Utility-Lifetime Trade-off in Self-regulating Wireless Sensor Networks: A Cross-Layer Design Approach

Hithesh Nama
WINLAB, Rutgers University
Email: hithesh@winlab.rutgers.edu

Mung Chiang
Princeton University
Email: chiangm@princeton.edu

Narayan Mandayam
WINLAB, Rutgers University
Email: narayan@winlab.rutgers.edu

***Abstract*— The performance of wireless sensor network applications is often dependent up on the amount of data collected by the individual sensors and delivered to a set of collectors through multihop routing within the network. However, the energy-constrained nature of the nodes limits the operational lifetime of the network since energy is dissipated both in sensing and in communicating data across the network. There is thus an inherent trade-off in simultaneously maximizing the network lifetime and the application performance (characterized as network utility). In this paper, we characterize this trade-off by considering a cross-layer design problem in a wireless sensor network with orthogonal link transmissions. We compute an optimal set of source rates, network flows, and radio resources at the transport, network, and radio resource layers respectively, while jointly maximizing the network utility and lifetime. Using dual decomposition techniques, we show that the cross-layer optimization problem decomposes vertically into three subproblems - a joint transport and routing problem, a radio resource allocation problem, and a network lifetime maximization problem, all of which interact through the dual prices for capacities of links and battery capacities of nodes.**

I. Introduction

We consider a wireless sensor network deployed in an information field. Each node in the network extracts data from the field and delivers it to a gateway node through multiple hops over intermediate nodes. All the nodes are battery powered and dissipate energy in sensing, transmitting, and receiving data. Typically, the miniature size of the sensor nodes allows very limited energy storage capabilities and in some applications it might even be infeasible to replace the batteries on these nodes. Thus energy is a scarce resource in wireless sensor networks and it is imperative that energy-aware protocols be designed across different layers of the protocol stack in order to prolong the operational lifetime of the network.

In certain sensor network applications, the application performance depends strongly on the amount of data gathered from each sensor node in the network. For instance, more data from a video sensor could mean a better quality image while high-precision samples of a temperature field result in a more accurate characterization of the physical process. However, higher data rates result in greater sensing and communication costs across the sensor network resulting in reduction in the overall network lifetime. There is thus an inherent trade-off between the application layer performance and the network lifetime.

This work is funded in part by the NJ Commission on Science and Technology under the MUSE project.

In this paper, we characterize the application layer performance using a generic utility function that monotonically increases with the data rate. Using an optimization framework, we study the cross-layer design problem that computes an optimal set of source rates, network flows, and radio resources at the transport, network, and radio resource layers respectively, while jointly maximizing the network utility and lifetime.

Energy-aware protocols for ad hoc wireless networks in general [1], [2] and wireless sensor networks in particular [3], [4] have been studied extensively in recent years. In [5] the lifetime of a network is defined as the time until the death of the first node and the problem of optimal routing is considered using a network flow based approach. In Section II-D we use a similar definition for network lifetime and consider the lifetime maximization problem with routing and radio resource constraints.

Cross-layer design of transport, network, and radio resource layers for maximizing the throughput or network utility of ad hoc wireless networks has been studied in [6], [7], [8]. Using dual decomposition techniques, [7] and [8] show that the cross-layer design problems decompose vertically into subproblems that interact through link congestion prices. In our work, we additionally consider the effect of power dissipation at the nodes on the network lifetime and utility. A cross-layer design approach to maximize the network lifetime in the context of interference-limited wireless sensor networks is considered in [9]. Since nodes are assumed to have fixed source rates, it is likely that the network cannot sustain these rates for the given system resource constraints. In our problem we consider a self-regulating network in which nodes can adapt their source rates so that the network operates at an optimal set of source rates that jointly maximizes the network utility and lifetime.

A main contribution of our work is the development of a unifying framework to understand the trade-off between the application layer performance and the lifetime of a sensor network. To our knowledge there has been no prior work that studies this important trade-off in sensor networks using a rigorous framework as has been done in our work.

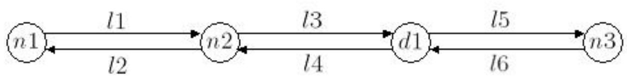

Fig. 1. Connectivity Graph

Section II describes the system model and states the separate problems of network utility maximization and network lifetime maximization. We pose the network utility-lifetime trade-off problem in Section III. Section IV describes a dual decomposition based solution of the cross-layer design problem. We present simulation results in Section V to illustrate the utility-lifetime trade-off in sensor networks.

II. System Model

We consider a wireless sensor network with orthogonal link transmissions which could include FDMA based networks or spread spectrum based networks with transmissions using orthogonal codewords. The network consists of a set of sensor nodes denoted by $\mathcal{N}$ and a set of destination nodes (also called collectors or sinks) denoted by $\mathcal{D}$. The sensor nodes are the sources that collect data from the information field and deliver it to any of the collectors, possibly over multiple hops. Nodes are capable of varying their source rates which allows the application performance to be tuned and also helps avoid any congestion in the network.

We model the sensor network using a connectivity graph, $G(\mathcal{V}, \mathcal{L})$, where the vertex set, $\mathcal{V} = \mathcal{N} \bigcup \mathcal{D}$, includes both the sensor nodes and the destination nodes. The set of edges, $\mathcal{L}$, represents logical bidirectional communication links between nodes. Fig. 1 shows the connectivity graph for a wireless sensor network consisting of three sensor nodes ($n1, n2, n3$), one destination node ($d1$), and six logical links ($l1, \ldots, l6$).

The set of outgoing links and the set of incoming links corresponding to a node n are denoted by $\mathcal{O}(n)$ and $\mathcal{I}(n)$ respectively. Thus, in Fig. 1, $\mathcal{O}(n2) = \{l2, l3\}$ and $\mathcal{I}(n2) = \{l1, l4\}$.

A. Routing and Source Rate Control: A multi-commodity flow approach

We model source rate adaptation and routing of data in the network using multi-commodity flows that are distinguished based on their destination nodes. Let r_n^d denote the average non-negative source rate from sensor node $n \in \mathcal{N}$ to destination (or collector) $d \in \mathcal{D}$. Apart from being sources of data, nodes also route data from other nodes toward the collectors. The average non-negative flow over link l toward destination d is given by f_l^d. Data transmission is assumed to be lossless and there is no compression of data at any node inside the network. Thus each flow commodity must satisfy the flow conservation constraint (1) at all sensor nodes.

$$\sum_{l \in \mathcal{O}(n)} f_l^d - \sum_{l \in \mathcal{I}(n)} f_l^d = r_n^d, \quad \forall\, d \in \mathcal{D}, n \in \mathcal{N} \tag{1}$$

The net average flow on link l is given by

$$f_l = \sum_{d \in \mathcal{D}} f_l^d. \tag{2}$$

Since flows terminate at the collectors, there are no outgoing flows from any collector d. The average rate at which information is gathered by a collector is equal to the net flow on all its incoming links, i.e., $\sum_{l \in \mathcal{I}(d)} f_l$.

B. Radio resource allocation

The physical and MAC layers of the protocol stack are concerned with allocation of radio resources like power, bandwidth, and the fraction of time a link is in operation. We use the notion of transmission modes to characterize the radio resource allocation problem. Each transmission mode m corresponds to a distinct set of links that are scheduled simultaneously. There are thus a maximum of $2^{|\mathcal{L}|}$ possible transmission modes, where $|\mathcal{L}|$ denotes the cardinality of set $\mathcal{L}$. However, not all of these $2^{|\mathcal{L}|}$ transmission modes are *feasible*.

Like in most practical transceivers, we assume that a node can at most operate on a single communication link at any time, i.e., it can either transmit on an outgoing link or receive on an incoming link. Hence, a feasible mode can only consist of independent links i.e., links that are not connected to the same node. Conversely, in our system model, every set of independent links forms a feasible mode because these links do not interfere with each other due to orthogonal transmissions. Thus, in the network of Fig. 1 there are eleven feasible modes: $\{l1, l5\}$, $\{l1, l6\}$, $\{l2, l5\}$, $\{l2, l6\}$, $\{l1\}$, $\{l2\}$, $\{l3\}$, $\{l4\}$, $\{l5\}$, $\{l6\}$, and the idle mode with no link in operation.

An indicator function $1^m(l)$ is used to indicate whether a particular link l is active in mode m. If active, $1^m(l) = 1$ else it is zero. We use P_l^m to denote the transmit power of link l in mode m and $\mathbf{P}^m \in \mathbf{R}^{|\mathcal{L}|}$ to denote the vector of link powers in mode m. Note that $1^m(l) = 1$ iff $P_l^m > 0$. We assume that each active link uses a fixed bandwidth W during transmission. Further, the transmit powers of links must satisfy a peak power constraint given by

$$P_l^m \leq P_l^{max}, \forall\, l \in \mathcal{L},\, m \in \mathcal{M}. \tag{3}$$

Since we assume orthogonal link transmissions, the capacity C_l^m of an active link l in mode m depends only on the transmit power of that link i.e.,

$$C_l^m(P_l^m, W) = W \log_2(1 + \frac{P_l^m K d_l^{-\alpha}}{N_0 W}),$$

where K is a constant that depends up on the transmission frequency, d_l is the separation between the transmitter and receiver on link l and α is the path-loss exponent. It can be seen that the link capacity is a concave function of the link power P_l^m. The fraction of time for which each mode is in operation is denoted by ν_m, so that

$$\nu_m \geq 0, \forall\, m \in \mathcal{M}; \sum_{m \in \mathcal{M}} \nu_m = 1. \tag{4}$$

The fraction of time for which each link l is in operation can then be computed as $\sum_{m | 1^m(l)=1} \nu_m$. For a given set of source rates $\{r_n^d, \forall n \in \mathcal{N}, d \in \mathcal{D}\}$ and network flows $\{f_l^d, \forall l \in \mathcal{L}, d \in \mathcal{D}\}$, the net average flow (2) on all links must satisfy the average link capacity constraint

$$f_l \leq \sum_{m \in \mathcal{M}} \nu_m C_l^m(P_l^m, W), \forall l \in \mathcal{L}. \tag{5}$$

C. Network utility maximization

In sensor networks, the application performance is often dependent on the amount of data gathered by the network. We characterize this dependency using a generic network utility function $U_n^d(r_n^d)$ corresponding to node n and destination d that is strictly concave and increasing. The network utility maximization problem can be formulated as:

$$\begin{aligned} &\max \sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} U_n^d(r_n^d) \\ &\text{subject to constraints (1) through (5).} \end{aligned} \tag{6}$$

The constraints in (6) represent a convex set if only one of either the link powers $\{P_l^m\}$ or the mode time-fractions $\{\nu_m\}$ are allowed to vary keeping the other constant. In such a case, the network utility maximization problem can be solved efficiently using convex optimization techniques. However, maximizing the network utility as above can lead to widely varying power dissipation levels across different nodes in the network. Such an allocation can result in shorter lifetimes for some nodes which could potentially lead to a disconnected network when the battery supplies of these nodes get exhausted.

D. Energy dissipation and network lifetime maximization

We now study the network lifetime maximization problem which results in a more uniform power dissipation across the nodes in the network. In a typical sensor network, sensor nodes have much tighter energy constraints than the collectors and hence we will focus only on the energy dissipated in the sensor nodes. We characterize the energy dissipation in each sensor by the energy consumed per bit during sensing ($\mathcal{E}_s$), in the receiver electronics ($\mathcal{E}_{rx}$), in the transmitter electronics ($\mathcal{E}_{tx}$), and the power radiated by the transmitter for reliable communication. The total average power dissipated in a node n is given by

$$\begin{aligned} P_n^{avg} = &\sum_{l \in \mathcal{O}(n)} \sum_{m \in \mathcal{M}} \nu_m P_l^m + \sum_{l \in \mathcal{O}(n)} f_l \mathcal{E}_{tx} \\ &+ \sum_{l \in \mathcal{I}(n)} f_l \mathcal{E}_{rx} + \sum_{d \in \mathcal{D}} r_n^d \mathcal{E}_s. \end{aligned} \tag{7}$$

If $E_n > 0$ denotes the initial energy of node n, then the lifetime of the node is given by $t_n = E_n / P_n^{avg}$. The network lifetime is defined as the time until the death of the first node i.e. $t_{nwk} = \min_{n \in \mathcal{N}} t_n$. The problem of maximizing the network lifetime can then be stated as

$$\begin{aligned} &\max_{t \geq 0} \ t \\ &\text{subject to constraints (1) through (5),} \\ &\text{and } P_n^{avg} \leq E_n/t, \forall n \in \mathcal{N}, \end{aligned} \tag{8}$$

where t denotes a lower bound on the node lifetimes. We introduce a lifetime-penalty function $F_n(1/t_n)$ for each node n to be a strictly convex and increasing function (e.g. $F(x) = x^2$), so that maximizing the node lifetime is equivalent to minimizing its lifetime-penalty function. Using a variable transformation $s = 1/t$ which denotes an upper bound on the inverse-lifetimes of nodes, the network lifetime maximization problem (8) can be equivalently stated as

$$\begin{aligned} &\min_{s \geq 0} \sum_{n \in \mathcal{N}} F_n(s) \\ &\text{subject to constraints (1) through (5),} \\ &\text{and } P_n^{avg} \leq E_n s, \forall n \in \mathcal{N}. \end{aligned} \tag{9}$$

It will be shown in Section IV-C that using a strictly convex function $F_n(s)$ simplifies the optimization problem by allowing analytical solutions.

As stated earlier, the constraints in (9) do not represent a convex set if both the mode time-fractions $\{\nu_m\}$ and powers $\{P_l^m\}$ are allowed to vary simultaneously. However if only one of them is allowed to vary while keeping the other constant, then the network lifetime maximization problem can be solved efficiently using convex optimization techniques.

Two cases corresponding to node source rates $\{r_n^d\}$ being either arbitrary constants or problem variables are of interest in the context of network lifetime maximization. In the former case, it might not be feasible to deliver any arbitrary set of source rates to the collectors due to congestion in the network. In the latter case, allowing the source rates to vary while maximizing the network lifetime would result in the trivial allocation of zero source rates throughout the network, which results in the worst case application performance. A cross-layer design problem that results in an optimal allocation of source rates to guarantee a desired application performance while also prolonging the lifetime of the network is studied in the next section.

III. Network utility-lifetime trade-off

As explained in the previous sections, there is an inherent trade-off between utility and lifetime in sensor networks. We introduce a system design parameter $\gamma \in [0, 1]$ that controls the desired trade-off between the network utility (6) and the network lifetime (9). We now pose the cross-layer design problem that computes an optimal set of source rates, network flows, and radio resources at the transport, network, and radio resource layers respectively, while jointly maximizing the network utility and lifetime.

$$\begin{aligned} &\max_{\{s, r_n^d, f_l^d, P_l^m\} \geq 0} \gamma \sum_{n \in \mathcal{N}} \sum_{d \in \mathcal{D}} U_n^d(r_n^d) - (1-\gamma) \sum_{n \in \mathcal{N}} F_n(s) \\ &\text{subject to constraints (1) through (5),} \\ &\text{and } P_n^{avg} \leq E_n s, \forall n \in \mathcal{N}. \end{aligned}$$

The above weighted utility-lifetime objective function is a concave function since $U(.)$ is concave and $F(.)$ is convex. The extreme case of $\gamma = 0$ and $\gamma = 1$ correspond to network lifetime maximization and network utility maximization respectively. The network designer could choose an appropriate value of γ that strikes the desired balance between the utility and lifetime of the network based on the application at hand.

IV. Dual Decomposition Based Solution

In this section, we consider the cross-layer design problem in which link powers $\{P_l^m\}$ are allowed to vary keeping the mode time-fractions $\{\nu_m\}$ fixed. The cross-layer optimization problem is stated in its entirety below.

$$\begin{aligned}
\max_{\{s,r_n^d,f_l^d,P_l^m\}\geq 0} \ & \gamma \sum_{n\in\mathcal{N}}\sum_{d\in\mathcal{D}} U_n^d(r_n^d) - (1-\gamma)\sum_{n\in\mathcal{N}} F_n(s) \\
\text{subject to } & \sum_{l\in\mathcal{O}(n)} f_l^d - \sum_{l\in\mathcal{I}(n)} f_l^d = r_n^d, \quad \forall\, d\in\mathcal{D}, n\in\mathcal{N} \\
& f_l = \sum_{d\in\mathcal{D}} f_l^d, \quad \forall\, l\in\mathcal{L} \\
& f_l \leq \sum_{m\in\mathcal{M}} \nu_m C_l^m(P_l^m, W), \quad \forall\, l\in\mathcal{L} \\
& P_l^m \leq P_l^{max}, \forall\, l\in\mathcal{L}, m\in\mathcal{M} \\
& P_n^{avg} \leq E_n s, \forall\, n\in\mathcal{N}
\end{aligned} \tag{10}$$

The constraint set in (10) represents a convex set and can be solved efficiently using dual decomposition techniques. Slater's condition for strong duality [10] requires that the non-linear inequalities in (10) be strictly feasible. If E_n and P_l^{max} are strictly positive, strong duality holds i.e. the optimal values of the primal and dual problems are equal. We can solve the primal problem (10) indirectly by first solving the dual problem and then recovering the desired primal variables $\{s, r_n^d, f_l^d, P_l^m\}$.

A. The Dual Problem

We state the dual problem by first formulating the partial Lagrangian dual function. It can be seen from (10) that the link capacity constraint and the average power constraint couple the optimization variables. We form the partial dual function by introducing Lagrange multipliers $\boldsymbol{\lambda}\in\mathbf{R}^{|\mathcal{L}|}$ and $\boldsymbol{\mu}\in\mathbf{R}^{|\mathcal{N}|}$ corresponding to these two inequality constraints. The partial dual function,

$$\begin{aligned}
& D(\boldsymbol{\lambda},\boldsymbol{\mu}) = \\
& \max_{\{s,r_n^d,f_l^d,P_l^m\}\geq 0} \gamma \sum_{n\in\mathcal{N}}\sum_{d\in\mathcal{D}} U_n^d(r_n^d) - (1-\gamma)\sum_{n\in\mathcal{N}} F_n(s_n) \\
& - \sum_{l\in\mathcal{L}} \lambda_l \left\{ f_l - \sum_{m\in\mathcal{M}} \nu_m C_l^m(P_l^m, W) \right\} \\
& - \sum_{n\in\mathcal{N}} \mu_n \left\{ P_n^{avg} - E_n s \right\} \\
\text{subject to } & \sum_{l\in\mathcal{O}(n)} f_l^d - \sum_{l\in\mathcal{I}(n)} f_l^d = r_n^d, \quad \forall\, d\in\mathcal{D}, n\in\mathcal{N} \\
& f_l = \sum_{d\in\mathcal{D}} f_l^d, \quad \forall\, l\in\mathcal{L} \\
& P_l^m \leq P_l^{max}, \forall\, l\in\mathcal{L}, m\in\mathcal{M}
\end{aligned} \tag{11}$$

The dual problem corresponding to the primal problem in (10) is then given by

$$\min_{\boldsymbol{\lambda}\succeq\mathbf{0},\boldsymbol{\mu}\succeq\mathbf{0}} D(\boldsymbol{\lambda},\boldsymbol{\mu}) \tag{12}$$

where $\succeq$ denotes component-wise inequality.

B. Dual Decomposition

Using (7), the dual function $D(\boldsymbol{\lambda},\boldsymbol{\mu})$ can be decomposed into the following three subproblems which are evaluated separately in the transport and network variables $\{r_n^d \text{ and } f_l^d\}$, the radio resource variables $\{P_l^m\}$, and the inverse-lifetime bound 's'.

$$\begin{aligned}
D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = & \max_{\{r_n^d,f_l^d\}\geq 0} \gamma \sum_{n\in\mathcal{N}}\sum_{d\in\mathcal{D}} U_n^d(r_n^d) - \sum_{l\in\mathcal{L}} \lambda_l f_l \\
& - \sum_{n\in\mathcal{N}} \mu_n \left\{ \sum_{l\in\mathcal{O}(n)} f_l \mathcal{E}_{tx} + \sum_{l\in\mathcal{I}(n)} f_l \mathcal{E}_{rx} + \sum_{d\in\mathcal{D}} r_n^d \mathcal{E}_s \right\} \\
\text{subject to } & \sum_{l\in\mathcal{O}(n)} f_l^d - \sum_{l\in\mathcal{I}(n)} f_l^d = r_n^d, \quad \forall\, d\in\mathcal{D}, n\in\mathcal{N} \\
& f_l = \sum_{d\in\mathcal{D}} f_l^d, \quad \forall\, l\in\mathcal{L}
\end{aligned} \tag{13}$$

$$\begin{aligned}
D_2(\boldsymbol{\lambda},\boldsymbol{\mu}) = & \max_{\{P_l^m\}\geq 0} \sum_{l\in\mathcal{L}} \lambda_l \left\{ \sum_{m\in\mathcal{M}} \nu_m C_l^m(P_l^m, W) \right\} \\
& - \sum_{n\in\mathcal{N}} \mu_n \left\{ \sum_{l\in\mathcal{O}(n)} \sum_{m\in\mathcal{M}} \nu_m P_l^m \right\} \\
\text{subject to } & P_l^m \leq P_l^{max}, \forall\, l\in\mathcal{L}, m\in\mathcal{M}
\end{aligned} \tag{14}$$

$$D_3(\boldsymbol{\lambda},\boldsymbol{\mu}) = \max_{s\geq 0} \sum_{n\in\mathcal{N}} \mu_n E_n s - (1-\gamma) F_n(s) \tag{15}$$

Note that in our system model, two nodes that can communicate with each other have two directed links between them, one for each flow direction. Hence, summing over all links $l\in\mathcal{L}$ is equivalent to summing over all outgoing links of all nodes. We thus have

$$\sum_{l\in\mathcal{L}} \lambda_l f_l = \sum_{n\in\mathcal{N}}\sum_{l\in\mathcal{O}(n)} \lambda_l f_l. \tag{16}$$

Using (16) and then eliminating the total link flow expression (2) from the constraints in (13), we rewrite this subproblem as

$$\begin{aligned}
D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = & \max_{\{r_n^d,f_l^d\}\geq 0} \sum_{n\in\mathcal{N}}\sum_{d\in\mathcal{D}} \Bigg\{ \gamma U_n^d(r_n^d) - \mu_n r_n^d \mathcal{E}_s \\
& - \sum_{l\in\mathcal{O}(n)} (\lambda_l + \mu_n \mathcal{E}_{tx}) f_l^d - \sum_{l\in\mathcal{I}(n)} \mu_n \mathcal{E}_{rx} f_l^d \Bigg\} \\
\text{subject to } & \sum_{l\in\mathcal{O}(n)} f_l^d - \sum_{l\in\mathcal{I}(n)} f_l^d = r_n^d, \quad \forall\, d\in\mathcal{D}, n\in\mathcal{N}
\end{aligned} \tag{17}$$

Similarly, subproblem (14) can be rewritten as

$$D_2(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \max_{\{P_l^m\} \geq 0} \sum_{n \in \mathcal{N}} \sum_{l \in \mathcal{O}(n)} \sum_{m \in \mathcal{M}} \nu_m \lambda_l C_l^m(P_l^m, W) - \nu_m \mu_n P_l^m \quad (18)$$

$$\text{subject to } P_l^m \leq P_l^{max}, \forall\, l \in \mathcal{O}(n), m \in \mathcal{M}, n \in \mathcal{N}$$

(17), (18), and (15) correspond to a *vertical* decomposition of the cross-layer optimization problem (10), into a joint transport and network layer problem, a radio resource layer problem, and a network lifetime problem respectively, which are coordinated by the master dual problem (12) using the dual variables $\boldsymbol{\lambda}, \boldsymbol{\mu}$.

C. Subgradient-based solution of the dual problem

Since the objective function in the primal problem (10) is not strictly convex in all the primal variables $\{s, r_n^d, f_l^d, P_l^m\}$, the dual function (11) is usually piecewise differentiable and the corresponding dual problem (12) is a non-differentiable convex optimization problem. Hence the usual gradient methods used for differentiable problems cannot be used to solve the dual problem. We use the subgradient method [11] that iteratively solves the dual problem.

For a given set of feasible dual variables, the three subproblems (15), (17), and (18) are solved in each iteration. Using these solutions, the subgradients of the dual function (11) are computed and are used to evaluate a new set of dual variables for the next iteration. Note that the strictly concave objective functions in subproblems (15) and (18) simplify their computation by allowing analytical solutions. We now present details of the subgradient based method and the conditions for convergence of the primal and dual variables to the optimal solutions.

Given a convex function $g\colon \mathbf{R}^n \rightarrow \mathbf{R}$, a vector $d \in \mathbf{R}^n$ is a subgradient of g at a point $u \in \mathbf{R}^n$ if $g(v) \geq g(u) + (v-u)^T d$, $\forall v \in \mathbf{R}^n$. Let, $\{r_n^{d*}(\boldsymbol{\lambda}, \boldsymbol{\mu}), f_l^{d*}(\boldsymbol{\lambda}, \boldsymbol{\mu})\}$, $\{P_l^{m*}(\boldsymbol{\lambda}, \boldsymbol{\mu})\}$, and $s^*(\boldsymbol{\lambda}, \boldsymbol{\mu})$ denote the optimal solutions of (17), (18), and (15) respectively. It can be verified that the subgradient of the dual function (11) with respect to the dual variable $\boldsymbol{\lambda}$ at point $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, denoted by $\boldsymbol{g} \in \mathbf{R}^{|\mathcal{L}|}$, is given by

$$g_l = \sum_{m \in \mathcal{M}} \nu_m C_l^m(P_l^{m*}, W) - f_l^*, \forall\, l \in \mathcal{O}(n), \forall\, n \in \mathcal{N}. \quad (19)$$

Similarly the subgradient with respect to $\boldsymbol{\mu}$ at point $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, denoted by $\boldsymbol{h} \in \mathbf{R}^{|\mathcal{N}|}$, is given by

$$h_n = E_n s^* - \left\{ \sum_{l \in \mathcal{O}(n)} \sum_{m \in \mathcal{M}} \nu_m P_l^{m*} + \sum_{l \in \mathcal{O}(n)} f_l^* \mathcal{E}_{tx} + \sum_{l \in \mathcal{I}(n)} f_l^* \mathcal{E}_{rx} + \sum_{d \in \mathcal{D}} r_n^{d*} \mathcal{E}_s \right\}, \forall n \in \mathcal{N}. \quad (20)$$

Using the subgradient corresponding to the dual variables at the k^{th} iteration, the dual variables are updated at the $(k+1)^{th}$ iteration as follows

$$\lambda_l(k+1) = [\lambda_l(k) - \beta_k g_l(k)]^+, \forall\, l \in \mathcal{O}(n), \forall\, n \in \mathcal{N}$$
$$\mu_n(k+1) = [\mu_n(k) - \alpha_k h_n(k)]^+, \forall\, n \in \mathcal{N}. \quad (21)$$

Here, $[.]^+$ denotes projection on the nonnegative orthant, and α_k, β_k are positive scalar stepsizes. Convergence to the optimal dual variables is guaranteed if the stepsizes are chosen such that [11]

$$\alpha_k \rightarrow 0, \sum_{k=1}^{\infty} \alpha_k = \infty \text{ and } \beta_k \rightarrow 0, \sum_{k=1}^{\infty} \beta_k = \infty.$$

The dual variables $\{\lambda_l, \mu_n\}$ and the subgradients $\{g_l, h_n\}$ have an intuitive economic interpretation. Let λ_l represent the price for utilizing the link capacity (cost per unit flow) and μ_n the price for utilizing the node's battery capacity (cost per Watt). The subgradient g_l (19) represents the excess capacity on link l, while the subgradient h_n (20) represents the excess battery capacity at node n. From the dual variable updates (21) at each iteration, it can be seen that the link and node-battery prices increase if the total link flow or total power dissipation exceeds the link capacity or the battery capacity respectively.

The joint transport and routing problem (17) maximizes the net utility function discounted by the cost of sensing at the nodes, the cost of congestion and transmission on all outgoing links, and the cost of reception on all incoming links. Thus higher link and node-battery prices result in greater penalty in the objective function in (17) forcing the source rates and flows to reduce.

The radio resource layer problem (18) maximizes the net revenue from the capacities that it supports discounted by the cost of battery reserves used in the process. While higher link prices allow higher revenue for the same increase in link powers, there is a corresponding cost to pay for utilizing the battery capacity proportional to the node-battery price.

The network lifetime problem (15) maximizes the revenue from supporting the battery capacities discounted by the lifetime-penalty function due to the resulting reduction in the lifetime. Thus, though higher node-battery prices allow higher revenue for the same increase in battery capacities (by increasing 's'), there is a corresponding penalty incurred due to the consequent lower lifetimes.

In its current form, the subgradient based approach described here does not allow a fully distributed solution of (10). However it sets a benchmark to study the convergence properties of distributed heuristics that might attempt to solve this cross-layer design problem. An important advantage of the vertical decomposition obtained above is that it allows several existing heuristics at transport, network, and radio resource layers to be separately evaluated in the context of joint utility and lifetime maximization.

V. Simulations

In order to illustrate the network utility-lifetime trade-off, we consider a sensor network consisting of four nodes located at the corners of a square of side 1000 m with the sink node at

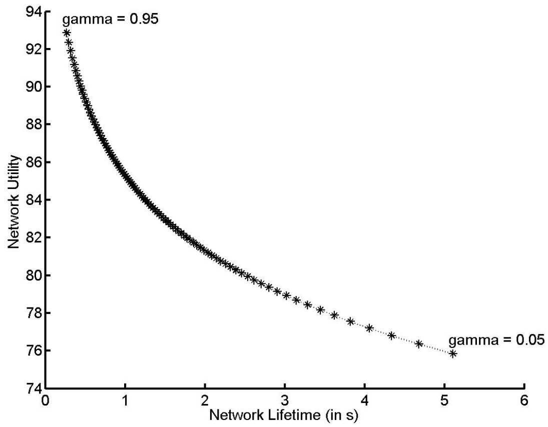

Fig. 2. Trade-off curve between network utility and network lifetime

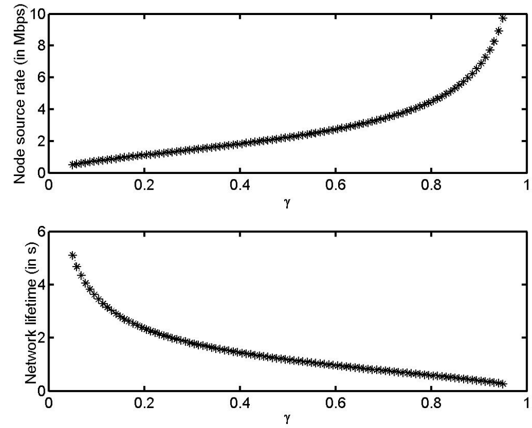

Fig. 3. Node source rate and lifetime as a function of γ

the center. We directly solve the primal optimization problem (10) with utility function $U_n^d(r_n^d) = \log_2(r_n^d)$, $\forall n \in \mathcal{N}$, $d \in \mathcal{D}$ and lifetime penalty function $F_n(s) = s^2$, $\forall n \in \mathcal{N}$. Since feasible modes correspond to sets of independent links in our system model, no feasible mode can have more than two active links in this five node network. All feasible modes except the idle mode are used for equal fractions of time.

The parameters for the node energy dissipation model $\mathcal{E}_s$, $\mathcal{E}_{rx}$, and $\mathcal{E}_{tx}$ are chosen to be 50 nJ/bit, 135 nJ/bit, and 45 nJ/bit [3] respectively. We assume a peak power P_l^{max} of 2 mW for all links, bandwidth W of 22 MHz, and path-loss exponent is set to 2. All four nodes have equal initial energies of 0.25 J each i.e. a total energy of 1 J in the network.

Fig. 2 shows the network utility ($\sum_{n=1}^{4} \log_2(r_n)$) and network lifetime trade-off curve as the system design parameter 'γ' is varied from 0.05 to 0.95. The figure clearly illustrates the inherent trade-off between the application performance and the lifetime costs in energy-limited wireless networks. As mentioned earlier, the case of $\gamma = 0$ corresponds to a network lifetime maximization problem resulting in a trivial solution with node source rates, link flows, and transmit powers all set to zero and infinite network lifetime. At the other end, $\gamma = 1$ corresponds to a network utility maximization problem resulting in near-zero network lifetime, with the source rates and link flows bounded only by the peak power constraint.

Due to the symmetry in the positions of the four nodes with respect to the sink, all nodes have the same optimal source rates and the same optimal lifetimes. Fig. 3 shows the source rate and lifetime of each node as a function of γ. Based on the desired application performance, the system designer can choose the optimal operation point for the system from Fig .2 and Fig. 3 by choosing the appropriate value of 'γ' and solving (10) for the optimal set of system variables.

VI. Conclusion

In wireless sensor networks, there is an inherent trade-off between application layer performance and the network lifetime. In this paper, we proposed a framework for cross-layer design across transport, network, and radio resource layers to find the optimal set of source rates, network flows, and radio resource allocation that jointly maximizes the network utility and lifetime. In the case of orthogonal link transmissions, we showed that the cross-layer optimization problem decomposes vertically into three separate problems - the joint transport and routing problem, the radio resource allocation problem, and the network lifetime problem, which interact through the link and node-battery prices. These three problems are coordinated by a master dual problem, which was solved using subgradient methods. As a part of future work, we are currently working on developing a fully distributed algorithm to solve the cross-layer design problem.

References

[1] S. Singh, M. Woo, and C. S. Raghavendra. Power-aware routing in mobile ad hoc networks. In *Proc. ACM/IEEE Int. Conf. on Mobile computing and networking*, pages 181–190, October 1998.

[2] V. Rodoplu and T. H. Meng. Minimum energy mobile wireless networks. In *Proc. IEEE Int. Conf. on Communications*, pages 1633–1639, June 1998.

[3] M. Bhardwaj and A. Chandrakasan. Bounding the Lifetime of Sensor Networks Via Optimal Role Assignments. In *IEEE Infocom'02*.

[4] W. Ye, J. Heidemann, and D. Estrin. Medium access control with coordinated, adaptive sleeping for wireless sensor networks. In *IEEE Transactions on Networking*, April 2004.

[5] J. H. Chang and L. Tassiulas. Routing for maximum system lifetime in wireless ad-hoc networks. In *37th Annual Allerton Conference on Communications, Control and Computing*, September 1999.

[6] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu. Impact of interference on multi-hop wireless network performance. In *Proc. ACM Mobicom*, 2003.

[7] L. Chen, S. Low, M. Chiang, and J. Doyle. Optimal cross-layer congestion control, routing, and scheduling design in ad hoc wireless networks. Preprint, 2005.

[8] L. Xiao, M. Johansson, and S. Boyd. Simultaneous Routing and Resource Allocation via Dual Decomposition. *IEEE Transactions on Communications.*, 52:1136–1144, July 2004.

[9] R. Madan, S. Cui, S. Lall, and A. Goldsmith. Cross-layer design for lifetime maximization in interference-limited wireless sensor networks. In *Proc. IEEE INFOCOM*, March 2005.

[10] S. Boyd and L. Vandenberghe. Convex Optimization. *Available: http://www.stanford.edu/people/boyd/cvxbook.html*, 2004.

[11] N. Z. Shor. Minimization methods for non-differentiable functions. In *Springer-Verlag*, 1985.