

Research Article

Energy-Efficient Route Optimization for Adaptive MPSK-Based Wireless Sensor Networks

Changmian Wang,¹ Liuguo Yin,² and Geir E. Øien¹

¹ Department of Electronics and Telecommunications, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

² School of Aerospace, TsingHua University, 100084 Beijing, China

Correspondence should be addressed to Changmian Wang, changmian.wang@gmail.com

Received 19 August 2009; Accepted 3 February 2010

Academic Editor: Athanasios Vasilakos

Copyright © 2010 Changmian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study a certain route configuration problem via optimization theory. We consider the optimal bit error rate (BER) and transmission rate allocations on each hop, subject to overall BER and delay constraints for a designated route. The pivot of the problem lies in the delay constraint, which divides the problem into two cases—the loose and the tight delay case. In the former, analytical solutions are obtained by applying the Karush-Kuhn-Tucker (KKT) theorem. Specifically, we discover in this case that for a given target BER, the optimum solutions are only related to the hop lengths in the route. When the delay constraint is tight, a mapping can be used to reduce the dimension of the problem by a factor of two; a numerical optimization algorithm has to be used to find the optimum. Simulation results show that by optimally configuring a chosen route, substantial energy savings could be obtained, especially under tight delay constraints. Simulation also reveals that a performance limit is reached as the number of hops increases. A parameter determining this limit is defined, and physical explanations are given accordingly.

1. Introduction

In the past few years, much research effort has been devoted to the field of wireless sensor networks (WSNs). Due to some of their unique features, such as low cost, small size, and limited transmission power, WSNs are in many ways different from typical mobile networks and hence pose many challenging problems. Some potential barriers to organization and coordination of such networks were identified by researchers (see, e.g., [1–3]). Usually, sensor nodes are battery driven and are supposed to function for a long-time period; possibly, for years. In many cases, it is economically undesirable or even impossible to recharge or replace the batteries. Hence one of the main concerns in designing such networks is how to operate the sensor nodes in an energy-efficient way.

In many applications of WSNs, the network is required to cover a large geographical area. In such cases, com-

munication between two nodes far apart will require the information to be relayed via several intermediate nodes. In addition, as the energy radiated in space decreases rapidly, roughly inversely proportional to at least the square of the transmission distance, from an energy-preserving perspective it can be worthwhile to forward the packets by means of multiple hops rather than a single hop. This motivates considering the problem of how to relay packets via multihops at minimum energy cost. In particular, we consider a scenario where such a multihop route is already formed from the source to the destination. This could be, for example, a predefined route chosen when the system was initially set up, an optimum route computed and established by a given routing algorithm in the network layer, or a route appointed by a well-powered data sink. It should be noted that contrary to the energy-aware routing algorithms for WSNs that have been heavily studied in the existing literatures, for example,

those in [4–6], in our work, we shall mainly set our focus on optimizing the physical layer resource allocations, such as the transmission power and the transmission rate.

In [7], a related route configuration problem for a multi-hop path was considered. Specifically, packet retransmissions using an automatic repeat request (ARQ) mechanism were studied, under a maximum transmission delay constraint, and a required packet delivery ratio constraint. In [8], minimization of the total energy consumption was considered, subject to a series of quality-of-service (QoS) constraints, by combination of certain adaptive transmission techniques. Results showed that large energy savings can be obtained by using optimized adaptive transmission techniques. However, the work in [8] was mainly based on simulation studies, and little was done regarding theoretical analysis. To provide more insights into the problem, in the present paper we study a similar problem, with more emphasis on theoretical analysis. Parts of the results in this paper were first published in [9].

We shall extend the ideas of [8] and show that by jointly considering per-hop bit error rate (BER) and transmission rate allocations under an end-to-end delay constraint and an end-to-end BER constraint, the route configuration problem can be formulated as an optimization problem. The delay constraint divides the whole problem into two cases—the loose and the tight delay case. In the former, an analytical solution can be obtained by using the Karush-Kuhn-Tucker (KKT) conditions [10]. However, under a tight delay constraint, the optimum cannot be acquired by using the KKT conditions alone. Therefore, a numerical optimization algorithm has to be employed to search for the minimum. In the latter case, it can however be shown that the number of unknown variables can be reduced by a factor of two through a unique mapping, which also enables the search algorithm to operate in an iterative manner. In addition, each iteration consists of two subconstrained convex optimization problems; hence the updating process converges rapidly. Simulations show that substantial energy savings can be attained with respect to a chosen benchmark scheme with constant per-hop BER and transmission rate. Results also reveal that a performance limit exists when the number of hops increases. A parameter quantifying this limit is defined, and explanations are given accordingly.

The remainder of this paper is organized as follows. In Section 2, we construct the general model for the problem. Section 3 solves the problem and evaluates the energy savings in the loose delay constraint case. Section 4 discusses the system performance under tight delay constraints, and a parameter measuring the maximum energy savings is defined. In Section 5 we discuss the effect of circuit power consumption and of adding peak power constraint. Conclusions are drawn in Section 6.

2. Model Construction and Problem Analysis

We consider ad hoc networks where there are multiple hop routes from the data sources to the sink. In particular, we

focus on networks with low traffic density properties. Many ad hoc networks fall within this category. For example, in some event-monitoring WSNs, hundreds of wireless sensor nodes will be randomly deployed to oversee a certain area. To save energy, all nodes shall keep silent until a target event is detected by one or a few nodes. Once the relatively rare event is detected in a certain part of the geographical area, data might be locally centralized within a cluster first, and subsequently be forwarded to the data sink.

In such a network, most traffic occurs only when an event is detected. Due to this low traffic density nature, there will mostly be almost negligible queuing delay, and nodes typically forward the packet immediately upon receiving it. In addition, links in such networks would also have few simultaneous transmissions. Therefore we simplify the problem by assuming negligible queuing delay and packet collision probability in all nodes throughout this work.

For WSNs, it is usually required that the sensor nodes be low cost, and small in size. Hence it is preferable to choose low complexity circuitry. M -ary Quadrature Amplitude Modulation (MQAM) schemes require the power amplifier (PA) to have accurate linearity due to multilevel amplitudes, and such amplifiers are usually expensive in practice. In M -ary Phase Shift Keying (MPSK), however, the constellation points are circled around the origin, which relaxes the PA linearity demands. For this reason, we shall assume that sensor nodes are using MPSK as their chosen modulation scheme.

In our work, adaptive radio is also assumed to be installed on each node. Results in [11] have shown that by implementing adaptive radio in a sensor node, substantial energy savings are possible in a single hop link. In [8], using adaptive MQAM modulation, the authors of the present paper have shown that this is also true for a multihop link.

A well approximated bit error rate bound for MPSK on an AWGN channel under various constellation sizes M was given in [9, Chapter 9], by $P_b = 0.2 \exp[-7\gamma/(2^{1.9k} + 1)]$, where $k = \log M$ is the number of bits per modulation symbol, γ is the signal-to-noise ratio (SNR), and P_b is the probability of bit error. Hence, to meet a certain target BER P_b , the required SNR at the receiver would be $\gamma = (1/7) \ln(1/5P_b)(2^{1.9k} + 1)$.

A deterministic κ th-power path-loss model was used in [11] to model the wireless link, resulting in an AWGN channel between any pair of nodes. To emphasize the energy consumption analysis rather than the packet error rate analysis, this model is also used in the current work. Assuming that the path-loss exponent κ is the same across the route, the channel gain can be modeled as $g(d) = 1/G_1 d^\kappa M_l$, where d is the physical distance between two nodes, G_1 is the gain factor at $d = 1$ m, κ is the path-loss exponent, and M_l is the link margin [11].

Hence, the SNR at the receiving node can be expressed as

$$\gamma = \frac{\bar{P}_t g(d)}{N_0 B}, \quad (1)$$

where \bar{P}_t [W] is the average transmit power, B is the system bandwidth, and N_0 is the power spectral density [W/Hz] of the additive white Gaussian noise (AWGN) at the receiver. It should again be noted that we have assumed negligible interference, hence we take only noise into account. If however, interference cannot be ignored, then N_0 can be set as the average power spectral density of the interference plus noise. In that case the conclusions of the present work still apply as long as the interference has approximately Gaussian characteristics.

In many WSNs, there are mechanisms designed for each node to discover its position. For example, the nodes could estimate their locations by measuring the signal strength to a few known points, or by exchanging information with neighboring nodes [12]. We assume that accurate locationing methods are applied in the network; thus the hop length information can be assumed to be known to the participating nodes before the transmission starts. Under the given channel model, all channel gains are then deterministically known to the transmitters.

In a WSN, we should also take the relevant QoS constraints into account. One of the most common, and perhaps the most important, constraints is the delay constraint, which implies that the packets should be delivered to the data sink or target node within a certain time limit. We denote this time limit as T [s]. It is also required that the packets should be delivered to the sink with a certain quality. That may, for example, imply that after passing through several relays, the packet should still satisfy an target end-to-end BER, denoted by P_{tar} .

We now assume that there are L hops from the source to the sink, and denote the BER at each hop, and the transmission rate in bits per symbol to be used in each hop, by $\mathbf{x}^T = [x_1, \dots, x_L]$, and $\mathbf{y}^T = [y_1, y_2, \dots, y_L]$, respectively. The packet length is assumed to be Q bits. Then the problem can be formulated as a minimization of total energy across all hops in the route:

$$\begin{aligned} \text{Minimize} \quad & \sum_i^L \bar{P}_{ti} \frac{Q}{By_i} = \left[(N_0 Q) \left(\frac{1}{7} \right) G_1 M_l \right] \\ & \times \sum_i^L d_i^k \ln \left(\frac{1}{5x_i} \right) (2^{1.9y_i} + 1) \frac{1}{y_i} \\ \text{subject to} \quad & \text{suitable QoS constraints.} \end{aligned} \quad (2)$$

It should be noted that the logarithm function $\ln(1/5x_i)$ requires $0 < x_i < 0.2$. Assuming that each node performs relaying in a decode-and-forward manner, the overall BER constraint requires that $[1 - \prod_i^L (1 - x_i)] = P_{\text{tar}}$, which can be rewritten as

$$\sum_i^L \ln(1 - x_i) - \ln(1 - P_{\text{tar}}) = 0. \quad (3)$$

It is of course also necessary that the BER in each hop satisfies $x_i > 0$ for $i = 1, 2, \dots, L$. Furthermore, in practice, P_{tar} is usually much lower than 0.2, and given any practical such value, constraint (3) will imply $x_i < P_{\text{tar}}$. Therefore we only require that $x_i > 0$, without demanding $x_i < 0.2$ specifically. The delay constraint can be written as

$$\sum_i^L \frac{Q}{By_i} \leq T, \quad \text{or equivalent as} \quad \frac{TB}{Q} - \sum_i^L \frac{1}{y_i} \geq 0. \quad (4)$$

We also assume from now on that uncoded BPSK is the lowest allowable modulation scheme in terms of rate that can be used in each node. This requires $y_i \geq 1$ for $i = 1, \dots, L$. It has been shown in [9, Chapter 9] that the performance of discrete-rate link adaptation schemes can often be accurately assessed by assuming that the rate is a continuous variable. Since this simplifies analysis, we also impose this assumption in this work by assuming y_i to be a real-valued continuous variable. If we define the positive constant $C = (N_0 Q)(1/7)G_1 M_l$, the problem can now finally be formulated as follows:

$$\text{Minimize} \quad f(\mathbf{x}, \mathbf{y}) = C \sum_i^L d_i^k \ln \left(\frac{1}{5x_i} \right) (2^{1.9y_i} + 1) \frac{1}{y_i} \quad (5)$$

$$\text{subject to} \quad x_i > 0 \quad \text{for } i = 1, 2, \dots, L, \quad (6)$$

$$\sum_{i=1}^L \ln(1 - x_i) - \ln(1 - P_{\text{tar}}) = 0, \quad (7)$$

$$y_i - 1 \geq 0 \quad \text{for } i = 1, 2, \dots, L, \quad (8)$$

$$\frac{TB}{Q} - \sum_{i=1}^L \frac{1}{y_i} \geq 0. \quad (9)$$

To find the minimum $f(\mathbf{x}^*, \mathbf{y}^*)$ of the objective function, and the point $(\mathbf{x}^*, \mathbf{y}^*)$ where it is attained, we apply the Karush-Kuhn-Tucker conditions (see, e.g., [10, Chapter 12]). They state that in this minimum, $\nabla f(\mathbf{x}^*, \mathbf{y}^*) = \sum_{i \in \mathcal{A}} \lambda_i^* \nabla c_i(\mathbf{x}^*, \mathbf{y}^*)$, where ∇ is the gradient operator, \mathcal{A} denotes the set of active constraints, $c_i(\mathbf{x}, \mathbf{y})$ are the constraints, and the parameters λ_i^* are the Lagrange multipliers. First, we assume that all the inequality constraints are active, and to simplify the expression, we define the function q' to be $q'(y_i) = (1.9 \ln 2) 2^{1.9y_i} (1/y_i) - (2^{1.9y_i} + 1)(1/y_i^2)$. It is the first-order derivative of function q , which is defined as $q(x) = (2^{1.9x} + 1)(1/x)$. In the appendix we show that $q'(y_i)$ is strictly positive for $y_i \geq 1$. Writing down the KKT equations, we obtain the following

equation:

$$\begin{aligned}
 & -C \begin{bmatrix} d_1^k \frac{1}{x_1} (2^{1.9y_1} + 1) \frac{1}{y_1} \\ \vdots \\ d_L^k \frac{1}{x_L} (2^{1.9y_L} + 1) \frac{1}{y_L} \\ d_1^k \ln(5x_1) q'(y_1) \\ \vdots \\ d_L^k \ln(5x_L) q'(y_L) \end{bmatrix} \\
 & = \lambda_d \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \frac{1}{y_1^2} \\ \vdots \\ \frac{1}{y_L^2} \end{bmatrix} + \lambda_e \begin{bmatrix} \frac{1}{x_1 - 1} \\ \vdots \\ \frac{1}{x_L - 1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
 & + \lambda_1 \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \dots + \lambda_L \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{10}
 \end{aligned}$$

In (10), λ_d is the Lagrange multiplier for the delay constraint (9), and λ_e is the Lagrange multiplier for the overall bit error rate constraint (7). $\lambda_1, \dots, \lambda_L$ are the Lagrange multipliers for the constraints in (8). Generally the optimal solution does not necessarily require all the inequality constraints to be simultaneously active, and in principle, we have to include all possible combinations of the inequality constraints as equalities and check which combination yields the smallest value of $f(\mathbf{x}^*, \mathbf{y}^*)$. However, we may simplify the problem in a different way, as shown in the next two sections, discussing, respectively, the cases of the loose and tight delay constraint.

3. Loose Delay Constraint

3.1. Closed Form Optimum Solutions. Let us first assume that at the KKT point, the delay constraint (9) is inactive, that is, $\lambda_d = 0$. Since x_i represents the BERs, it is easy to verify that $-C \cdot d_i^k \ln(5x_i) q'(y_i) > 0$, as long as $P_{\text{tar}} < 0.2$. Hence, the positivity of $-C \cdot d_i^k \ln(5x_i) q'(y_i)$, together with the nonnegativity of all λ_i for $i = 1, \dots, L$, implies that all the inequality constraints in (8) will become active, since otherwise no solution exists. In this case, the optimal solution for the per-hop bit rate in hop i , y_i is $y_i^* = 1$ for

$i = 1, \dots, L$. Therefore, the optimal per-hop BER in hop i , x_i , is the solution of the equation

$$\underbrace{-C \cdot (2^{1.9} + 1)}_{C'} d_i^k \frac{1}{x_i^*} = \lambda_e \frac{1}{x_i^* - 1}. \tag{11}$$

The optimal solution can be written as

$$\mathbf{x}^* = \left[\frac{C' d_1^k}{C' d_1^k - \lambda_e}, \dots, \frac{C' d_L^k}{C' d_L^k - \lambda_e} \right]^T, \quad \mathbf{y}^* = [1, \dots, 1]^T, \tag{12}$$

where λ_e can be obtained by substituting the optimum solution into constraint (7), which yield that λ_e should be the root of the polynomial equation

$$\prod_{i=1}^L (\lambda_e - C' d_i^k) - \frac{1}{1 - P_{\text{tar}}} = 0. \tag{13}$$

The fact that the optimum solution requires $y_i = 1$, for $i = 1, \dots, L$, means that using BPSK in every node is the best choice. This solution also makes sense in practice. Since lower-order modulation consumes less power compared to high-order modulations, it is preferable to use the lowest possible transmission rate as long as the resulting delay is acceptable. In our model, we have restricted ourselves to uncoded BPSK as the lowest possible rate.

We should re-emphasize that the above closed-form optimum solutions have been attained under the condition that the delay constraint is inactive, which happens *if and only if the time is sufficiently large*. The minimum under this condition is then *uniquely* given by (12). It might seem obvious, but we shall see that the contrapositive of this proposition can help us to identify one active constraint at all candidate KKT points in the *tight* delay constraint case.

The above optimal solutions have another layer of meaning. Note that the solution \mathbf{x}^* is optimal when $y_i^* = 1$ for all i . Hence, in this case, the problem has been converted to the question of how to optimally allocate the per-hop BERs in the route, when every node adopts the same modulation scheme. Solution \mathbf{x}^* in (12) is the answer to this. To get a better understanding about this optimum solution we shall, in the following, also develop an *approximate* solution, which will indicate clearly that for a given P_{tar} , the optimal per-hop BER allocation scheme is only related to the distance and is roughly κ th order inversely proportional to the ratio of its own distance to the overall distance of the entire route. Before getting into this, we will however first show the performance of the optimal configurations in the loose delay case through simulations.

3.2. Numerical Results and Approximation. Table 1 shows the numerical example parameters used in the simulations (we borrow some of these practical values from [11]). In many situations, the sensor nodes in a WSN are deployed in a random fashion. The hop length in a particular route will then behave like a random variable. In our simulations, we have assumed that the hop length obeys a *Rayleigh*

TABLE 1: System parameters.

$G_1 = 30$ dB	$M_l = 40$ dB
$\kappa = 3.5$	$B = 10$ kHz
$N_0 = -173.7$ dBm/Hz	$Q = 2$ kb

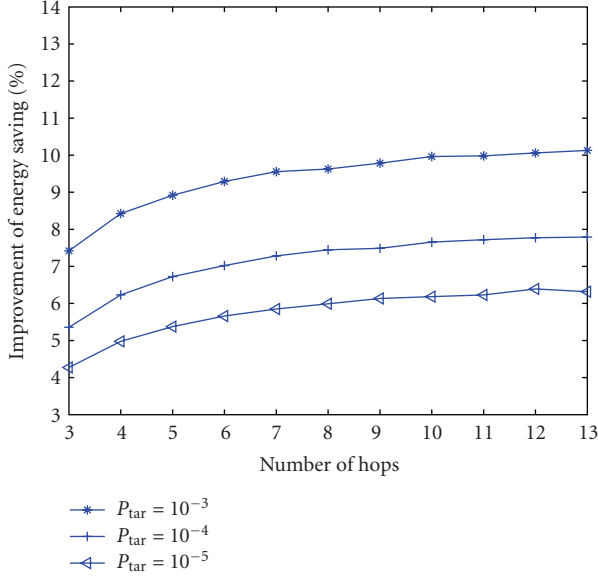


FIGURE 1: Average improved performance due to optimum BER allocations under different hop lengths for the loose delay case.

distribution, as this is shown in [8] to match very well the empirical hop length distribution when nodes are uniformly distributed across a finite square. The mean value of the hop length is set equal to 10 meters.

Figure 1 shows the simulation result regarding energy savings for different P_{tar} under different numbers of hops in the loose delay case. For a given P_{tar} the performance improvements are drawn as comparisons between optimally allocated BERs in all hops according to \mathbf{x}^* in (12), and the case when the per-hop BERs are evenly assigned as in [8], that is, by $P_{bi} = 1 - \sqrt[\kappa]{1 - P_{\text{tar}}}$. Because of the random nature of the route lengths, the gain for each route realization will be different; each dot drawn in Figure 1 is hence plotted after taking an average over 1000 randomly generated routes. Figure 1 shows that by optimally allocating the BER along the route, the total energy consumption across the route could indeed be reduced compared to uniformly assigning BER, roughly by around 7% on the average.

One prominent feature we can see from Figure 1 is that the average energy gains clearly drop when P_{tar} decreases. The reason for this could be seen from the objective function (5), where the BERs contribute to the energy budget by the terms $\ln(1/5x_i)$, where x_i represents the BER in each hop. The logarithm function is insensitive to changes in its argument when it is large. Therefore, when a lower target bit error rate is set up, the optimal BER assigned to each hop will accordingly become smaller. Hence, even if we optimally choose the BER, the performance gain will not be as obvious as in the larger P_{tar} case.

The second important feature from Figure 1 is that the performance gain grows larger as the number of hops increases. We may also explain this by means of the approximation $\ln(1-x) \approx -x$, which holds whenever $x \approx 0$. Since x represents the BER, which is typically a very small value, we could then approximate the constraint (7) well as a linear equality constraint:

$$\sum_{i=1}^L x_i = P_{\text{tar}}. \quad (14)$$

Now, we can use this constraint to compute λ_e instead of using (13). By recalculating the above steps and following exactly the same argument, this will give us the optimum solution $x_i^* = C' d_i^\kappa / \lambda_e$, where the Lagrange multiplier λ_e could be obtained by satisfying the equality constraint in (14). This leads us to the approximate optimum solution

$$\mathbf{x}^* = P_{\text{tar}} \left[\frac{d_1^\kappa}{\sum_{i=1}^L d_i^\kappa}, \dots, \frac{d_L^\kappa}{\sum_{i=1}^L d_i^\kappa} \right]^T, \quad \mathbf{y}^* = [1, \dots, 1]^T. \quad (15)$$

It can be seen from \mathbf{x}^* in (15) that for a given small P_{tar} , the BER allocation scheme is only related to the hop lengths in the route. We should also note that in (15), for one particular target route, the denominator $\sum_{i=1}^L d_i^\kappa$ will become a constant. Therefore, the optimum BER allocation for a specific hop i is proportional to the current hop length d_i powered by the path-loss exponent κ . This implies that we should allow larger error probabilities on those hops with long transmission distances, but that be very restrictive on the short hops. Thus, compared to the case where the per-hop BERs are evenly distributed, the energy savings are obtained by using relatively more power on the short hops, and relatively less on the long ones, as it costs much more transmit power to attain the same BER on a long hop than on a short hop. The result, compared to \mathbf{x}^* in (12), directly and profoundly reveals the relation between the optimum results and the hop lengths, and it also matches our intuition.

This analytical result can also be used to understand the second feature mentioned above: why the energy saving gains become greater as the number of hops increases. This can be seen through a simple example, where we assume a route consisting only of two hops. Denote their lengths by d_1 and d_2 , respectively, The total length of the route is set to be fixed, that is, $d_1 + d_2 = c$. It is known from above that the optimum BER allocations of this two-hop route is

$$P_{b1} = P_{\text{tar}} \frac{d_1^\kappa}{d_1^\kappa + d_2^\kappa}, \quad P_{b2} = P_{\text{tar}} \frac{d_2^\kappa}{d_1^\kappa + d_2^\kappa}. \quad (16)$$

We now explore the worst case where we have the maximum total energy consumption. Equivalently, this means that we need to solve the following constrained optimization problem:

$$\begin{aligned} \text{Maximize} \quad & C' \cdot \left[d_1^\kappa \ln \left(5P_{\text{tar}} \frac{d_1^\kappa}{d_1^\kappa + d_2^\kappa} \right) + d_2^\kappa \ln \left(5P_{\text{tar}} \frac{d_2^\kappa}{d_1^\kappa + d_2^\kappa} \right) \right] \\ \text{subject to} \quad & d_1 + d_2 = c. \end{aligned} \quad (17)$$

The function $x \log x$ is known to be strictly convex in x . Since the objective function is the sum of two such strictly convex functions with a sign reversal (C' is negative), it is a strictly concave function. It is not difficult to prove that this function reaches its maximum when $d_1 = d_2 = c/2$ which implies that the power savings are minimum when the hop lengths are equally distributed, and any unevenly distributed hop lengths will result in a potential for further energy gain. In other words, the more the hop lengths deviate from the constant length, the more gain we can obtain.

As mentioned earlier, in the above simulations, we have assumed the hop lengths to be i.i.d. Rayleigh random variables. As the number of hops increases, we are then more likely to experience large deviations in the hop lengths across the route, which after optimization will result in a higher performance gain. However, Figure 1 also shows that all performance gain curves saturate as the hop number grows greater than 10. The reason is that, as the number of hops increases, though we generally deviate more from a uniform distribution of the hop length, at the same time there is also a larger probability of having a noticeable number of hops which will have approximately the same length, which actually contribute very little to the overall performance improvement. The relative gain is thus becoming more stable. In the next section, we will be able to quantify this effect more exactly, through a single parameter quantifying the maximum possible gain.

4. Tight Delay Constraint

4.1. Two Convex Subproblems. As previously emphasized, if the delay constraint is inactive, then the time must be sufficiently large. The contrapositive of this implies that *if the time is not sufficiently large, then the delay constraint must be active.* This means that the optimal solution will stay the same if we rewrite the inequality constraint (9) as an equality, and it also means that we will have one more active constraint when writing the KKT equations. However, this will not help us further, since we still need to check the combinations of the rest of the inequality constraints. In this case, therefore numerical optimization algorithms must be used to search for the minimum. The change from inequality to equality does however facilitate the converging speed of the numerical algorithm (we used the MATLAB function *fmincon* to solve the optimization problem).

To aid the speed of the numerical optimization even more, further simplifications can also be explored, as will now be explained. If we assume that \mathbf{y} is given (i.e., we know the modulation order used in each hop), then the problem left will be the following optimization problem:

$$\text{Minimize } f(\mathbf{x} | \mathbf{y}) = C \sum_{i=1}^L (2^{1.9y_i} + 1) \frac{1}{y_i} d_i^\kappa \ln\left(\frac{1}{5x_i}\right) \quad (18)$$

$$\text{subject to } x_i > 0, \quad \text{for } i = 1, 2, \dots, L, \quad (19)$$

$$\sum_{i=1}^L \ln(1 - x_i) - \ln(1 - P_{\text{tar}}) = 0. \quad (20)$$

For this model, since \mathbf{y} is given, the objective function is a weighted sum of $\ln(1/5x_i)$ over all i , which can easily be shown to be strictly convex with respect to x_i for $x_i < 0.2$. Constraint (19) is a strict inequality constraint; however, it is not hard to see that the function $f(\mathbf{x} | \mathbf{y})$ will be infinite and thus cannot reach its minimum at the point x whenever $x_i = 0$ for $i = 1, 2, \dots, L$. Therefore it is no harm for us to modify the constraint (19) to a constraint where equality is included, that is, $x_i \geq 0$, for $i = 1, 2, \dots, L$. Furthermore, we will again use the approximate linear equality constraint $\sum_{i=1}^L x_i = P_{\text{tar}}$, instead of the exact nonlinear constraint (20). Hence, the optimization problem for a given \mathbf{y} can be rewritten as

$$\text{Minimize } f(\mathbf{x} | \mathbf{y}) = C \sum_{i=1}^L (2^{1.9y_i} + 1) \frac{1}{y_i} d_i^\kappa \ln\left(\frac{1}{5x_i}\right) \quad (21)$$

$$\text{subject to } x_i \geq 0, \quad \text{for } i = 1, 2, \dots, L, \quad (22)$$

$$\sum_{i=1}^L x_i = P_{\text{tar}}. \quad (23)$$

It should be noted that the replacement of the nonlinear constraint by the linear constraint will not introduce much of a gap between the exact and approximate optimal gains. We can see this from Figure 1 where the order of magnitude of variation on the target BER only results in a few percentage changes in the gain. This suggests that BER allocations in general have limited impact in the energy savings, and we will show later that compared with the gain achieved by introducing optimal transmission rate allocations, this gap can be ignored.

The constraints we have now consist of several sets of linear inequality constraints, and one linear equality constraint. Since the linear functions are also convex, the constraints (22) and (23) together define a convex feasible domain. Hence, by the convex KKT theorem (see, e.g., [13]), any KKT point will also be the global minimum. The previously noted fact that the minimum cannot exist at any point with $x_i = 0$, for $i = 1, 2, \dots, L$, implies that none of the constraints in (22) will be active at the minimum point, and thus they could be excluded when applying the KKT conditions. On the other hand, we know that the BER constraint (23) will surely be active; therefore, applying the KKT theorem gives us the following equation:

$$-C \begin{bmatrix} (2^{1.9y_1} + 1) \frac{1}{y_1} d_1^\kappa \frac{1}{x_1^*} \\ \vdots \\ (2^{1.9y_L} + 1) \frac{1}{y_L} d_L^\kappa \frac{1}{x_L^*} \end{bmatrix} = \lambda_e \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (24)$$

By solving this equation, we get the following optimum solution for any given \mathbf{y} :

$$x_i^* = -\frac{C d_i^\kappa (2^{1.9y_i} + 1) (1/y_i)}{\lambda_e}, \quad (25)$$

where λ_e satisfies the equality constraint $\sum_{i=1}^L x_i^* = P_{\text{tar}}$. Therefore, this gives us the following global minimum for problem (21):

$$x_i^* = \frac{P_{\text{tar}} d_i^\kappa (2^{1.9y_i} + 1) (1/y_i)}{\sum_{j=1}^L d_j^\kappa (2^{1.9y_j} + 1) (1/y_j)}. \quad (26)$$

One meaning of the above derived solution is that it tells how to find the optimum BERs along the route, given the transmission rate of all the nodes. Another is that it suggests that the existence of a set of mappings, denoted by $x_1^* = m_1(y)$, $x_2^* = m_2(y)$, \dots , $x_L^* = m_L(y)$, which together make sure that \mathbf{x}^* is the global minimum for the function $f(\mathbf{x} | \mathbf{y})$ defined above. We can thus eliminate x_i in the problem formulation, and the problem is now reduced to finding the \mathbf{y}^* which can minimize the original problem (5). The whole problem under a tight delay constraint can now be written as the following optimization problem:

$$\text{Minimize } f(\mathbf{y}) = C \sum_{i=1}^L d_i^\kappa \ln \left(\frac{1}{5m_i(y)} \right) (2^{1.9y_i} + 1) \frac{1}{y_i} \quad (27)$$

$$\text{subject to } y_i - 1 \geq 0, \quad \text{for } i = 1, 2, \dots, L, \quad (28)$$

$$\sum_{i=1}^L \frac{1}{y_i} \leq \frac{TB}{Q}. \quad (29)$$

Thus, we have reduced the number of unknowns by a factor of two, which will help to increase the numerical algorithm's search speed. The first constraint above is a set of linear inequality constraints, which naturally define a convex set. The second constraint is in fact a set of nonlinear inequality constraints. Hence, if we define $c(\mathbf{y}) = \sum_{i=1}^L (1/y_i)$, the Hessian (see, e.g., [10, Appendix A]) of $c(\mathbf{y})$ will be

$$\begin{bmatrix} \frac{1}{y_1^3} & 0 & \cdots & 0 \\ 0 & \frac{1}{y_2^3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{y_L^3} \end{bmatrix}. \quad (30)$$

Since by (28) $y_i \geq 1$ for all i , this matrix is diagonal with all entries strictly positive. Hence, it is a positive definite matrix, which in turn implies that $c(\mathbf{y})$ is a strictly convex function. The argument that the level set of a convex function is also convex (see, e.g., [14, Chapter 1]) makes (29) define a convex set. Moreover, since (28) is a set of linear inequality constraints, the argument of the intersection of convex sets being convex guarantees that the problem is defined on a convex feasible domain.

It should be noted that, as we have argued above, (29) can be written as an equality. However, to prove the convexity, it is admissible if we extend the domain to include the inequality cases, since we are still in the feasible domain. Furthermore, given an initial \mathbf{y} value, due to the mapping of

(26), the problem in (27)–(29) is equivalent to the following optimization problem:

$$\text{Minimize } f(\mathbf{y} | \mathbf{x}) = C \sum_{i=1}^L (2^{1.9y_i} + 1) \frac{1}{y_i} d_i^\kappa \ln \left(\frac{1}{5x_i} \right) \quad (31)$$

$$\text{subject to } y_i - 1 \geq 0, \quad \text{for } i = 1, 2, \dots, L, \quad (32)$$

$$\sum_{i=1}^L \frac{1}{y_i} \leq \frac{TB}{Q}. \quad (33)$$

This objective function can be proven to be convex with respect to \mathbf{y} (see the appendix). It suggests that we can find a unique global minimum \mathbf{y} that corresponds to the given \mathbf{x} , which is also uniquely determined by the previous \mathbf{y} through (26). Hence, the search can operate in an iterative manner. From an initial \mathbf{y}_1 we find an optimum \mathbf{x}_1^* through (26), and this \mathbf{x}^* in turn update \mathbf{y}_1 to a new value of \mathbf{y}_2 by solving problem (31). In particular, both subproblems are constrained convex optimization problems, which ensure that the extremum found is unique and global for each iteration. The above iterative process repeats until both \mathbf{x} and \mathbf{y} converge. In order to find the global minimum to the whole problem, different initial values of \mathbf{y} (which should comply with the constraints (32) and (33)) should be tested.

4.2. Numerical Results and Analysis. The system performance under optimum configuration is depicted in Figure 2, where average energy savings are obtained by jointly optimizing the transmission rate and the BER allocation. In particular, the target BER is set to $P_{\text{tar}} = 10^{-3}$ for all curves. The gain is plotted as a comparison between the optimal configuration and the case when both the BERs and transmission rates are evenly assigned along the route. To be consistent with simulation in the loose delay case, the hop length was still assumed to obey a Rayleigh distribution with mean value 10 meters. The curve is drawn over different values of the delay constraint. In particular, $T = 1$ implies the situation that if every node adopts BPSK, the delay constraint will be just met.

From the figure, all the curves indicate that we can have a significant average energy saving gain when the delay constraint is tight. For instance, in the 11 hops case, when the delay constraint is 0.1 (i.e., the acceptable delay is only 10% of what we would have when BPSK is used) the gain could be as much as 63%. Even in the 3 hops case, the gain could be close to 50%. This suggests that it is crucial to use adaptive transmissions if the delay requirement is tight and an energy efficient solution is required.

A typical sensor node usually consists of a sensing unit and a central processing unit, a transceiver unit, and energy spent on transmission is often dominant in the total energy cost. An improvement in transmission energy savings as seen in Figure 2 could therefore almost double a node's lifetime on the average. We also notice from Figure 2 that generally, the gain goes down as the delay constraint become less tight. This is again because, as the delay becomes larger, the optimum strategy is to better utilize the time and choose a lower modulation order, and the modulation in all nodes will

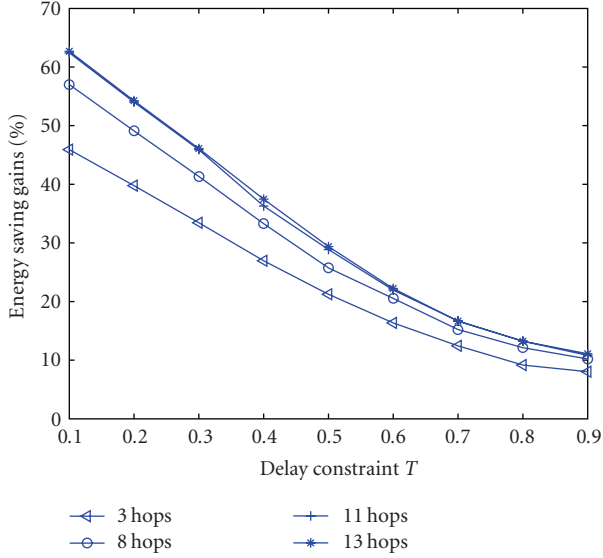


FIGURE 2: Energy savings for different number of hops under a tight delay constraint with target BER = 10^{-3} .

gradually converge to BPSK. Thus the system's performance will eventually approach the previously discussed loose delay case, in which only the per-hop BERs and corresponding transmit powers are optimally allocated.

Attention should also be paid to the gain performance as the hop number increases. It first improves fast over the whole range of delays; then, however, very little performance gain is obtained by going from 11 hops to 13 hops. This suggests that there is a performance limit; that is, the gain converges to a constant value as the number of hops grows very large. This phenomenon can be connected with the loose delay case where the BERs were optimally allocated. In Figure 1 we also observed that the improvement slows down when the number of hops grows very large. This implies that the system's maximum gain is limited by some inherent factor, to be explained next.

4.3. The D Parameter and Its Impact. It is important to understand the reason why substantial energy savings can be acquired when the number of hops goes from 3 to 8 in Figure 2. In the loose delay case, we demonstrated that the gain comes from the uneven distribution of the hop lengths. Similarly, if all hops had the same length, then the best strategy would be to let all hops use the same transmission rate and be assigned the same BER, which would produce no gain at all.

Also in the tight delay case, the gain actually comes from the differences in the hop lengths. These differences are ultimately controlled by the degree of deviation from the uniform hop length distribution. To characterize this effect, we hereby define this *degree of deviation*, D , by the ratio of the variance to the square of the mean:

$$D = \frac{\mathbb{E}[d^2] - \mathbb{E}^2[d]}{\mathbb{E}^2[d]} = \frac{\mathbb{E}[d^2]}{\mathbb{E}^2[d]} - 1. \quad (34)$$

Specifically, the D parameter for the Rayleigh distribution is in fact a constant, $(4 - \pi)/\pi \approx 0.273$. This explains why the simulation above indicates a performance limit, since as the hop number increases, the degree of deviation experienced in the hop length distribution of one route realization will statistically converge to this value. In the case shown in Figure 2, 11 hops are actually already large enough for this statistical convergence to occur, hence there is very limited gain by going from 11 hops to 13 hops. With a smaller number of hops, this situation of deviation effect is relatively less evident, which explains why the gain increment is more obvious for fewer hops.

The above analysis, on the other hand, also suggests that a higher performance gain could be obtained if the D parameter grows large. To verify this, we assume that the distance d obeys a Gamma distribution instead of Rayleigh and carry out a new simulation under this assumption. The probability density function of the Gamma distribution is defined by two parameters k and θ and is given by

$$f(d; k, \theta) = d^{k-1} \frac{e^{-d/\theta}}{\theta^k \Gamma(k)}, \quad \text{for } d > 0, k, \theta > 0, \quad (35)$$

where Γ denotes the gamma function. The mean value is $k\theta$, and the variance is $k\theta^2$. Therefore, the D parameter for Gamma distribution is $D = 1/k$. To be consistent with the simulation carried out previously, again we set the mean distance to be 10 meters, but with several different k and θ values such that the D parameter is varying. The P_{tar} is set to be 10^{-3} , and we choose a route consisting of 11 hops with delay set to be 0.3.

Figure 3 shows the results. We see a clear improvement when increasing the D parameter. It should be noted that the energy gain depicted in this figure is an average, since the hop lengths in a route are random variables. Figure 3 should thus be treated as an indicator of the average possible energy gain on the network level, since in an ad hoc network, communication can take place between arbitrary nodes, and hop lengths in the routes given by the routing algorithm can be random. It should also be noted that, in reality, if one particular route is given, it is more proper to configure the route according to the situation, since we might encounter the case where the network as a whole has a small overall D value, but a specific route inside might have a large hop length deviation, implying that a large energy saving gain can be achieved. Therefore, it is crucial to measure the D parameter of a specific route of interest.

Another thing we have not investigated so far is how the target BER P_{tar} affects the system performance in the tight delay constraint case. As we saw in the above analysis, for a Rayleigh hop length distribution, the system approaches a limit after approximately 11 hops. Therefore, we choose two representative numbers of hops, 3 and 11 hops, and run the simulations under different P_{tar} values.

Figure 4 shows the average system performance gain for different delay constraints and under different P_{tar} . In

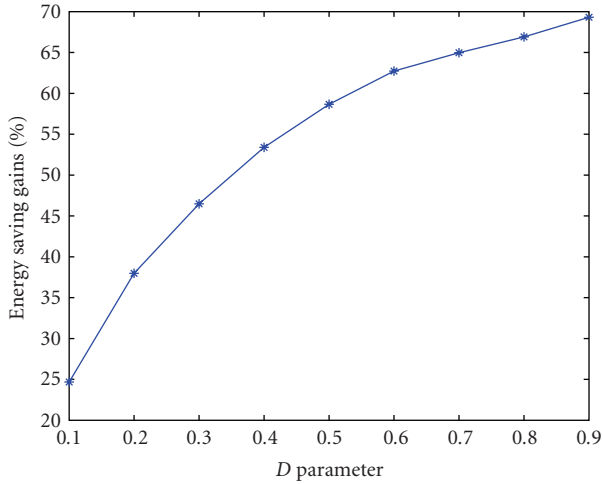


FIGURE 3: Performance improvement for different D parameters under a 11 hops route with delay constraint $T = 0.3$ and target BER = 10^{-3} .

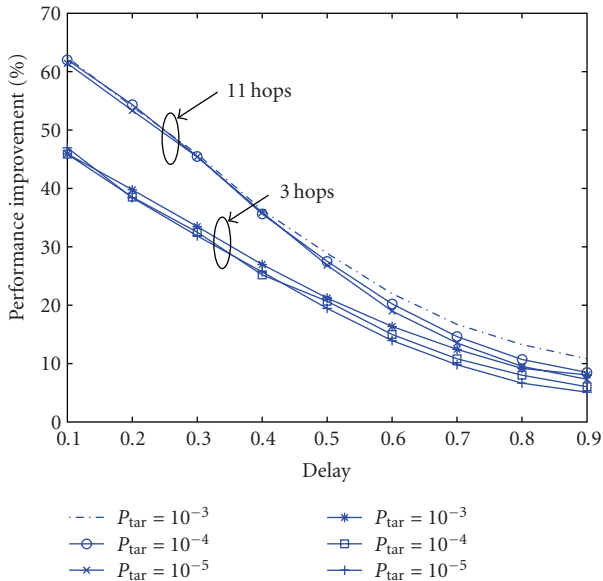


FIGURE 4: Performance improvement for different delay constraints under different P_{tar} .

general, the average performance is nearly the same as in Figure 2. Within each hop number, the curves are very close to each other, which means that there is no significant difference in relative energy saving gains on the average as we vary P_{tar} . We also notice that as the delay become larger, the discrepancies of the curves within the same cluster become more apparent. This is because, when the delay constraint is tight, the optimal BER assignment does not affect the average performance significantly, since at this point, the modulation order is very high. Thus the energy consumed by the modulation is dominant; whereas when the delay requirement is sufficiently loose, the modulation order will drop, and the energy spent on BER adaptation starts to come into play.

5. Circuit Power Consumption and Peak Power Constraints

When the communication distance is short, it has been shown that the circuit power consumption is approximately on the same order of magnitude as the transmission power and should be taken into account in the system optimization [11]. To keep this work fit for general ad hoc networks where transmission energy is most often dominant, circuit power consideration was however not particularly addressed so far. Also, the circuit power consumption, including the power of filters and power amplifiers, will be dependent on the practical parameters in use, such as bandwidth and PA efficiency. The power consumption in the circuitry can however be modeled as a given constant for one specific transceiver pair during the transmission. If we denote this constant by P_{cir} , the energy dissipated on the circuit along the whole route can be written as

$$\frac{P_{cir}Q}{B} \sum_{i=1}^L \frac{1}{y_i}, \quad \text{for } y_i \geq 1. \quad (36)$$

Equation (36) is a function of \mathbf{y} and it is of similar form as (29), which was proven to be convex. Hence, taking circuit power consumption into account preserves the convexity.

Similar to circuitry power consumption, peak power constraints have not been addressed either. One reason was that this would depend very much on specific implementation details such as the type of battery installed, the power shared by the other components other than the transceiver such as the sensing unit, and the central processing unit. Another reason was that the peak power constraint might be different from node to node in a particular route, since nodes may have different residual energy. However, we can show that adding a peak power constraint also preserves the convexity. If we denote the maximum available transmission power on node i as P_{ni} , then this peak power constraint on node i can be formulated as

$$\frac{1}{7}(N_0B) \ln\left(\frac{1}{x_i}\right) (2^{1.9y_i} + 1) \leq P_{ni}, \quad \text{for } y_i \geq 1. \quad (37)$$

It can be seen that the constraint above defines a convex set if x_i is given. Similarly, x_i is linked to y_i through (26). Therefore, for each iteration under an given initial value of \mathbf{y} , the problem in (27) remains convex. Moreover, it is foreseeable that the circuit power consumption and the peak power constraint only function as setting up an upper and lower bound for the feasible \mathbf{x}, \mathbf{y} , and neither of them will lead to a fundamental change of the results we get so far.

6. Conclusions

In this work, we have discussed a particular route configuration problem under a series of QoS constraints. It can be

seen that by using the KKT theorem, the delay constraint divides the problem into two cases: the loose delay and tight delay case. In the loose delay case, an analytical solution was obtained, and the problem was proved to be equivalent to the question of how to optimally assign BERs along the route. Under a pure path-loss channel model, we have also shown that the optimal BER allocation scheme for a given target bit error rate, perhaps somewhat surprisingly, is only related to the hop lengths d_i .

In the tight delay case a numerical optimization algorithm has to be employed to solve the problem. We could however reduce the dimensions of the original problem by a factor of 2 through a unique mapping, which enabled the search algorithm to operate in an iterative manner. Simulations showed that large energy savings are possible, especially when the delay constraint is tight. At the same time, we also noticed that the energy saving performance seems to converge to a certain limit as the number of hops increases. A parameter D measuring this limit was defined accordingly. Simulation results substantiate the validity of the parameter definition by indicating that more energy saving gain is possible as D increases.

It should be noted that, to keep the problem's generality, the problem has been analyzed on the basis of the most general physical layer QoS constraints, that is, the packet quality constraint (in terms of BER) and the delay constraint. Peak power constraints and circuit power consumption were not addressed, as these two parameters are highly dependent on the specific hardware implementations. However, we can show that these two constraints will not alter the convexity of the feasible domain, which implies that adding these two constraints would not fundamentally influence the analysis in the paper. Adding these two constraints can be thought of as imposing two additional bounds on the domain of \mathbf{x} and \mathbf{y} , again indicating that they will not change the results fundamentally.

It should also be noted that in this paper, the transmission rate at each hop is assumed to be a continuous variable. However, in practical implementations, a discretization operation is needed since only integer number of constellation size can be supported. How to optimally do the discretization and to analyze in detail the performance after discretization are beyond the scope of this paper and is a topic for further research. Work on discrete rates to achieve the maximum average spectral efficiency can be found in [15]. In addition, analysis based on other modulation schemes other than MPSK is definitely worthy of studying. This can be useful for example in well-powered ad hoc networks such as in multimedia WSNs, where spectral efficiency is more emphasized. In addition, due to the dense traffic properties in such cases, packet collisions and queuing delay should also be taken into account.

Finally, our proposed work can also be combined with routing protocol design. Specifically, it can be added as a measure for the route decision mechanism and help to identify the least energy-consuming route among several options after route configuration.

Appendix

The Function q and Its Properties

The function q is defined as

$$q(x) = (2^{1.9x} + 1) \frac{1}{x}. \quad (\text{A.1})$$

Now we prove that the function $q(x)$ is a positive, monotonically increasing convex function for $x \geq 1$. The first-order derivative of $q(x)$ is

$$q'(x) = \frac{[(1.9 \ln 2)x - 1]2^{1.9x} - 1}{x^2}. \quad (\text{A.2})$$

It is easy to verify that $q'(x)$ is greater than 0 for $x \geq 1$. Meanwhile, $q(1) \approx 2.732 > 0$. Hence $q(x)$ is strictly positive and monotonically increasing. The second-order derivative of $q(x)$ is

$$q''(x) = \frac{(1.9 \ln 2)^2 2^{1.9x} x^3 - [(1.9 \ln 2) \cdot x - 1]2^{1.9x} - 1}{x^4} 2x \quad (\text{A.3})$$

$$> \frac{[(1.9 \ln 2)^2 x^2 - 2(1.9 \ln 2)x + 1]x 2^{1.9x}}{x^4}. \quad (\text{A.4})$$

The term $[(1.9 \ln 2)^2 x^2 - 2(1.9 \ln 2)x + 1]$ in (A.4) is a second-order polynomial, which is increasing and strictly positive for $x \geq 1$. The linear function x and the exponential function $2^{1.9x}$ are both increasing and positive for $x \geq 1$. Hence $q''(x)$ is strictly positive when $x \geq 1$. We conclude that $q(x)$ is strictly convex for $x \geq 1$. Therefore, $q(x)$ is monotonically increasing and convex for $x \geq 1$.

Acknowledgments

This work was carried out and supported by the NORDITE/NFR(VERDIKT) project CROPS. The authors are grateful to Dr. Anna Kim for helpful discussions.

References

- [1] G. Karlsson, S. Lindfors, M. Skoglund, and G. Øien, "Cross-layer optimization in short-range wireless sensor networks (CROPS)," Technical Part, <http://www.ee.kth.se/commth/projects/CROPS/docs/>.
- [2] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, no. 4, pp. 8–27, 2002.
- [3] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 40–50, 2002.
- [4] M. Younis, M. Youssef, and K. Arisha, "Energy-aware routing in cluster-based sensor networks," in *Proceedings of the 10th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS '02)*, Fort Worth, Tex, USA, October 2002.
- [5] R. Shah and J. Rabaey, "Energy aware routing for low energy ad hoc sensor networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '02)*, Orlando, Fla, USA, March 2002.

- [6] S. D. Muruganathan, D. C. F. Ma, R. I. Bhasin, and A. O. Fapojuwo, "A centralized energy-efficient routing protocol for wireless sensor networks," *IEEE Communications Magazine*, vol. 43, no. 3, pp. S8–S13, 2005.
- [7] L. Yang, H.-C. Yang, and K. Wu, "Minimum-energy route configuration for wireless ad hoc networks," in *Proceedings of the 25th IEEE International Performance, Computing, and Communications Conference (IPCCC '06)*, pp. 9–14, Phoenix, Ariz, USA, April 2006.
- [8] E. Krogsvæn, C. Wang, G. E. Øien, and S. Lindfors, "Energy-efficient adaptive route configuration in short-range wireless ad-hoc networks," in *Proceedings of the 4th IEEE International Symposium on Wireless Communication Systems (ISWCS '07)*, pp. 292–296, Trondheim, Norway, October 2007.
- [9] C. M. Wang, L. G. Yin, and G. Øien, "Adaptive route configuration for increased energy efficiency in wireless sensor networks," in *Proceedings of the 5th International Conference on Broadband Communications, Networks, and Systems (CONET '08)*, pp. 16–23, London, UK, September 2008.
- [10] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, 2nd edition, 2006.
- [11] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.
- [12] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6–28, 2004.
- [13] H. E. Krogstad, "The Karush-Kuhn-Tucker Theorem," <http://www.math.ntnu.no/~hek/Optimering2007/KKT2007ver21feb.pdf>
- [14] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Cambridge, Mass, USA, 2003.
- [15] A. Gjendemsjø, G. E. Øien, and H. Holm, "Optimal power control for discrete-rate link adaptation schemes with capacity approaching coding," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, pp. 3498–3502, St. Louis, Mo, USA, November 2005.