Joint Routing and Link Rate Allocation under Bandwidth and Energy Constraints in Sensor Networks

Maggie Cheng, Xuan Gong, and Lin Cai

Abstract—In sensor networks, both energy and bandwidth are scarce resources. In the past, many energy efficient routing algorithms have been devised in order to maximize network lifetime, in which wireless link bandwidth has been optimistically assumed to be sufficient. This article shows that ignoring the bandwidth constraint can lead to infeasible routing solutions. As energy constraint affects how data should be routed, link bandwidth also affects not only the routing topology but also the allowed data rate on each link. In this paper, we discuss the sufficient condition on link bandwidth that makes a routing solution feasible, then provide mathematical optimization models to tackle both energy and bandwidth constraints. We first present a basic mathematical model to address using uniform transmission power for routing without data aggregation, then extend it to handle nonuniform transmission power, and then routing with data aggregation. We propose two efficient heuristics to compute the routing topology and link data rate. Simulation results show that these heuristics provide more feasible routing solutions than previous work, and provide significant improvement on throughput and lifetime.

Index Terms—Sensor networks, routing, cross layer design, optimization.

I. INTRODUCTION

WIRELESS sensor networks are resource scarce, which is manifested in both energy and link bandwidth, as well as computing power etc. While it has been widely accepted that energy constraint limits the total amount of data being transmitted, and plays an important role for sensor network lifetime, bandwidth constraint has long being ignored. In previous work related to energy efficient routing and data aggregation [1]–[10], wireless link bandwidth is often optimistically assumed to be large enough. Actually, in a sensor network where every node transmits towards the sink, the aggregated bandwidth requirement can be surprisingly high. Even for a single path with three or more hops between a source and a sink, if the link bandwidth is B, the allowed source rate is only 1/3 B, because the transmission of the source node is conflictive with that of the next two hops. It could be worse in a complicated network topology. If the total required data rate is higher than the link bandwidth on any particular link, the source rate cannot be supported, and network congestion is doomed.
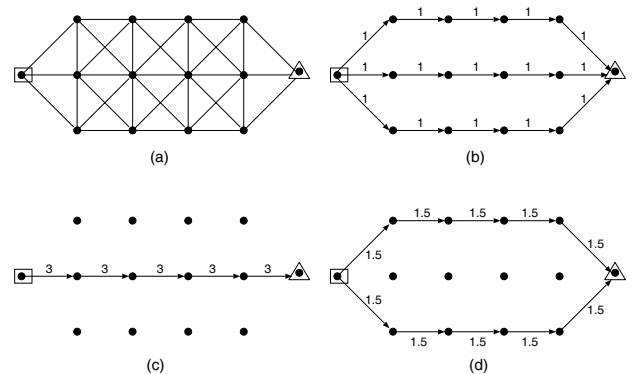
Fig. 1. For the network shown in (a), nodes within each other's transmission range are connected with a line. The three routing options (b), (c) and (d) lead to the same lifetime, but (b) and (c) demand higher bandwidths than (d).

In most previous work on energy efficient routing, routing decisions are made to optimize the energy aspect and tend to ignore the bandwidth limitation. For the network shown in Fig. 1(a), a maximum lifetime routing algorithm would choose any of the routing topologies shown in Fig. 1(b),(c) and (d) because they all lead to the same lifetime. However, (b) and (c) demand much higher bandwidth than (d). Suppose that there exists an optimal MAC layer solution that requires the minimum bandwidth to support a given routing. If the source is generating 3 units of data per second, (b) requires a bandwidth of 6 units per second by the optimal solution (and 9 units per second by our sufficient condition defined in Section II); (c) requires 9 units per second by the optimal solution (and 9 units per second by our sufficient condition); and (d) only requires 4.5 units per second by the optimal solution (and 4.5 units per second by our sufficient condition). In a slightly different scenario shown in Fig. 2, the solution that provides the longest lifetime is actually the worst in terms of bandwidth requirement. A shortest path routing algorithm would choose (b) to maximize lifetime, but the required bandwidth may be too high to accommodate.

From the two examples above, we observed that for a randomly deployed network, usually the one that is likely to be used as a relay node is at the core of the network (if every one chooses what is best for itself selfishly), which unfortunately is also the most interfered area due to the broadcast nature of wireless transmissions. Sending a lot of data to the core is likely to congest the network, but sending every packet along the outlier is not the best solution either.

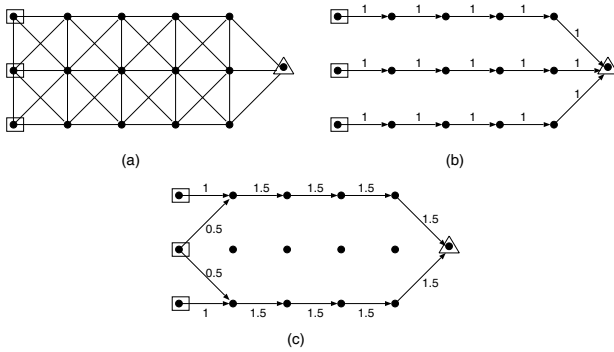What should be the maximum lifetime routing solution that

Fig. 2. For the network shown in (a), both (b) and (c) use shortest paths routing; (b) is optimal in terms of lifetime, but is the worst in terms of bandwidth; (c) is the best in terms of bandwidth, but is suboptimal in terms of lifetime.

is feasible with link bandwidth constraint? Apparently there is no generic policy such as shortest path routing or minimum energy routing that can lead to the maximum lifetime and be accommodated by the link bandwidth. To answer this question, we first consider for an arbitrary network topology, what condition should hold in order to ensure all source data can be put through, with each source generating data at a given rate. In this article, we discuss the sufficient condition on the link bandwidth, and use the bandwidth constraint to decide not only the routing topology but also the actual data rate on each link. The routing topology and link data rate are computed by solving an optimization problem that includes both energy and bandwidth constraints.

The major contributions of this work are that we formulated the energy and bandwidth-constrained routing problem as a multi-constraint optimization problem and provided efficient heuristic solutions to it. In addition, a companion time slot assignment algorithm is proposed to support the resulting routing solution at the MAC layer.

The rest of the paper is organized as follows: Section II discusses the sufficient conditions on link bandwidth; Section III formally describes the energy-bandwidth constrained routing problem and provides a mathematical model for the problem; Section IV presents two heuristics for joint optimization of energy and bandwidth; Section III-E addresses how to use the mathematical model to address in-network data aggregation; Section V provides numerical simulation results that show the comparison of algorithms in terms of throughput and lifetime; Section VI briefly surveys the related work, followed by concluding remarks and further research issues in Section VII.

II. A SUFFICIENT CONDITION FOR COLLISION-FREE COMMUNICATION

Let R_{ij} denote the data rate from node i to node j. Assume that the MAC layer uses an efficient TDMA scheme in which the number of time slots assigned to link (i, j) is proportional to R_{ij}. For any node i's reception to be successful, the TDMA schedule must satisfy that (1) when node i is receiving, it cannot be sending, and (2) when node i is receiving from j, none of its neighbors except j should be sending. Let N_i denote the neighbors of node i, and B the wireless link bandwidth. These two necessary conditions can be written as:

1) $\sum_{j \in N_i} (R_{ij} + R_{ji}) \leq B, \forall i;$

2) $\sum_{j \in N_i} R_{ji} + \max_{j \in N_i} \left(\sum_{k \in N_j, k != i} R_{jk} \right) \leq B, \forall i.$

However, these two are only necessary but not sufficient conditions, i.e., satisfying these two conditions does not guarantee that conflicting transmissions can always be assigned to different slots. In this paper, we prove that the sufficient condition to guarantee a global collision-free schedule is:

$$\sum_{j \in N_i} R_{ij} + f_i \cdot \sum_{j \in N_i} \sum_{k \in N_j} R_{jk} \leq B, \quad \forall i \quad \texttt{(sufficient)}$$

where $f_i = 1$ if node i is a receiver, and $f_i = 0$ otherwise. The proof of the sufficient condition is included in the Appendix.

The sufficient condition may require more bandwidth than necessary, but if this condition is satisfied at each node, it guarantees that a conflict-free time slot assignment can be found, which provides guaranteed data rate for each node. If each node injects data into the network at a rate below the guaranteed source rate, the network will be congestion-free. Moreover, since every node transmits at its scheduled time slot, there will be predictable delay at each hop, and hence bounded delay from the source to the sink. In the following sections, we base our discussion on the sufficient condition only.

III. MATHEMATICAL MODEL

A. Problem Definition

We assume that in a sensor network of n nodes, wireless link capacity is B (bits per second), and each node i has initial battery energy E_i (J). Each node i generates sensory data at a rate of R_i bits per second ($R_i > 0$ if node i is a source, $R_i = 0$ if it is a pure relay node, and $R_i < 0$ if it is a sink). We assume that nodes consume energy on transmitting, receiving and sensing (i.e., generating sensory data), and their energy consumption rates are Pt, Pr, and Ps J per bit respectively. We further assume that Pr and Ps are constants, but Pt is handled differently in the two models: in the uniform model, each node transmits at the same power level Pt; in the nonuniform model, each node can transmit at different power level from others but the transmission power used by node i is still fixed, denoted by Pt_i.

The energy-bandwidth constrained maximum lifetime routing problem can be formally stated as follows: Suppose that sources are preselected and each node i's rate R_i is known, but the transmission rate from node i to node j is unknown. Let T be the total network lifetime. The rate allocation problem is to compute the data rate R_{ij} on each link (i, j), given each node i's E_i, R_i and link capacity B, such that the total network lifetime T is maximized and the rate allocation can be accommodated by wireless link capacity and energy reserve.

B. With Uniform Transmission Power

In this model every node uses the same transmission power, therefore links are all symmetric. We use N_i to denote the neighbors of i excluding i itself, and f_i as an indicator of the receiver, as defined in Section II:

$$f_i = 1, \quad \text{if } \sum_{j \in N_i} R_{ji} > 0;$$

$$f_i = 0, \quad \text{otherwise.}$$

To maximize lifetime T is equivalent to minimize $1/T$. Thus, we can formulate the rate allocation problem as follows.

minimize

$$1/T \qquad (1)$$

subject to

$$\sum_{j \in N_i} (R_{ij} - R_{ji}) = R_i, \qquad \forall i \qquad (2a)$$

$$PsR_i + \sum_{j \in N_i} (PrR_{ji} + PtR_{ij}) \le E_i/T, \qquad \forall i \qquad (2b)$$

$$\sum_{j \in N_i} R_{ij} + f_i \cdot \sum_{j \in N_i} \sum_{k \in N_j} R_{jk} \le B, \qquad \forall i \qquad (2c)$$

$$0 \le R_{ij} \le B, \qquad \forall i, \forall j \qquad (2d)$$

$$f_i = \{0, 1\}, \qquad \forall i \qquad (2e)$$

In this formulation, equality (2a) indicates that data rates R_{ij} satisfy flow conservation at each node; Inequality (2b) is the energy constraint, and inequality (2c) defines the bandwidth constraint. In wireless networks, the capacity constraint is different from that in a flow network, where each link (u, v) has a fixed link capacity $c(u, v)$ and flow $f(u, v) \le c(u, v)$ is the only capacity constraint on each individual link. In wireless networks, because of the broadcast nature of transmission, the capacity constraint needs to be considered on a collision domain, rather than on each link separately. In other words, how much can be transmitted over one link depends on not only the link capacity B, but also the amount of data transmitted over other links in its collision domain. Inequality (2c) ensures that all links possibly in the same collision domain have a total demand less than B— If node i is a sender but not a receiver, it only needs to satisfy that the sum of the flow going out of i is bounded by B; If node i is a receiver, it needs to satisfy that node i's sending, receiving and other interfering nodes' transmission have a total demand of at most B; If node i is neither a sender nor a receiver, (2c) is automatically satisfied. Inequalities (2d) and (2e) are constraints for the variables.

C. With Nonuniform Transmission Power

In this model, we assume that each node still uses fixed transmission power, but node i can use Pt_i to transmit and node j can use Pt_j to transmit, and it is possible $Pt_i \ne Pt_j$. The inequality (2b) of the above linear program is modified as in (3a) to reflect the individual transmission power.

With this model, network topology is predetermined, but the links can be unsymmetrical. To deal with asymmetrical links, we use N_i^+ to denote the neighbors that can receive from node i; and N_i^- to denote the neighbors that node i can receive from. Therefore the inequality (2c) is modified as in (3b) to reflect the change on the collision domain.

$$PsR_i + \sum_{j \in N_i} (PrR_{ji} + Pt_iR_{ij}) \le E_i/T, \qquad \forall i \qquad (3a)$$

$$\sum_{j \in N_i^+} R_{ij} + f_i \sum_{j \in N_i^-} \sum_{k \in N_j^+} R_{jk} \le B, \qquad \forall i \qquad (3b)$$

D. With Double Disk Model

The models presented in section III-B and III-C both assume a single disk model, i.e., the effective transmission range is the same as the interference range. In reality, the interference range is usually larger than the effective transmission range. For example, a radio's transmission range is 500 meters, but the nodes located 800 meters away still are interfered by this node's transmission. Between 500 meters and 800 meters, the signal is not strong enough to be decoded, but strong enough to cause intereference at others. In this section we modify our model to reflect this phenomena.

We use the double disk model with the uniform transmission power. In terms of energy constraint, the inequality (2b) remains the same, since the transmission range remains the same; in terms of the bandwidth constraint, the definition of neighbors is changed. We use N_i to denote the nodes that are in the transmission range of node i, N_{i_F} to denote the nodes that are in the interference range of node i. Since the interference range is larger than the transmission range, apparently $N_i \subseteq N_{i_F}$. Since all links are symmetrical, if $i \in N_{j_F}$, then $j \in N_{i_F}$. The bandwidth constraint is changed to:

$$\sum_{j \in N_i} R_{ij} + f_i \sum_{j \in N_{i_F}} \sum_{k \in N_j} R_{jk} \le B, \qquad \forall i \qquad (4a)$$

E. Data Aggregation

Section III-B - III-D gives a mathematical formulation for a basic data forwarding scheme without data aggregation. In sensor networks, sometimes data aggregation is used to reduce the number of transmissions. In this section we show that this model can be extended to compute the optimal routing and link rate allocation for data aggregation as long as the data aggregation scheme is given.

A well known data aggregation scheme is to aggregate data from different sources when they arrive at a relay node at a close time-frame. The idea is similar to that used in Opportunistic Network Coding [11]. In Fig. 3, suppose source node i generates data at a rate of 5 packets per second, and input link (j, i) has a rate of 3 packets per second, and (k, i) has a rate of 2 packets per second, then the output flow of node i has a total rate of 5 packets per second, because each packet from the the low-rate flows can be combined with a packet of the high rate flow and get a "free ride".

Thus, the flow conservation constraint in equality (2a) is changed to

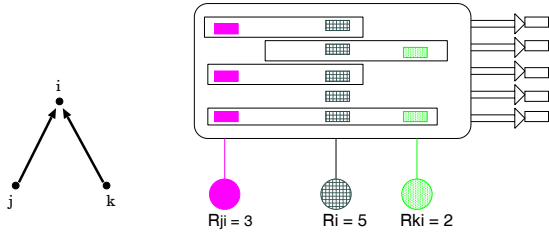$$\sum_{j \in N_i} R_{ij} = \max_{j \in N_i} \{R_i, R_{ji}\}, \qquad \forall i \qquad (5a)$$

Fig. 3. Opportunistic Data Aggregation. Low rate streams are aggregated into the high rate stream.

Fig. 4. The most contentious node v requires 7a units. If link bandwidth is B units, then a=B/7.

IV. HEURISTICS

The mathematical model defined by objective (1) and inequalities (2a) - (2e) considers the bandwidth constraint while optimizing sensor network lifetime, therefore the solution to this model contains the optimal solution to the energy-bandwidth constrained maximum lifetime routing problem. However it is not linear because f_i is also a variable. In the following, we will present two heuristics that both work around the nonlinear problem by using information from the shortest paths (in terms of hops) from sources to the sink. The shortest paths represent the minimum-energy routing topology if data is not aggregated [12]. Heuristic I bears the characteristics of the shortest path routing, and Heuristic II bears the characteristics of the mathematical-programming based optimal solution, and they both consider bandwidth constraints.

A. Heuristic I: Scalable Rate Allocation on Shortest Paths

The first heuristic starts from the shortest paths from sources to the sink, but the rate on each link is determined by the available bandwidth.

1) Compute the shortest path from each source node to the sink;
2) Assume source rate is one unit, check against condition (2c) for each node, and find the most bandwidth-contentious node i. Let LHS=required bandwidth of node i's collision domain. Then compute the scale factor a (see Fig. 4): $a = B/LHS$. Set $\Delta f = \min\{a/2, R_i\}$;
3) Push out Δf amount of flow from each source to the sink then update the remaining input flow $R_i' = R_i - \Delta f$ for each source i;
4) Repeat 5) - 7) until we push through R_i' for each source i or the network is fully saturated:
5) Find the shortest paths for nodes with $R_i' > 0$ based on the current available nodes and links. Nodes that are saturated on (2c) and their neighbors are not eligible for relaying. In case of a tie, give higher priority to nodes with more remaining energy; if there is still a tie, give higher priority to nodes with smaller degree;
6) Decide the scale factor a in a similar manner as in step 2). If pushing $\min\{a, R_i'\}$ units does not decrease lifetime, then set $\Delta f = \min\{a, R_i'\}$; otherwise, set $\Delta f = \min\{a/2, R_i'\}$;
7) Push out Δf amount of flow from each source with $R_i' > 0$ then update the remaining input flow $R_i' = R_i' - \Delta f$.
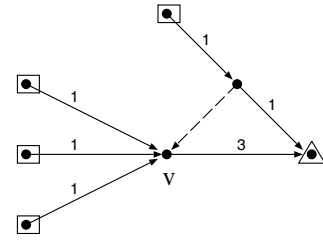
In steps 2) and 6), this algorithm uses $a/2$ when computing Δf for the purpose of load balancing, which makes the network last longer. A simplified version is to use a instead of $a/2$ when computing Δf. It finishes faster but leads to shorter lifetime.

B. Heuristic II: Optimizing Lifetime With Bandwidth Constraint

Since the mathematical model defined in (1) and (2a) - (2e) has an objective of maximizing lifetime, if we can convert it to a linear program in a controlled manner, it is likely to produce a close-to-optimal solution in terms of lifetime. The following describes an algorithm that chooses the likely-to-be relay nodes and set their $f_i = 1$ to make the program linear.

Heuristic II will either terminate with a valid solution or report "infeasible". There will not be endless iterations in line 4. If the given source rates R_i are very low, it terminates at line 1. In most of the simulations, it requires solving the linear program two to four times to get a suboptimal solution. If it does become infeasible, it is likely because the given source rates R_i are higher than what the network can support.

Heuristic II is presented as follows:

1) Set $f_i = 1$ for sink and $f_i = 0$ for all other nodes; solve the LP; update $f_i = 1$ if $\sum_{j \in N_i} R_{ji} > 0$; if (2c) is satisfied $\forall i$, return link rates R_{ij}; otherwise, go to line 2.
2) Compute shortest paths from sources to the sink.
3) Set $f_i = 1$ for receiving nodes; solve the LP; if $\sum_{j \in N_i} R_{ji} > 0$ and $f_i = 0$, update $f_i = 1$.
4) Repeat line 3 until there is no update for f_i or the linear program becomes infeasible.
5) If it converges, output R_{ij} for all links (i, j).
6) If it is infeasible: if $\sum_{j \in N_i} R_{ji} = 0$ and $f_i = 1$, set $f_i = 0$ and $R_{ji} = 0, \forall j \in N_i$, solve the linear program again; if it is still infeasible, report infeasible.

V. SIMULATION

A. With Uniform Transmission Power

In the simulation study, we investigate how the bandwidth constraint can change the routing decision and eventually affect the lifetime of the sensor network. First, we compare the existing algorithms with the proposed heuristics and observe which algorithm is more likely to cause network congestion and fail to push through the applied load.

Nodes are randomly deployed in a 100×100 square region, and transmission range is set to 30. In the first simulation (Fig. 6(a) and (c), we use 50 nodes in total. We randomly select 4 source nodes and apply increasing source rate on them. Source rate is set to be a percentage of link bandwidth. The proposed schemes Heuristic I and II are compared with MaxLife from [1], and shortest path routing (labeled as SPR). The reason we choose MaxLife is because it computes the maximum lifetime without considering bandwidth constraint. When there is enough bandwidth, MaxLife represents the optimal solution. SPR uses the shortest paths from sources to the sink, with link weight representing the transmission power of the node. In the uniform transmission power setup, each link has weight 1.

We found that when each source node's data rate R_i is increased to $12\% \sim 13\%$ of the given link bandwidth, MaxLife starts to congest, i.e., some collision domain requires more bandwidth than what is available, and SPR starts to congest when it is increased to 15%. Heuristic I can push through without congestion until the load is increased to 18% and Heuristic II can support as much as 16%. The vertical lines in Fig. 6(a) and (c) indicate after this point, increased data rate cannot be put through. Fig. 6(a) shows the average ratio of the required bandwidth in each collision domain to the offered bandwidth. The lower the average, the more bandwidth efficient of the scheme. Fig. 6(c) shows the maximum ratio. A scheme stops working when the maximum ratio reaches 1. We can get the maximum throughput of the network at the stop point.

Fig. 6(a) shows which scheme is more bandwidth efficient from a different angle. If a routing scheme violates the necessary condition, there is absolutely no way to push through the applied traffic load; when it violates the sufficient condition, there is no guarantee we can find a valid transmission schedule at the MAC layer to support the routing.

In the second simulation (Fig. 6(b) and (d), we show the results with 100 nodes and 10 source nodes. The four algorithms show similar behavior as in the first simulaiton, except that the per node throughput is lower because there are more source nodes. The total throughput of the network is close to that in the first simulation.

In Fig. 7 we compare four algorithms on their contribution toward lifetime. The results show when there is enough bandwidth, MaxLife does not have bandwidth violations and achieves the optimal solution, and Heuristic II achieves the same optimal solution; However when bandwidth does pose a constraint, Heuristic II can still push through 33% more data than MaxLife, and Heuristic I can push through 50% more data than MaxLife. Heuristic II achieves the best performance on lifetime and second best on throughput; Heuristic I achieves the best performance on throughput, which is consistent with our observation from Fig. 6. Networks with 100 nodes can achieve longer lifetime than networks with 50 nodes because the workload is shared among more nodes.

B. Nonuniform Transmission Power

In this simulation, transmission range is randomly selected between 25-35. With asymmetrical edges, the performance comparison of the four algorithms in Fig. 9 is consistent

with the uniform power case in Fig. 6. Network lifetime is reduced because the disparity in energy consumption is severe. Since the nonuniform power distribution is captured in the optimization model given in section III-C, as a result, Heuristic II shows more performance gain in `lifetime` over other algorithms.

C. With Double Disk Model

In this simulation, we choose transmission range 30, interference range $1.7 \times$ transmission range, with everything else the same as in section V-A. Fig. 10(a)-(c) show the throughput performance. With a larger interference range, there is less chance for spatial reuse of channel, therefore the network throughput is less, but the lifetime is increased due to the lower data rate as shown in Fig. 10(d).

D. Data Aggregation

In this simulation, we test how much improvement we can achieve through mathematical optimization on a chosen data aggregation method. Using the opportunistic aggregation method outlined in section III-E, we compare our solution with the shortest path tree and the minimum spanning tree, and the results show dramatic improvement on network lifetime as shown in Fig. 8. LP-SPT results from applying Heuristic II using an initial shortest path tree at step 2, and LP-MST results from applying Heuristic II using an initial minimum spanning tree at step 2. SPT and MST are fixed-route aggregation on the shortest path tree and the minimum spanning tree respectively.

Fig. 8(a) shows that LP-SPT and LP-MST can push data through until source rate is 20% of link bandwidth, while SPT and MST stop working (due to congestion) when source rates are 15% and 17% of link bandwidth respectively. This indicates a throughput gain of 33% over SPT and 17% over MST.

In Fig. 8(b), we use networks of different sizes to show the maximum network throughput. Each source sends at a rate $0.01 \leq R_i \leq 10$ with link bandwidth=10, and we try to maximize $\sum R_i$. Our observation is consistent with that in Fig. 8(a) — LP-MST and LP-SPT have the same throughput, and both are better than MST and SPT.

VI. RELATED WORK

The most related work includes our previous work on edge coloring for transmission scheduling [13], maximum lifetime routing [1], and throughput optimization [14]. In [13], we precisely depicted the conflict relation among transmissions with each color corresponding to one time slot at the MAC layer. It guarantees conflict-free time slot assignment if each edge carries the same load. However, edge coloring by itself is NP-complete, and it assigns one color to each edge which implies that it works best for uniform traffic load. Link rate allocation in this article is an extension from color assignment, and it works well for arbitrary traffic load because the number of time slots that each edge gets is proportional to the traffic load on the edge; Furthermore, we consider nodes' energy constraint for link rate allocation. In [14], a linear programming model is used to optimize system throughput subject to the
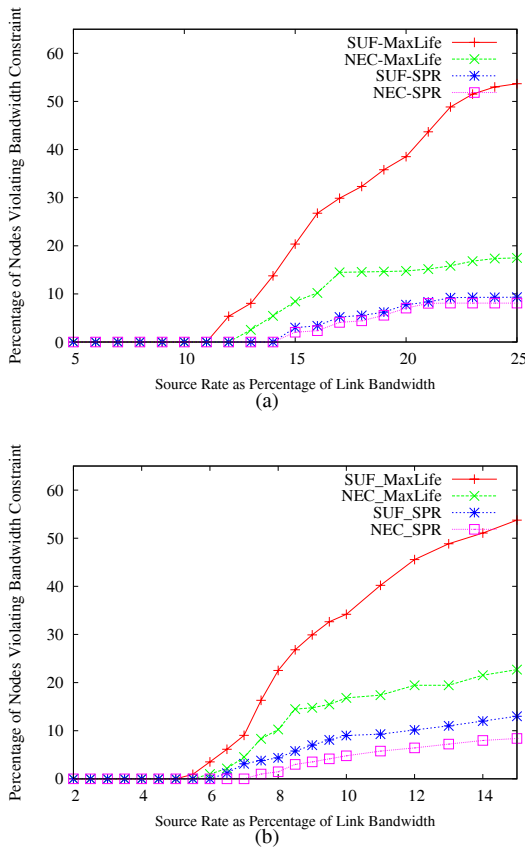
Fig. 5. Percentage of nodes violating necessary and sufficient conditions. (a) 50 nodes; (b) 100 nodes

fairness constraint. In this paper, energy is not considered as a constraint, and a network flow model is used that characterizes the capacity constraint: $f(e) \leq c(e)$ on a link e, instead of using the accurate capacity constraint on a collision domain as discussed in this paper. An earlier work [15] also falls in this category and only considers a very simple interference model: when a node sends, it cannot receive. In [1], the authors proposed a distributed algorithm to compute link rates with an objective of maximizing the network lifetime. The major contribution is on the distributed implementation of the optimization algorithm. However, like most previous work on energy efficient routing in sensor networks, bandwidth is not taken into consideration in their model. Similar work along this line includes [2]–[10] and many others.

In [2], the proposed routing algorithms select the routes and the corresponding power levels such that the network lifetime is maximized. In [3], the routing problem is formulated as a linear programming problem, where the objective is to maximize the network lifetime, which is equivalent to the time until the network partition due to battery outage. Packet aggregation techniques were proposed to further reduce the energy consumption rate [4], [5], [7]. In [6], it was proposed to deploy a network clustering scheme and assign a less-energy-constrained gateway node to act as a centralized network manager to further improve the energy efficiency and maximize network lifetime. Cui et al. further considered energy-efficient routing, scheduling, and link adaptation strategies together to maximize the network lifetime in [8], but the authors did not
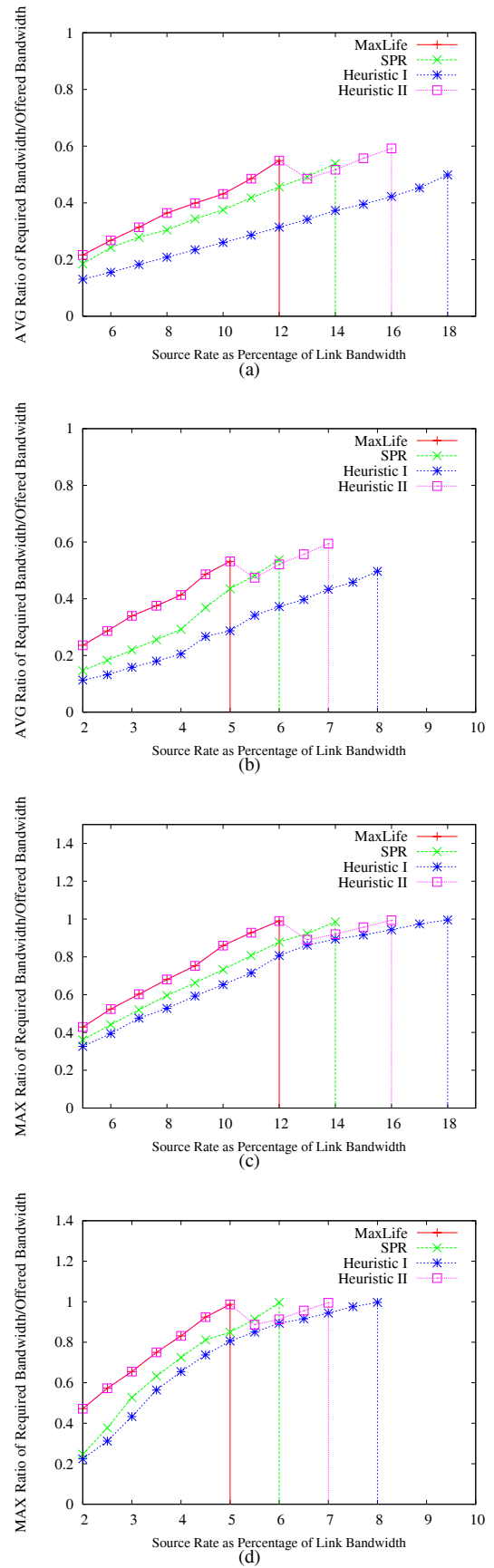

Fig. 6. (a)-(b) The average ratio of required bandwidth/offered bandwidth; (c)-(d) The maximum ratio of required bandwidth/offered bandwidth. (a) and (c) for 50 nodes, (b) and (d) are for 100 nodes.
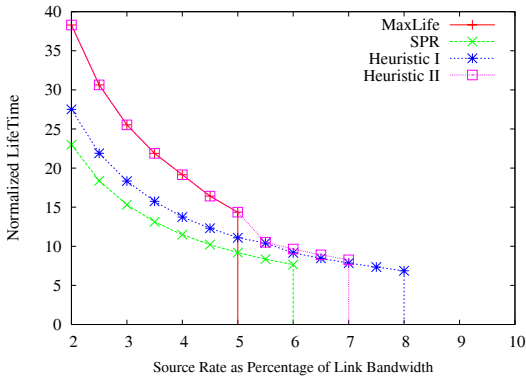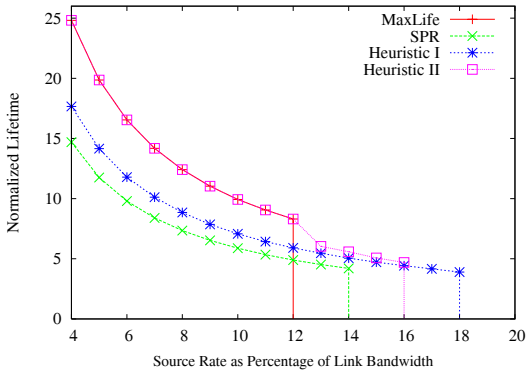
Fig. 7. Normalized lifetime for data forwarding without aggregation, assuming sending one unit of data consumes 10% total energy. (a) 50 nodes; (b) 100 nodes.

Fig. 8. Opportunistic aggregation for data forwarding, assuming sending one unit of data consumes 10% total energy. (a) normalized lifetime; (b) throughput.
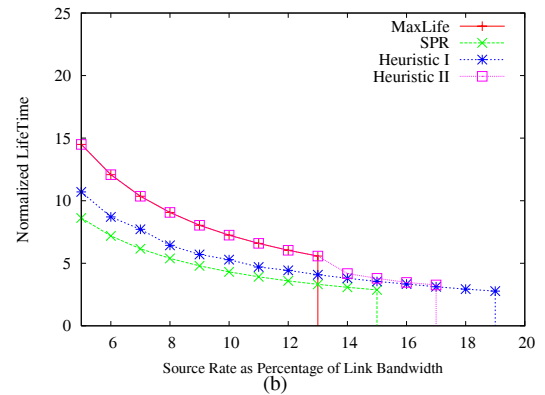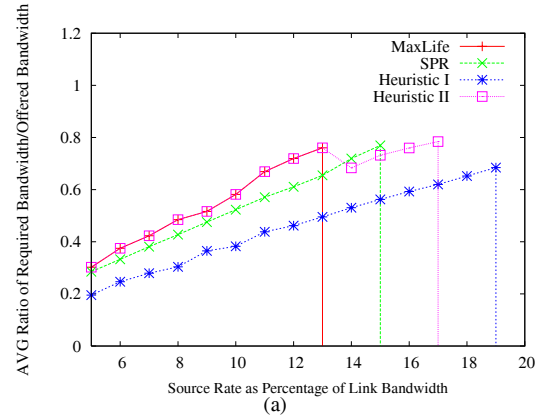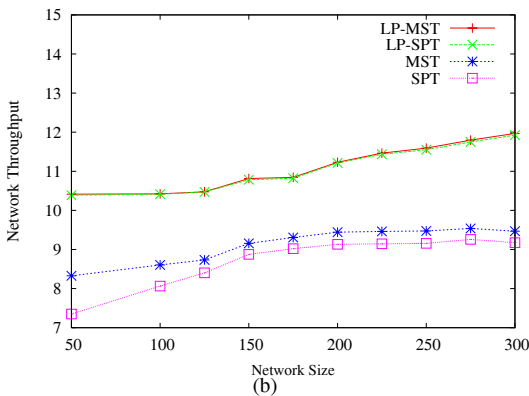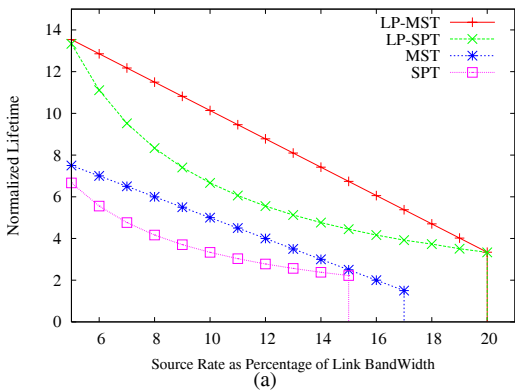

Fig. 9. With nonuniform transmission power, (a) the average ratio of required bandwidth/offered bandwidth; (b) normalized lifetime.

explicitly consider the bandwidth constraint in an arbitrary topology as we do. How to arrange the location of base-stations for WSN and select relay paths to maximize the network lifetime was discussed in [9], [10].

Along the direction of cross-layer design and optimization, we found [8], [16]–[21] and many others. Optimization problems in multihop wireless networks are naturally cross-layer problems ([16], [17]). It involves PHY layer coding, modulation and error control, MAC/link layer resource (both bandwidth and power) management, network layer routing, and transport layer flow and congestion control. Many of the related work in cross-layer design focused on how to minimize energy consumption under various constraints [8], [18]–[20]. Reference [18] proposed to adjust the transmission powers of nodes in a multihop wireless network to create a desired topology, aimed to minimize power used while maintaining network connectivity. Cruz and Santhanam studied the problem of joint routing, link scheduling and power control to support high data rates for broadband wireless multihop networks in [19]. The main objective is still to minimize the total average transmission power. Since most cross-layer optimization problems are too complex to solve, distributed algorithms with suboptimal (and potentially distributed) scheduling component were studied in [16], [20].

Although this paper aims to provide maximum lifetime routing under energy and bandwidth constraints, the resulting solution naturally satisfies guaranteed data rate for each source and hence guaranteed fairness. Previous works ([21]–[23])
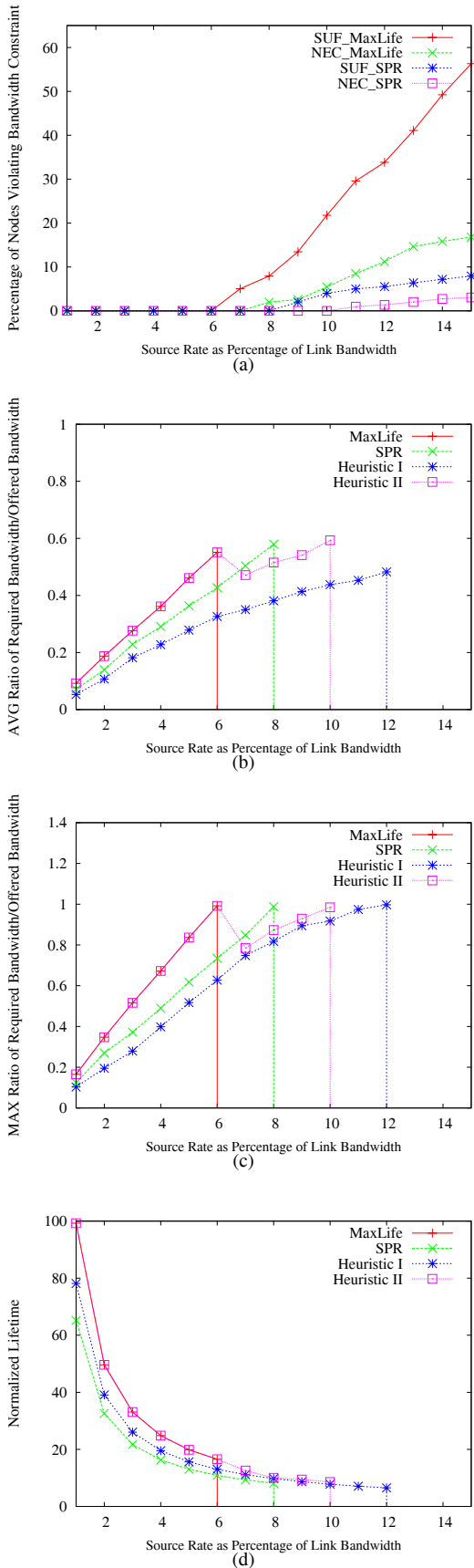
Fig. 10. With double disk model, (a) percentage of nodes violating necessary and sufficient conditions; (b) the average ratio of required bandwidth/offered bandwidth; (c) the maximum ratio of required bandwidth/offered bandwidth; (d) normalized lifetime.

addressed the fairness issue through different mechanisms, such as packet scheduling, distributed layer-2 fairness solution (by modifying the contention and back-off mechanisms of CSMA/CA), joint power allocation and routing etc.

VII. CONCLUSION AND DISCUSSION

This article has provided a generic mathematical model for the maximum lifetime routing problem in energy and bandwidth-constrained sensor networks. Using the sole energy constraint sometimes leads to unrealistic solutions that cannot be accommodated by the link capacity. In this paper we have provided a sufficient condition that a given traffic load can be put through a given network and jointly considered energy and bandwidth constraints for routing and link rate allocation. Joint optimization guarantees that there exists a conflict-free time slot assignment to support the given routing solution. To the best of our knowledge, this is the first work that explicitly considers bandwidth constraint in solving a maximum lifetime routing problem in sensor networks. The basic mathematical optimization model can be easily extended to address heterogeneous sensor networks where nodes have different initial energy or different transmission power levels, and to work with various data aggregation schemes.

The proposed heuristics are centralized. To apply mathematical optimization on large scale sensor networks, hierarchical scheme can be used, such as to divide the network into areas or clusters, and then apply the algorithms within the area or cluster. This will compromise the global optimality, but the solution is still better than the pure discentralized algorithms in terms of energy and channel efficiency.

VIII. ACKNOWLEDGEMENT

APPENDIX: PROOF OF THE SUFFICIENT CONDITION

To prove that the condition in Section II is sufficient for collision free communication, we first introduce a time slot assignment algorithm. The algorithm requires that input link rates satisfy the sufficient condition and outputs a conflict-free schedule.

A. A Slot Assignment Algorithm

SlotAssignment(G(V,E), R)

1) Scale the link rates R_{ij} to integers and scale B proportionally; Let slot size $\tau = 1$.
2) Find the most bandwidth-contentious node v according to the sufficient condition, and compute the required bandwidth B_v at node v's collision domain:
$$v = argmax_{i \in V} \left(\sum_{j \in N_i} R_{ij} + f_i \cdot \sum_{j \in N_i} \sum_{k \in N_j} R_{jk} \right); \text{ and}$$
$$B_v = \sum_{j \in N_v} R_{vj} + f_v \cdot \sum_{j \in N_v} \sum_{k \in N_j} R_{jk}$$
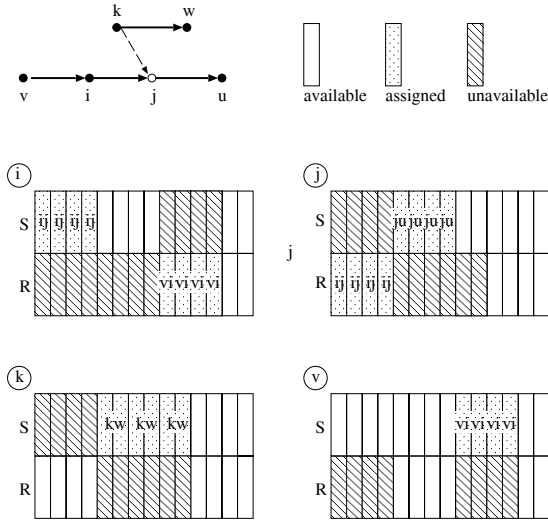3) Let frame size $F = B_v$. Number the slots from 1 to F.

Fig. 11.   A walk-through example for the SlotAssignment algorithm. Suppose $R_{vi} = R_{ij} = R_{ju} = 4$, $R_{kw} = 6$, so node j is the most bandwidth-contentious node; frame size F=14 slots; the order that nodes are randomly picked at step 6 is i, j, k, v.

4) Create a table of $2 \times F$ associated with each node's sending and receiving schedules, use S row for sending and R row for receiving.

5) Let $L = V$. Repeat the following until $L = \phi$:
   (a) Randomly pick a node i from L;
   (b) For each node $j \in N_i$, if $R_{ij} > 0$, assign R_{ij} slots to link (i, j), starting from the smallest available slot. A slot is available if it is available in both the S row of table[i] and the R row of table[j]; Mark those slots unavailable in the S row of table[j]; For each $k \in N_j$, if $k \neq i$, mark those slots unavailable in the S row of table[k];
   (c) Mark those slots unavailable in the R row of table[i];
   (d) For each node $j \in N_i$, mark those slots unavailable in the R row of table[j], if they are not previously assigned;
   (e) Remove i from L.

6) Update frame size F to be the largest slot number used.

In Fig. 11, the sufficient condition requires F=14 slots, but actually it only needs 12 slots by allowing the transmissions on (k, w) and (j, u) to occur at the same time. The sloppiness in the sufficient condition guarantees no matter whether there is a link between (j, w) or not, there are always enough slots to use regardless of the order that nodes are picked. This property makes it easy to implement the algorithm in a localized and distributed manner.

Lemma 1: The SlotAssignment algorithm generates a collision-free schedule.

Proof: Lemma 1 has two folds:

1) There are always sufficient number of slots to use, i.e., at step 5(b), the number of available slots $\geq$ the number of slots needed for any node i being considered, and
2) The resulting schedule is collision-free.

The second statement is obvious because all conflicting transmissions are scheduled at different time— when i is sending to j, j is not sending, and other neighbors of j are not sending, so there is no collision at j according to step 5(b);
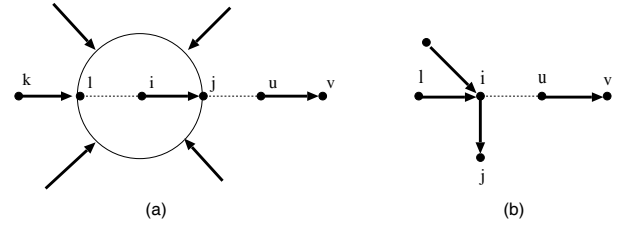

Fig. 12.   (a) with $f_i = 0$; (b) with $f_i = 1$

i is not receiving according to 5(c) so there is no collision at i; other neighbors of i are not receiving according to 5(d) so there is no collision at i's neighbors.

The first statement is proved as follows. Let N_1 be the total number of slots that are needed for sending when a random node i is picked at step 5(a), so $N_1 = \sum_{j \in N_i} R_{ij}$, and let N_2 be the number of slots that are still available for sending at this time.

- Case (1), when i is not a receiver ($f_i = 0$): the only reason that i's S row is marked unavailable is when a neighbor l is receiving from another node k (Fig. 12(a)). Let $C = \{(k, l)\}$ be the maximum set of such conflicting transmissions, so the total unavailable slots in i's S row is $\sum_{(k,l) \in C} R_{kl}$. Similarly, for each receiver node j of i, the only reason that the R row of j is marked unavailable is because j's neighbor u is transmitting. Transmissions on (k, l) and (u, v), if not conflicting with each other, can be arranged at the same slot. Therefore, as long as the sufficient condition holds at node l with $f_l = 1$ and at node j with $f_j = 1$, the number of available slots N_2 for i's transmission is still $\geq \sum_{j \in N_i} R_{ij}$. Therefore, $N_2 \geq N_1$ is held.

- Case (2), when i is a receiver ($f_i = 1$): from case (1) to case (2), there will be $\sum_{l \in N_i} R_{li}$ additional slots marked unavailable in the S row of i, according to step 5(b); others remain unchanged. As long as the sufficient condition holds at node i with $f_i = 1$, the number of available slots N_2 for i's transmission is still $\geq \sum_{j \in N_i} R_{ij}$. Therefore $N_2 \geq N_1$ is held. Because during the iteration in step 5, $F = \max_i \{ \sum_{j \in N_i} R_{ij} + f_i \cdot \sum_{j \in N_i} \sum_{k \in N_j} R_{jk} \}$, so N_2 is sufficient for any node i.

Next we will see that even though the sloppiness of the sufficient condition requires more slots than necessary, the SlotAssignment algorithm itself does not prevent non-conflicting transmissions from happening at the same time.

Lemma 2: The SlotAssignment algorithm can completely avoid the exposed terminal problem.

Proof:

In Fig. 13, if node B is picked first by the algorithm to use the first slot, transmission on (C, D) can still use the first slot because B's transmission in slot 1 only marked the R row of node C unavailable, the S row is still available. If node C is picked first, the result is the same.

Theorem 1: The following condition is sufficient to have a TDMA schedule that completely avoids collision and the exposed terminal problem in a multihop wireless networks
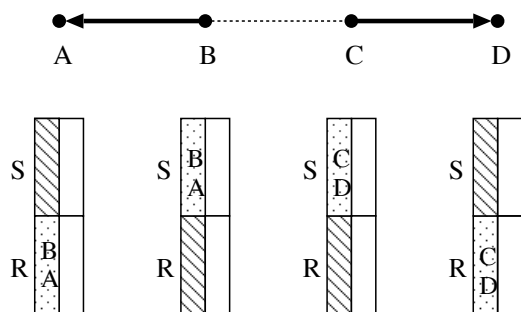
Fig. 13.   The `SlotAssignment` algorithm would allow $C \rightarrow D$ and $B \rightarrow A$ to occur at the same time.

with omnidirectional antenna:

$$\sum_{j \in N_i} R_{ij} + f_i \cdot \sum_{j \in N_i} \sum_{k \in N_j} R_{jk} \leq B, \ \forall i$$

Proof: By constructive proof, using the `SlotAssignment` algorithm described above, we can always find a TDMA schedule that is collision-free (by Lemma 1) and completely avoids the exposed terminal problem (by Lemma 2), as long as the given input R_{ij} satisfies $\sum_{j \in N_i} R_{ij} + f_i \cdot \sum_{j \in N_i} \sum_{k \in N_j} R_{jk} \leq B, \ \forall i$.

REFERENCES

[1] R. Madan and S. Lall, "Distributed algorithms for maximum lifetime routing in wireless sensor networks," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2004, pp. 748–753.

[2] J. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in *Proc. IEEE INFOCOM 2000*, vol. 1, 2000, pp. 22–31.

[3] ——, "Maximum lifetime routing in wireless sensor networks," *IEEE/ACM Trans. Networking*, vol. 12, no. 4, pp. 609– 619, Aug. 2004.

[4] C. Schurgers and M. B. Srivastava, "Energy efficient routing in wireless sensor networks," in *Proc. MILCOM 2001*, vol. 1, 2001, pp. 357–361.

[5] M. Bhardwaj, T. Garnett, and A. Chandrakasan, "Upper bounds on the lifetime of sensor networks," in *Proc. IEEE ICC'01*, vol. 3, 2001, pp. 785–790.

[6] M. Younis, M. Youssef, and K. Arisha, "Energy-aware routing in cluster-based sensor networks," in *Proc. IEEE MASCOTS 2002*, 2002, pp. 129–136.

[7] B. Krishnamachari, D. Estrin, and S. Wicker, "Modelling data-centric routing in wireless sensor networks."

[8] S. Cui, R. Madan, A. Goldsmith, and S. Lall, "Joint routing, MAC, and link layer optimization in sensor networks with energy constraints," in *Proc. IEEE ICC'05*, vol. 2, May 2005, pp. 725–729.

[9] I. Kang and R. Poovendran, "Maximizing static network lifetime of wireless broadcast ad hoc networks," in *Proc. IEEE ICC'03*, vol. 3, 2003, pp. 2256–2261.

[10] J. Pan, T. Hou, L. Cai, Y. Shi, and X. Shen, "Topology control for wireless sensor networks," in *Proc. ACM MOBICOM'03*, 2003, pp. 286–299.

[11] S. Katti, D. Katabi, W. Hu, H. Rahul, and M. Medard, "The importance of being opportunistic: practical network coding for wireless environments," in *Proc. 43rd Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sept. 2005.

[12] M. X. Cheng and L. Yin, "Energy-efficient data gathering algorithm in sensor networks with partial aggregation," *Int. J. Sensor Networks,*, 2008.

[13] M. Cheng and L. Yin, "Transmission scheduling in sensor networks via directed edge coloring," in *Proc. IEEE ICC 2007*, pp. 3710–3715.

[14] M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," in *Proc. ACM MobiCom*, 2005, pp. 58–72.

[15] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problem," in *Proc. ACM MobiCom*, 2003, pp. 42–54.

[16] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, 2006.

[17] S. Toumpis and A. J. Goldsmith, "Performance, optimization, and cross-layer design of media access protocols for wireless ad hoc networks," in *Proc. IEEE ICC'03*, vol. 3, 2003, pp. 2234–2240.

[18] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit poweradjustment," in *Proc. IEEE Infocom'00*, vol. 2, 2000, pp. 404–413.

[19] R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multihop wireless networks," in *Proc. IEEE Infocom'03*, vol. 1, Mar. 2003, pp. 702–711.

[20] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proc. IEEE Infocom'05*, vol. 3, Mar. 2005, pp. 1804–1814.

[21] M. Johansson and L. Xiao, "Scheduling, routing and power allocation for fairness in wireless networks," in *Proc. IEEE VTC 2004-Spring*, vol. 3, May 2004, pp. 1355–1360.

[22] H. Luo, S. Lu, and V. Bharghavan, "A new model for packet scheduling in multihop wireless networks," in *Proc. ACM Mobicom 2000*, pp. 76–86.

[23] V. Gambiroza, B. Sadeghi, and E. W. Knightly, "End-to-end performance and fairness in multihop wireless backhaul networks," in *Proc. ACM Mobicom'04*, pp. 287–301.

Maggie Cheng is an Assistant Professor in the Department of Computer Science at Missouri University of Science and Technology. She obtained her Ph.D. in Computer Science from the University of Minnesota-Twin Cities in 2003. Her research interests are in the areas of Computer Networks, Wireless Ad Hoc Networks and Sensor Networks. She is on the Editorial Board of INTERNATIONAL JOURNAL OF SENSOR NETWORKS, and has been on the executive committee of IEEE International Performance, Computing, and Communications Conference, International Conference on Nano-Networks, and has been on the technical program committee of several major conferences in communications and networks including INFOCOM, ICC and Globecom.

Xuan Gong obtained his Master degree in Computer Science from Missouri University Science & Technology in 2006. He is currently a Ph.D. student under the supervision of Dr. Maggie Cheng. His research interests are wireless networks, ad hoc and sensor networks.

Lin Cai (S'00-M'06) received the M.A.Sc. and Ph.D. degrees (with Outstanding Achievement in Graduate Studies Award) in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 2002 and 2005, respectively. Since July 2005, she has been an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Victoria, British Columbia, Canada. Her research interests span several areas in wireless communications and networking, with a focus on network protocol and architecture design supporting emerging multimedia traffic over wireless, mobile, ad hoc, and sensor networks. She serves as the Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2007-), EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING (2006-), and INTERNATIONAL JOURNAL OF SENSOR NETWORKS (2006-).