Wesley, 1st edition, Dec. 2000.
- [17] S.-J. Lee and M. Gerla, "Dynamic load-aware routing in ad hoc networks," in *Proc. IEEE ICC*, Helsinki, Finland, June 2001, vol. 10, pp. 3206–3210.
- [18] M. R. Pearlman, Z. J. Haas, P. Sholander, and S. S. Tabrizi, "On the impact of alternate path routing for load balancing in mobile ad hoc networks," in *Proc. ACM MobiHOC*, Boston, MA, Aug. 2000, vol. 1, pp. 3–10.
- [19] I. Chlamtac, A. Faragò, A. D. Myers, V. R. Syrotiuk, and G. Záruba, "ADAPT: A dynamically self-adjusting media access control protocol for ad hoc networks," in *Proc. IEEE GLOBECOM*, Rio De Janeiro, Brazil, Dec. 1999, vol. 1, pp. 11–15.
- [20] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Tech.*, vol. 42, no. 4, pp. 641–646, Apr. 1993.
- [21] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems," in *Proc. 5th WINLAB Workshop*, Rutgers University, New Brunswick, NJ, Oct. 1993, vol. 1, pp. 249–257.
- [22] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad hoc networks," in *Proc. IEEE INFOCOM*, New York, NY, June 2002, vol. 2, pp. 976–985.
- [23] S. Toumpis and A. J. Goldsmith, "Performance, optimization, and cross-layer design of media access protocols for wireless ad hoc networks," in preparation for submission in *IEEE Trans. Wireless Commun.*