Transactions Papers

Device Placement for Heterogeneous Wireless Sensor Networks: Minimum Cost with Lifetime Constraints

Quanhong Wang, *Member, IEEE,* Kenan Xu, *Student Member, IEEE,* Glen Takahara, *Member, IEEE,* and Hossam Hassanein, *Senior Member, IEEE*

***Abstract*— Device placement is a fundamental factor in determining the coverage, connectivity, cost and lifetime of a Wireless Sensor Network (WSN). In this paper, we explore the problem of relay node placement in heterogeneous WSNs. We formulate a generalized node placement optimization problem aimed at minimizing the network cost with constraints on lifetime and connectivity. Depending on the constraints, two representative scenarios of this problem are described. We characterize the first problem, where relay nodes are not energy constrained, as a minimum set covering problem. We further consider a more challenging scenario, where all nodes are energy limited. As an optimal solution to this problem is difficult to obtain, a two-phase approach is proposed, in which locally optimal design decisions are taken. The placement of the first phase relay nodes (FPRNs), which are directly connected to Sensor Nodes (SNs), is modeled as a minimum set covering problem. To ensure the relaying of the traffic from the FPRNs to the base station, three heuristic schemes are proposed to place the second phase relay nodes (SPRNs). Furthermore, a lower bound on the minimum number of SPRNs required for connectivity is provided. The efficiency of our proposals is investigated by numerical examples.**

***Index Terms*— Cost, connectivity, device placement, facility location problem, lifetime, minimum set covering, wireless sensor networks.**

I. Introduction

WE consider a formulation, in which (a) the number and positions of sensing spots are deterministic and assumed known in advance [1][2]; (b) three types of devices, Sensor Nodes (SNs), Relay Nodes (RNs), and a Base Station (BS), are used [3]; (c) devices can be installed at deliberately chosen spots [4]. Such WSNs can have a wide range of applications, such as building automation, environmental monitoring, traffic monitoring, etc.

A WSN can be logically decomposed into a sensing domain, which is dedicated to providing required coverage, and a communication domain, which is dedicated to ensuring the connectivity of the devices. Based on the WSN characteristics stated above, design of the sensing domain is application determined (e.g., [5]). In this paper, assuming the coverage-oriented placement has been done, we will concentrate on the connectivity-oriented placement in the communication domain.

The major contributions of this paper are as follows. Firstly, we introduce a device placement design framework for WSNs with the objective of minimum device cost under the constraints of coverage, connectivity and lifetime. We formulate a generalized optimization problem and exemplify two design scenarios with different constraints. Secondly, we propose optimal and heuristic solutions to two representative design scenarios. We model the problem in the case where SNs are energy constrained but RNs are not by the minimum set covering problem [6][7]. For a more challenging scenario, where both SNs and RNs are energy limited, we present locally optimal two-phase placement heuristics to approximate the overall optimal solution. The performance of different heuristic algorithms is evaluated and compared with a derived lower bound.

The remainder of this paper is organized as follows. In Section II, related research efforts are outlined. In Section III, the system models are described. The general device placement problem with the objective of minimizing device cost under the constraints of coverage, connectivity and lifetime, is formulated. In Section IV, the design problem with non-energy-constrained RNs is examined. In Section V, two-phase heuristic algorithms are proposed to solve the problem with

Manuscript received January 12, 2006; revised May 3, 2006 and October 10, 2006; accepted October 24, 2006. The associate editor coordinating the review of this paper and approving it for publication was R. Fantacci. This work is partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), and Communication and Information Technology Ontario (CITO), Canada.

Q. Wang and K. Xu are with the Department of Electrical and Computer Engineering, Queen's University, 99 University Avenue, Kingston, Ontario, K7L 3N6, Canada (e-mail: {wang, xu}@cs.queensu.ca).

G. Takahara is with the Department of Mathematics and Statistics, Queen's University, 99 University Avenue, Kingston, Ontario, K7L 3N6, Canada (e-mail: takahara@mast.queensu.ca).

H. Hassanein is with the School of Computing, Queen's University, 99 University Avenue, Kingston, Ontario, K7L 3N6, Canada (e-mail: hossam@cs.queensu.ca).

Digital Object Identifier 10.1109/TWC.2007.05357.

energy-constrained RNs. In Section VI, optimality of the First Phase Relay Node (FPRN) placement is discussed and a lower bound on the minimum number of Second Phase Relay Nodes (SPRNs) to guarantee connectivity is provided. In Section VII, the performance of the proposals is evaluated. Finally, the paper is concluded in Section VIII.

II. Related Work

Extending lifetime and conserving cost/energy are of paramount concern in WSN design. Conventionally, such issues have been addressed by devising various energy efficient networking protocols, whose objectives are to reduce energy wastage and/or balance the energy consumption ([8]-[9]). Promising performance is observed in analytical studies and simulations. On the other hand, device placement is another design space for WSNs, which has significant impact on energy efficiency and lifetime. Therefore, our interest is to study the placement problem in the context of an ideal set of energy efficient networking protocols, by which no energy is wasted. Recently, the optimal placement of heterogeneous devices has been considered ([1],[2],[10]-[11]). The research in [1] assumed the location and initial energy of SNs and application nodes (i.e., RNs) were known. In order to maximize the network lifetime, an optimal location was obtained for the BS through theoretical analysis. A relay traffic allocation scheme was developed to further extend the network lifetime. Upper and lower bounds on the maximum topological lifetime were also derived. In [10], based on a two-tiered network model, RN placement algorithms were proposed to guarantee network connectivity and/or ensure survivability in case of node failure. Lifetime constraints were not considered. In [2], aiming at maximizing the system lifetime under an energy budget, the joint design problem of energy provisioning and relay node placement was formulated as a mixed-integer nonlinear programming problem. To overcome the computational complexity, heuristic algorithms were introduced. The paper assumed that a RN could adapt its transmission energy to reach any other RN in the network, so the connectivity was not a problem. In contrast, the connectivity of nodes is an explicit constraint in our paper. In [12], the heterogeneous node placement problems were formulated with different objectives, i.e., minimizing the number of sensor nodes, minimizing total cost, minimizing energy consumption, maximizing network lifetime, and maximizing nodal utilization. The paper considered a discrete and finite set of feasible placement sites as the solution space. In contrast, our paper will search for the optimal solution in a continuous solution space. The research in [13] explored the tradeoff between cost and gains on lifetime and delivery rate when line-powered nodes are used in a WSN, though no specific optimization objective was considered.

While the efforts above attempt to extend the lifetime of a deterministic WSN, where nodes can be placed at deliberately chosen spots, some other efforts try to address similar problems for randomly deployed WSNs. In [14], with the objective of minimizing the cost, optimal nodal intensities and nodal initial energies are derived to guarantee a lifetime threshold in a randomly deployed WSN using two types of nodes. The research in [11] mapped the problem of placing transportation nodes (relay nodes) to an electrostatics problem. The optimal node deployment density was calculated in various scenarios.

Motivated by the benefits of device heterogeneity, as well as the wide range applications where device placement is deterministic, our research provides a feasible approach to provisioning WSNs of minimum cost under the constraint of guaranteed lifetime.

III. System Models

In many WSN applications, sensor nodes are expected to operate on duty cycle as low as 1% or less, such as home automation and industrial control [3]. Therefore, the traffic rate is low as compared to the bandwidth. This makes it easier for networking protocols to coordiate packet transmission so that collisions and network congestion are minimized. In other words, we assume the network to be energy limited as opposed to bandwidth limited. We model such a WSN as follows.

A. Network Model and Placement Problems

We consider a heterogeneous WSN composed of three types of devices, i.e., SNs, RNs and a BS. An SN generates data by sensing the environment and transmits the packets to a neighboring RN in one hop. Individual SNs may differ in a few aspects, for example, the physical properties they sense, the rate at which they generate data, and their transmission range, etc. However, all SNs have limited capability and unreplenished power supply. In order to conserve energy and reduce the complexity of circuits, SNs usually have fixed and limited transmission ranges and do not relay traffic for other nodes. One example of such an SN is the Reduced Function Device (RFD) defined in the IEEE802.15.4 standard[3]. Moreover, the traffic pattern of individual SNs is assumed predictable. An RN is assumed to have higher intelligence and carry more power supply than the SNs. It can function as a cluster header (CH) to collect data from SNs. In addition, it is capable of relaying and routing traffic, coordinating the media access in its proximity, and/or aggregating data. An RN can transmit information to the BS in one or multiple hops. A BS is located at a fixed position and it is the data sink for the system.

We assume the placement of SNs is application determined and is given. In this paper, we address the issue of the optimal placement of RNs. A general RN placement problem in a WSN is defined as follows: *Given a specific sensing task, determine the number and positions of heterogeneous devices, so that the total network cost is minimized while the constraints of lifetime and connectivity are satisfied.*

As design circumstances vary, the problem outlined above can be different from one case to another. Among all the factors, transmission range and energy supply of an RN are critical to distinguish the design scenarios. Depending on the transmission range of RNs, we envision three types of scenarios: Type I – RNs with adaptive transmission range, where the RNs can adjust their transmission power arbitrarily, so that the transmission from a source RN can reach any destination in the system; Type II – RNs with fixed transmission range, where the RNs always transmit at a fixed power, thus the transmission from an RN can only reach other RNs that fall in its fixed transmission range; and Type III – RNs with limited-adaptive transmission range, where the RN can adjust its transmission power within an upper bound. Note that the energy supply of an RN can be limited or unlimited. In Sections IV and V, we will describe several RN placement solutions.

B. Cost Model, Energy Model and System Lifetime

The cost of a device depends on its functionalities and power supply. The more functionalities a device has, the more complex and, thus, the more expensive it is. Also, there are various means of power supply (e.g., battery, solar panel, wall power) at different costs. In this paper, we assume that the costs of individual nodes of the same type are the same. We adopt the communication power consumption model used in [1][8]. The energy consumption for receiving a packet of length L is:

$$J_{rx} = L\beta \tag{1}$$

and the energy consumption for transmitting a packet of length L over the distance d is

$$J_{tx} = L\Big(\alpha_1 + \alpha_2 d^m\Big) \tag{2}$$

where α_1, α_2 and β are hardware specific parameters, and m is the path loss exponent.

This paper addresses the problem of minimizing cost with constraints on lifetime. In the literature, the lifetime of a node has been measured in various ways, such as the number of cycles over which the data is collected in a synchronized network [8], or the cumulative active time until the node depletes its energy [1]. In this paper, the nodal lifetime is represented by the cumulative traffic volume until its energy is depleted. In such a way, both synchronous and asynchronous traffic patterns are accommodated. We remark that our research can be easily modified for other lifetime measurements.

From the energy model above, it is clear that the communication energy consumption of a node depends on two factors, namely, the traffic volume[1] and the transmission distance. For an energy-constrained node, in order that it functions for a desired lifespan, its transmission radius and/or traffic volume should be confined.

1) Confined Transmission Range for SNs: Given the average traffic rate for an SN, which can be obtained through product specification or experimental measurement, the lifetime requirement can be converted into the traffic volume the SN should transmit. In order to do so, the transmission power of an SN has to be confined based on its power supply. Given a fixed initial energy J_{SN}, the confined transmission range corresponding to a certain lifetime requirement, L_{SN}, for an SN can be calculated as:

$$d_{SN} \leq \left(\frac{J_{SN}/L_{SN} - \alpha_{1,SN}}{\alpha_{2,SN}}\right)^{1/m}. \tag{3}$$

In reality, this confined transmission range can be enforced during manufacturing, as a part of product specification. Thus, each individual SN can be described by a transmission disc centered at the SN with the radius calculated by inequality (3). Therefore, in order that an SN can function for the required time and transmit its data to the BS via the RNs, an RN should be placed inside the transmission disc of the SN to relay its traffic.

2) Constraints for RNs: The major tasks of an RN are receiving traffic from other nodes and transmitting data to the BS or the next hop RN. For an RN, the incoming traffic is equal to the outgoing traffic (assuming no traffic aggregation is done).[2] Given a fixed energy supply J_{RN}, when its transmission range is fixed and known, the cumulative traffic amount an RN can handle can be bounded as:

$$L_{RN} \leq \frac{J_{RN}}{\alpha_{1,RN} + \alpha_{2,RN} d_{RN}^m + \beta_{RN}}. \tag{4}$$

Therefore, the initial energy of an RN determines the total cumulative traffic amount it can handle, which is referred to as the RN capacity, denoted by C in the remainder of this paper. If the initial energy is unlimited (e.g., wall powered) or sufficiently large relative to that of SNs, then C may be taken to be infinite. In the following, we assume that if the amount of traffic handled by an RN does not exceed C, its lifetime constraint is satisfied.

IV. Placement of Relay Nodes Without Energy Constraints

We first consider a basic version of the generalized problem. SNs are energy constrained and transmit data over a small distance. Individual SNs may have different transmission ranges depending on their initial energy and surrounding environment. RNs have large transmission distance and ample energy supply (wall powered, solar powered, or high capacity battery) so that their energy constraint is not a critical issue in contrast to the SNs. RNs can send data to the BS in one hop or multiple hops. In this scenario, as the connectivity and lifetime constraints on RNs are not factors, the optimization problem becomes: given a deployment of SNs, find a minimum number of RNs and their positions so that each SN can reach at least one RN in a single hop. We note that no assumptions regarding routing are required for this problem, as SN lifetime and connectivity constraints are satisfied if at least one RN is placed within the transmission radius of each SN. We formulate the problem of optimal RN placement (for the sake of connectivity) as a set covering problem.

Let $X = \{o_1, o_2, \ldots, o_N\}$ be the set of given SNs. If an RN is placed at a strategic position where multiple SN discs overlap, the corresponding SNs will all benefit from this RN. We make the following definitions.

Definition 1: *Region.* For a given set X of N SNs, let a subset s of X denote the intersection of the corresponding discs. Then s is called a *region* if this intersection is nonempty. For example, the subset $s = \{o_1, o_2\}$ denotes the nonempty intersection of the discs of o_1 and o_2. An RN placed in the intersection corresponding to s_i can serve at least as many SNs as an RN placed in the intersection corresponding to s_j if $s_j \subset s_i$, and vice versa. Note that each SN disc is also a region by itself.

[1]Only the uplink data traffic is considered for the energy consumption calculation, as it counts for the majority of traffic.

[2]With moderate modifications, the energy model can be extended to the case with data aggregation.

Definition 2: *Densest Region.* A region s_i is said to be a densest region if there is no region s_k, satisfying $s_i \subset s_k$.

Let f denote an individual densest region and F denote the set of all densest regions of the disc set X. For any region s_k of the set X, there must be a densest region $f_l \in F$, such that $s_k \subseteq f_l$. If an RN is placed in the intersection corresponding to s_k, it can serve the same number or more SNs if it is also placed in the intersection corresponding to f_l. Therefore, when designing the optimal placement of RNs, the non-densest regions can be ignored. Hereafter, we restrict our consideration to the intersections corresponding to densest regions as the candidate positions for RNs. As such, the problem of finding the optimal placement of RNs for a set of SNs becomes equivalent to finding a *minimum set cover* for the disc set X using its corresponding set of densest regions F [6][7]. In practice, any point in the intersection corresponding to a densest region is treated equally in the sense that an RN placed anywhere in the intersection incurs the same cost while satisfying the lifetime and connectivity constraints.

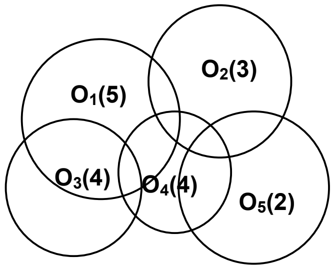
We state the minimum set covering problem as follows. We are given a finite set X and a family F of subsets of X such that for any $x \in X$ there is at least one $f \in F$ so that $x \in f$. A set cover is a subset $C \subseteq F$ whose members cover all of X, i.e., $X = \bigcup_{f \in C} f$. A set cover which has the smallest size is called a *minimum set cover* from F for X.

The above RN placement problem can be viewed as a special form of the un-capacitated facility location problem in the area of operations research [15][16]. As such, many techniques developed for the un-capacitated facility location problem in the operations research literature can be considered here [17][18]. The minimum set covering problem is NP-complete (so is the above RN placement problem). In [6], a polynomial-time approximation algorithm, called the GREEDY-SET-COVER algorithm, is presented. In [19], we propose an optimal recursive algorithm that requires lower computation than exhaustive enumeration. The proposed algorithm divides the overall minimum set covering problem into a series of minimum set covering problems of a smaller size iteratively and combines the results in an optimized manner.

V. Placement of Relay Nodes With Energy Constraints

In this section, a more general and challenging situation is considered. RNs are assumed to have limited energy and fixed transmission range. Therefore, RNs connecting to SNs can only relay a limited amount of traffic within a restricted range. In general, RNs may not be able to transmit data to the BS by themselves. Thus, a complete RN placement solution should not only provide connectivity to the SNs, but also ensure that each RN has at least one (multi-hop) path to the BS. In addition, the lifetime constraints of RNs, or the capacity constraints as described in Section III, should be satisfied. As RNs are assumed to have identical cost, the optimization objective is to minimize the number of RNs for a given deployment of SNs.

For an arbitrary SN placement, the solution space for the RN placement problem in this scenario is infinitely large and finding the optimal one is highly non-trivial. As such, we propose a two-phase RN placement approach. In each phase the number and locations of RNs to be added are decided in a locally optimal manner.[3] In the first phase, a minimum number of RNs are placed to ensure the connectivity of SNs. We therefore call them First Phase RNs (FPRNs). In the second phase, RNs are placed to provide a complete relay path for the existing RNs. They are called Second Phase RNs (SPRNs). In both phases, lifetime (i.e., capacity) and connectivity requirements have to be satisfied.



A set of SNs= $\{o_1, o_2, o_3, o_4, o_5\}$ and their traffic, denoted by the numbers in brackets. Assume a RN can relay a maximum traffic volume of 10 in order to meet the lifetime constraint. Subsets, such as $\{o_1, o_3, o_4\}$, $\{o_1, o_2, o_4\}$ are regions. Each region, such as $\{o_1, o_3, o_4\}$, has an associated intersection of the transmission disks. $\{o_1, o_3\}$, $\{o_1, o_4\}$, $\{o_3, o_4\}$ are also regions. However, $\{o_1, o_3, o_4\}$ is not energy feasible as the sum of the traffic of its associated nodes is greater than 10. $\{o_1, o_3\}$, $\{o_1, o_4\}$, $\{o_3, o_4\}$ are EFRs, as in each of them the total traffic of their component nodes does not exceed 10. However, the EFR $\{o_2, o_5\}$ is not DEFR since it belongs to the EFR $\{o_2, o_4, o_5\}$.

Fig. 1. An illustration of regions, EFRs, and DEFRs.

A. The Placement of First Phase Relay Nodes (FPRNs)

The objective of FPRN placement is to ensure the connectivity of SNs as they have limited transmission range. It is a similar problem as that described in the previous section. However, due to the power limitation on RNs, the amount of traffic that an RN can handle is limited (refer to Eq.(4)). As such, some densest regions may not be energy feasible in the sense that the total traffic volume of SNs associated with a region is greater than the capacity of an RN. For example, in Fig. 1, the total traffic volume of the SNs associated with region $f = \{o_1, o_3, o_4\}$ is 13 while the capacity of each RN is only 10. We enhance the criteria of candidate FPRN placement locations with an energy constraint. After finding the eligible candidate locations by applying the enhanced criteria, the minimum set covering model is applicable to the FPRN placement problem. The new candidate locations are called *densest energy-feasible regions*.

Definition 3: *Energy-Feasible Region (EFR) and Densest Energy-Feasible Region (DEFR).* For a set of SNs, $X =$

[3]The two-phase approach first appears in our earlier work in [20].

$\{o_1, o_2, \ldots, o_N\}$, a region is energy feasible if an RN deployed in the intersection corresponding to this region can relay all traffic from the associated SNs while meeting the lifetime constraint.[4] Such a region is called an Energy-Feasible Region (EFR). An EFR $R^\star$ is a DEFR if there is no other EFR R such that $R^\star \subset R$.

The concepts of region, EFR, and DEFR are illustrated in Fig. 1. It is safe to assume that the capacity of an RN is not less than the cumulative traffic volume of any individual SN. Therefore, each SN should be associated with at least one RN in a DEFR. By finding a minimum set covering for X among DEFRs using such algorithms as in [19], the FPRN placement solution can provide the connectivity to all SNs. The placement problem of FPRNs can also be viewed as a special form of the un-capacitated facility location problem.

B. The Placement of Second Phase Relay Nodes (SPRNs)

By the end of the first phase placement, every SN has found an RN to forward its traffic. Next, we need to place more RNs so that every FPRN will be able to find the neighbor(s) via which it relays traffic to the BS. To the best of our knowledge, there is no existing model for such a problem. We should note that with energy and transmission constrained RNs multiple hops are generally required to relay data from SNs to the BS. Therefore, any placement solution depends on the underlying routing protocol taking sufficient advantage of the RN placements to realize the desired network lifetime. In this paper we assume that the routing protocol used is sufficiently energy efficient so that the target network lifetime can be realized. Alternatively, in the construction of each of our placement solutions the level of traffic on each network link is specified, and a centralized routing scheme [21] that enforced these levels of traffic would ensure the target network lifetime. We first identify two essential design principles, then present three heuristic algorithms to implement the principles.[5]

1) Far-Near and Max-Min Principles:

Far-Near Principle. This refers to the principle that the placement decision in the second phase should first consider the RN which is farthest from the BS and evolves step-by-step to the RNs that are closest to the BS. The rationale is that data are to be forwarded towards the BS. Hence RNs that are closer to the BS should relay traffic for other farther nodes. This principle helps to avoid energy wastage incurred due to unnecessary detouring of relayed traffic.

Max-Min Principle. This refers to the principle to maximally utilize the capacity of existing RNs, while introducing a minimum number of new RNs. Specifically, from far to near to the BS, each RN will distribute its workload to other existing neighboring RNs first. Only when the existing neighboring RNs of a given RN cannot handle its traffic load, a new RN will be added. In order to implement this principle, we assume that a supportive energy-aware routing protocol or an optimal traffic allocation mechanism is available. This is outside the scope of this paper. However, such issues have been addressed in [21][9].

2) Localized Heuristic Algorithms: Following the principles above, three heuristic algorithms are proposed, which differ in the way the traffic load of an RN is forwarded and distributed to neighboring RNs.

We can abstract the SPRN placement design as a procedure of constructing a directed graph $G(V, A)$, where V is the set of RNs and A is the edge set. Initially, $V = \{v_1, v_2, \ldots, v_N\}$ is the set of FPRNs. Each RN has an associated weight which indicates the traffic load it handles. Let d_i denote the Euclidean distance from v_i to the BS, d_{ij} the Euclidean distance from v_i to v_j and r the fixed transmission radius of an RN. There would be a directed edge from v_i to v_j if $d_{ij} \leq r$ and $d_j < d_i$, i.e., if the transmission by v_i can reach v_j and v_j is closer to the BS than v_i. Furthermore, we define the workload, w_i, as the sum of v_i's relayed traffic loads, and its residual capacity as the difference between its capacity, C, and its workload.

A. Nearest-To-BS-First algorithm (NTBF): Starting from the farthest RN, say v_{far}, if the workload of v_{far}, say w_{far}, does not exceed the total residual capacity of its adjacent neighbors, then its workload is distributed to its adjacent neighbors, by first filling up the capacity of the node nearest to the BS, then to the node next nearest to the BS, and so on. In case there are two or more neighboring nodes having the same closest distance to BS, one is chosen arbitrarily. Otherwise, a new RN will be introduced as its next hop relay. The algorithm is described in Fig. 2.

Step 0: (Initialization) Set $U = V$.
Step 1: Pick a node v_i that is farthest from the BS among the nodes in the set U. If $d_i \leq r$, delete v_i from U, go to Step 3. Otherwise, go to Step 2.
Step 2: (Traffic Distribution with Nearest-To-BS-First) Calculate the remaining capacity (RC) of the neighboring nodes of node v_i as follows:
$RC = \sum(C - w_k)$, where the sum is over all k such that $d_{ik} \leq r$ and $d_k < d_i$.
If $RC \geq w_i$ then
v_i distributes its traffic load to the neighboring nodes by first filling up the capacity of the node nearest to the BS, then of the node next nearest to the BS, and so on. In case there are two or more neighboring nodes having the same closest distance to the BS, one is chosen arbitrarily. Delete v_i from U, go to Step 3.
Else
A new RN, v_{new}, is deployed on the line from v_i to the BS at the point that is distance r away from v_i. Set $w_{new} = w_i$, add v_{new} into U, delete v_i from U, and go to Step 3.
Step 3: If $U = \phi$, EXIT. Otherwise, go to Step 1.

Fig. 2. Nearest-To-BS-First algorithm.

B. Max-Residual-Capacity-First algorithm (MRCF): We observe that by using the NTBF algorithm, the workloads among the nodes could become unbalanced, since the traffic distribution is sensitive to the distance between a node and the BS. A tiny difference in distance to the BS could lead to significant variance in the workloads of two nodes. Thus, some potential traffic path segments may become jammed causing

[4] In this research, we assume an SN only transmits its traffic to one RN during its lifetime. In the case that an SN transmits its traffic to multiple RNs, one needs to find a fractional solution to this problem [16]-[18].

[5] The preliminary results of the proposed schemes were shown in our earlier work in [22].

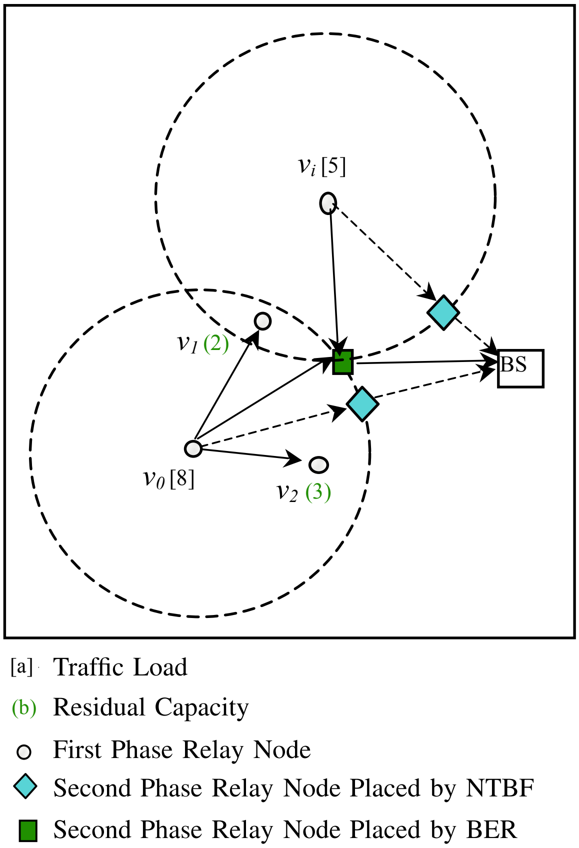


Fig. 3. An illustration of the Best-Effort-Relaying algorithm.

more new RNs to be added to set up new paths. As such, we introduce the MRCF algorithm to maintain better load balance among the RNs. The performance improvement by doing so will be considered in Section VII.

The MRCF algorithm is similar to the NTBF algorithm, except Step 2. When distributing the workload of a given RN to its neighbors, it first fills up the capacity of the neighboring RN with maximum residual capacity, then the neighboring RN with second to maximum residual capacity, and so on. If two or more neighboring nodes have the same highest residual capacity, then the one nearer to the BS is chosen first.

C. Best-Effort-Relaying algorithm (BER): In the previous algorithms, a new RN is added if the neighbors of the farthest RN cannot relay its workload. However, the capacity of its neighbors is potentially wasted as its workload is totally passed to the new RN without bothering the existing neighboring RNs. Therefore, we amend the previous algorithms to utilize the existing RNs to relay traffic in a best-effort manner. That is, the traffic relaying will be arranged even if an RN's neighbors cannot serve it all. In addition, when placing a new RN, the location is picked not only to make the RN be as close to the BS as possible, but also to be strategically placed so as to serve as many existing RNs as possible. For example, in Fig. 3, v_1 and v_2 are assigned a portion of the traffic load from v_0, even though they cannot handle all the traffic load of v_0. Meanwhile, one fewer SPRN is needed if using the BER algorithm than using the NTBF algorithm.

To describe the algorithm, we define the following concepts and notations. The residual load of a node refers to the

Step 0: (Initialization) Set $U = V$.
Step 1: Remove from U all RNs that have zero residual capacity and can reach the BS in one hop.
Step 2: From far to near, let each RN in U relay as much traffic load as possible to its neighboring RNs which are closer to the BS, using the Nearest-To-BS-First rule; Update the residual capacity and workload for each RN; Remove from U all RNs that are both fully loaded and have zero residual capacity.
Step 3: Up to this step, we have a set of RNs that cannot distribute their workload using the existing RNs. Therefore, we now add a new RN. Pick the RN in U farthest from the BS and let v_i denote this RN;
If v_i has zero residual capacity, remove v_i from U and go to Step 5;
Else if the residual load of v_i equals its capacity, go to Step 4;
Else
  Let $N = \mathit{INS}(v_i)$;
  If $(N = \phi)$ go to Step 4;
  Else
    Set $W = \{v_i\}$;
    Do the following:
      Pick the RN $v_j \in N$ that is farthest from the BS;
      If $\mathit{GIR}(\{v_j\}) \bigcap \mathit{GIR}(W) \neq \phi$ and the total residual load of RNs in $\{v_j\} \bigcup W$ is no more than the capacity of a new node, then update $W \leftarrow W \bigcup \{v_j\}$ and $\mathit{GIR}(W)$, and remove v_j from N;
    Until the total load in W is equal to the capacity of a new node or all $v_j \in \mathit{INS}(v_i)$ have been checked;
    Place the new node in $\mathit{GIR}(W)$ as close to the BS as possible and relay the load from RNs in the set W to the new RN;
    Update node residual capacities and workloads of RNs in U;
    Remove all nodes in W from U and add the new node to U;
    Go to Step 5.
Step 4: Place a new RN on the line joining v_i to the BS as close to the BS as possible, and assign the load from v_i to the new RN; remove v_i from U; add the new node to U.
Step 5: Repeat Steps 1, 2 and 3 until all nodes in U can reach the BS in one hop.

Fig. 4. Best-Effort-Relaying algorithm.

difference between its total traffic load and the traffic load for which next step RNs have been assigned. Representing an RN v_i by a disc with radius r centered at the location of v_i, let $\mathit{INS}(v_i)$ denote the intersecting neighbor set of v_i, i.e., $\mathit{INS}(v_i) = \{v_j \in V : d_{ij} \leq 2r, j \neq i\}$. Let $\mathit{GIR}(W)$ denote the geometrical intersecting region (i.e., the intersection) of the transmission discs of all RNs in the RN set W. The BER algorithm is described in Fig. 4.

VI. Theoretical Analysis

In Section V, we examined the placement problem for a heterogeneous WSN when both SNs and RNs are energy constrained. A two-phase placement approach is proposed. The first phase placement is modeled as a minimum set covering problem with DEFRs. One can easily verify that the FPRN placement is optimal in the sense that it uses the minimum number of RNs required to satisfy the condition that each SN can connect directly to at least one RN for the duration of its lifetime.

For the SPRN placement, an optimal solution is unknown to our best knowledge in a general scenario. We now derive a lower bound on the number of SPRNs required to ensure that each FPRN has at least one path to the BS. The notations in Section V-B.2 are used. Also we define $d_{\max} = \max_{v_j \in V} d_j$. Given a BS and a RN set V, draw M concentric circles of radii $r, 2r, \ldots, Mr$ centered at the BS, where $M = \lceil d_{\max}/r \rceil$. Thus, all FPRNs are within the circle of radius Mr. Define *shell* m, denoted by S_m, to be the region between the mth and $(m+1)$th circles, $m = 1, 2, \ldots, M-1$, and shell 0 to be the disc of radius r centered at the BS. For an FPRN v_j, we say that $v_j \in S_m$ if $mr < d_j \leq (m+1)r$.

Let w_j and rc_j be the workload and residual capacity of v_j. The total workload from all FPRNs in S_m is $W_m = \sum_{j:v_j \in S_m} w_j$ and the total residual capacity in S_m is $RC_m = \sum_{j:v_j \in S_m} rc_j$. Let $W_m^{M-1} = W_m + \ldots + W_{M-1}$ denote the total workload in shells $m, \ldots, M-1$, for $m = 1, \ldots, M-1$, and let C denote the total capacity of each RN.

Lemma. A lower bound on the optimal number of SPRNs is:

$$n_{\min} = \sum_{m=0}^{M-2} \max\left\{0, \left\lceil \frac{W_{m+1}^{M-1} - RC_m}{C} \right\rceil \right\} \quad (5)$$

Proof. As all traffic transmission is towards and terminates at the BS, traffic residing in S_m must be relayed across each of the inner shells $S_{m-1}, \ldots, S_1, S_0$. Thus, starting from the farthest shell S_{M-1}, W_{M-1} is the amount of workload that must pass through S_{M-2}. As the capacity of existing FPRNs is limited, some SPRNs may necessarily be added in S_{M-2} to carry W_{M-1}. Also, if the existing RNs are utilized completely before adding new nodes, the minimum number of SPRNs to be added in S_{M-2} is:

$$n_{\min}^{(M-2)} = \begin{cases} 0 & \text{if } RC_{M-2} \geq W_{M-1} \\ \left\lceil \frac{W_{M-1} - RC_{M-2}}{C} \right\rceil & \text{otherwise.} \end{cases} \quad (6)$$

Next the total workload relayed from S_{M-2} towards the BS will be W_{M-2}^{M-1}. In general, following the same line of logic, the total workload from S_{m+1} towards the BS will be the accumulated workload of all the nodes in the shells $S_{M-1}, S_{M-2}, \ldots, S_{m+1}$; i.e., W_{m+1}^{M-1}. Therefore, the minimum number of SPRNs to be added in the mth shell is:

$$n_{\min}^{(m)} = \begin{cases} 0 & \text{if } RC_m \geq W_{m+1}^{M-1} \\ \left\lceil \frac{W_{m+1}^{M-1} - RC_m}{C} \right\rceil & \text{otherwise.} \end{cases} \quad (7)$$

Then $\sum_{m=0}^{M-2} n_{\min}^{(m)}$ is a lower bound to the number of new nodes. □

We note that this lower bound depends only on the capacity, workloads and distances to the BS of the FPRNs. Further improvements to the bound are possible with more detailed knowledge about the physical locations of the FPRNs, though we do not pursue this here. Also, the bound is tight in that the equality is achievable for given capacity and workloads for some configurations of FPRNs. The bound is an equality, for example, when the FPRNs in shell S_m are fully utilized by the cumulative workload from shell $S_{m+1}, \ldots, S_{M-2}, S_{M-1}$, for $m = 0, 1, \ldots, M-2$.

VII. Performance Evaluation

As the placement of the FPRNs is optimal, in this section we will demonstrate the performance of our proposed algorithms for the placement of SPRNs by comparing with the lower bound in various scenarios. We first conduct the experiments in grid networks, and then check the algorithms in random networks and compare their performance in a statistical manner.

A. Performance Evaluation on Grid Networks

In order to obtain insight into the features of the three heuristic algorithms, we compare the number of SPRNs required by each algorithm in some representative scenarios, assuming FPRNs have been placed in a grid network.

The grid network is set up as follows. Sixteen FPRNs are placed at the vertices of a 4×4 grid with the BS to the left of the grid. As depicted in Fig. 5, a is the grid length, the BS is located at $(-b, \epsilon)$, where $\epsilon = 0.001a$ is a small perturbation value so that distances from FPRNs to the BS are distinct, and $b = 0.55a$. The nodes are numbered in an increasing order by their distance to the BS, i.e, node 16 is the farthest, while node 1 is the closest. All RNs have the same capacity of 8, namely, they can relay 8 units of traffic. The experimental scenarios are distinguished by the initial workload and transmission range of FPRNs.

We first conduct two groups of experiments with both the NTBF and MRCF algorithms. In the first group of experiments, the transmission range of RNs is $r = 1.2a$, so that each RN reaches its adjacent neighbors but not the diagonal neighbors. In the second group of experiments, the transmission range of RNs is enlarged to $r = 1.5a$. Thus, each RN has both adjacent neighbors and diagonal neighbors. Each group consists of two experiments, wherein the initial workloads on FPRNs (directly from SNs) are 2 and 4, respectively, in the two experiments. Accordingly, the lower bounds on the optimal number of SPRNs in the two experiments of the first group are 2 and 7, and 2 and 6 for the second group, respectively.

Fig. 5 illustrates the placement of SPRNs by the NTBF algorithm. The number of the SPRNs is 2 and 10, respectively, when the transmission range is small. Contrary to intuition, the number increases from 2 to 3 as the transmission range increases when the traffic load is 2. This undesired increase is due to the sensitiveness of traffic distribution to the distance. Even though the distance (to the BS) differences of two RNs are small, traffic will fill up the closer one first. For example, on the right side of Fig. 5, node 9 is only a little closer to the BS than node 10 is, but its final workload is twice that of node 10. This will further create an unbalanced traffic distribution for other nodes closer to the BS. This problem is due to the fact that the NTBF scheme prefers to fill up the capacity of the RNs closer to the BS. As a result, some traffic paths are inappropriately congested.

We examine the MRCF algorithm using the same two groups of experiments as above. The results when $r = 1.2a$ are the same as those for the NTBF algorithm. When $r = 1.5a$ and the traffic load is 2, the MRCF algorithm does not exhibit the load unbalance problem and the number of RNs is optimal (the number of SPRNs placed by the algorithm is equal to the

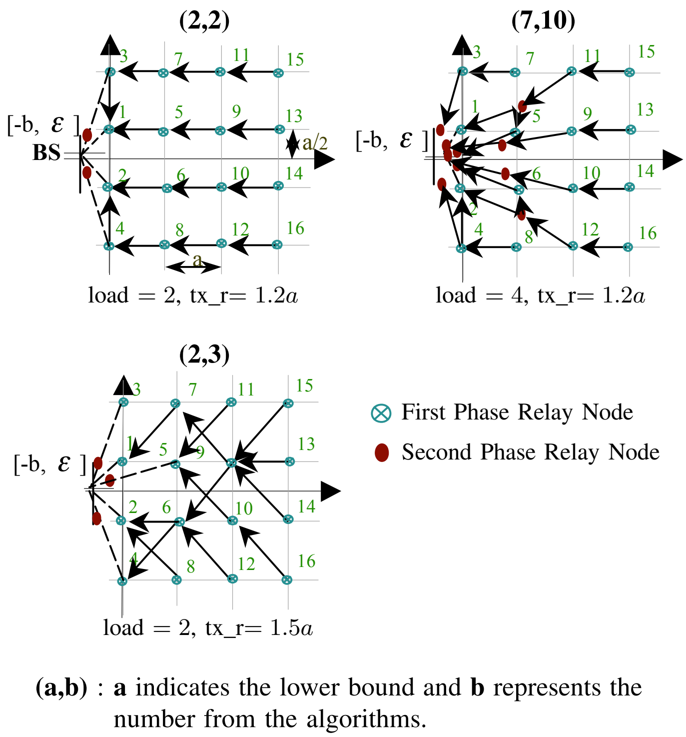


Fig. 5. SPRN placement and traffic distribution by the NTBF algorithm.

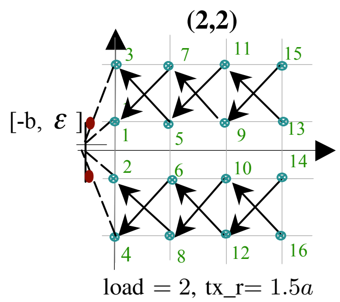


Fig. 6. SPRN placement and traffic distribution by the MRCF algorithm.

lower bound), as shown in Fig. 6. When the traffic load is 4, the MRCF algorithm requires the same number of RNs as the NTBF algorithm does.

However, the MRCF algorithm does not always yield better placement than what NTBF provides. We set up an experiment in which the FPRNs do not have equal initial workloads. Specifically, we set the FPRNs 1-4 with initial workload 2, 5-8 with initial workload 3, and 9-16 with initial workload 4. In the case $r = 1.2a$, the MRCF algorithm requires one more RN than the NTBF algorithm. The results are shown in Table I.

We further evaluate the BER algorithm in the above scenarios. In most cases, the BER algorithm achieves the optimal results. The results of all experiments and lower bounds are listed in Table I. The BER algorithm differs from the other two algorithms in two respects. First, an RN forwards its traffic load as much as possible to RNs closer to the BS. Thus, the capacity of existing RNs is maximally utilized before any new RN is added. Second, when adding a new SPRN at any time,

TABLE I
EXPERIMENTAL RESULTS IN GRID NETWORKS

Tx. range	Initial load	lower bound	BER	NTBF	MRCF
$1.2a$	2	2	2	2	2
	2,3,4	5	7	9	10
	4	7	8	10	10
$1.5a$	2	2	2	3	2
	2,3,4	5	5	6	6
	4	6	6	6	6

not only the farthest RN is considered, but also its intersecting node set is taken into account. In this way two or more RNs can share one common SPRN as their next hop relay and the RNs can achieve better utilization.

B. Performance Evaluation on Random Networks

As the number of potential design scenarios is immense, and the optimal placement is not available in general, in this section the three algorithms are executed on randomly generated networks. Their performance is evaluated and compared in a statistical manner.

The random network is on a square field of dimension 300×300, wherein the BS is located at the fixed position $(0, 0)$. A total of 100 FPRNs are distributed randomly in the field. The coordinates of the four corners of the square field are $(20, -150)$, $(320, -150)$, $(20, 150)$, and $(320, 150)$, the transmission range of a RN is 15, and the initial workload of the FPRNs has a uniform distribution on $[25, 75]$. Four experiments are executed, in which the capacity of RNs is set to 75, 100, 125, and 150 respectively. In each of the experiments, 500 independent random networks are generated and the results shown in this paper are the averages over the 500 runs. The simulation is written in C++. Simulation results in all scenarios hold a confidence interval no more than 2% of the average (over 500 runs) at a 95% confidence level. Four metrics are used to evaluate the performance of the three heuristic algorithms, as follows.

(1) Number of SPRNs. This criterion reflects the total device cost incurred in a system.
(2) Composite Energy Cost. This is the sum of capacities of all RNs required by each placement design. It indicates the system cost in a unified manner and makes the placement designs comparable across the four experiments.
(3) Average Capacity Utilization. This is the ratio of the sum of the workloads on RNs to the sum of their capacities. In this research, we calculate the average utilization of FPRNs and the average utilization of SPRNs separately.
(4) Coefficient of Variance (C.O.V.) of the capacity utilization. This is a measurement of load balance among the RNs, defined for a given network as the standard deviation of utilization of capacity divided by the average utilization over all nodes. The C.O.V. is computed separately for FPRNs and SPRNs.

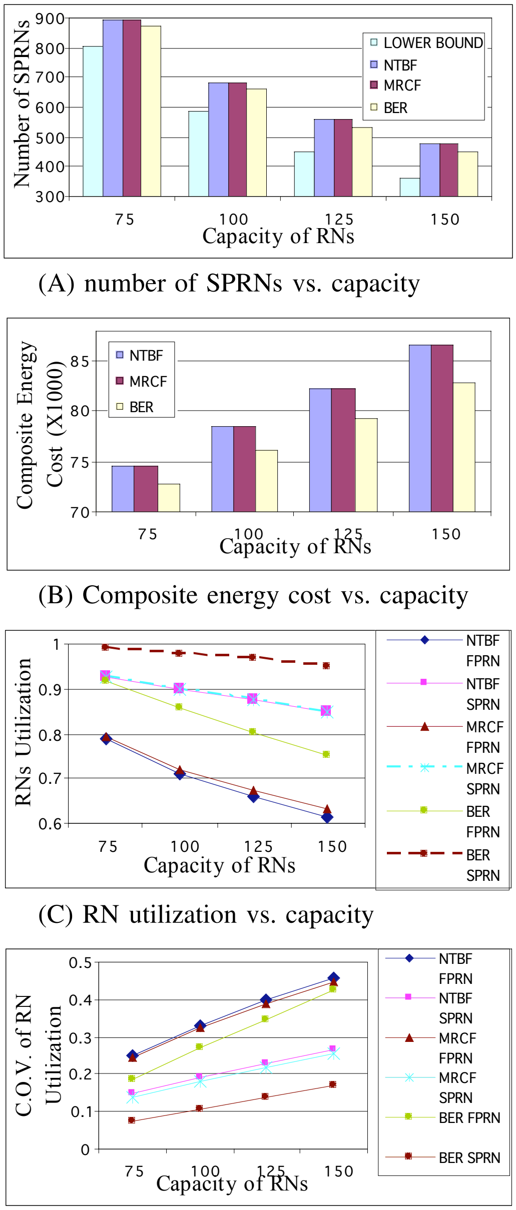
Fig. 7 (A-D) demonstrate the results by applying the three algorithms in random networks. From these figures, we observe the following.

1) All three algorithms require fewer SPRNs as the RNÕs capacity increases. This is a desirable property, as nodes with higher capacity would be able to relay more traffic than those with lower capacity.
2) For all three schemes the ratio of the number of SPRNs obtained from experiments to the theoretical lower bound is 1.1–1.3, which is quite acceptable.
3) The numbers of SPRNs used by the NTBF and the MRCF algorithms are similar to one another with the latter having smaller C.O.V. of the energy utilization, which supports the assertion that the MRCF algorithm achieves better load balancing among the nodes.
4) The BER algorithm outperforms the other two algorithms in terms of smaller number of SPRNs, higher utilization, and smaller C.O.V., though the improvement is moderate.
5) From Fig. 7 (B) and (C), we notice that with the increase of individual node capacity the composite energy cost increases while the resource utilization decreases. This suggests that an energy economical system may prefer to use a large number of devices with lower energy rather than using a smaller number of devices with higher energy. This could be helpful for the network designer to choose the devices.

VIII. Conclusion

In this paper, we explored the problem of optimal WSN device placement, aiming at minimizing the network cost with constraints on lifetime and connectivity. A general design problem was formulated and discussed. The placement problem with non-energy-constrained relay nodes was modeled as a minimum set covering problem. Furthermore, taking into account the energy and transmission range constraints on RNs, a comprehensive two phase approach was presented. Solutions for both phases were described. Based on the solution to problem one, an optimal solution was presented to place the first phase relay nodes. For the placement of second phase relay nodes, the Far-Near and Max-Min principles were proposed, and three heuristic schemes were developed accordingly. Analysis and numerical results indicate that the proposed algorithms can provide optimal or near optimal solution in some scenarios.

The relay node placement problem formulated in this paper is fundamental and critical in WSN design. The methodologies and results presented in this paper can provide a tangible guide for WSN provisioning. Future work includes investigating the optimal WSN design problem in further scenarios, such as when the transmission range of RNs is adaptive and/or the power supply (thus cost) of each RN is different. Also, of practical interest is the optimal traffic allocation issue to support the required traffic distribution. As the IEEE 802.15.4 becomes standardized, and with Zigbee's ongoing development for WSNs, we will also study how the optimal design can work with these standardized specifications.



(A) number of SPRNs vs. capacity

(B) Composite energy cost vs. capacity

(C) RN utilization vs. capacity

(D) C.O.V. of RN utilization vs. capacity

Fig. 7. Experimental results in random networks.

References

[1] J. Pan, Y. T. Hou, L. Cai, Y. Shi, and S. X. Shen, "Topology control for wireless sensor networks," in *Proc. ACM MOBICOM*, Sep. 2003, pp. 286–299.
[2] Y. T. Hou, Y. Shi, H. D. Sherali, and S. F. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2579–2590, Sep. 2005.
[3] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, IEEE Std. 802.15.4, 2003.
[4] K. Römer and F. Mattern, "The design space of wireless sensor networks," *IEEE Wireless Commun. Mag.*, vol. 11, no. 6, pp. 54–61, Dec. 2004.
[5] S. S. Dhillon and K. Chakrabarty, "Sensor placement for effective coverage and surveillance in distributed sensor networks," in *Proc. IEEE WCNC*, Mar. 2003, vol. 3, pp. 1609–1614.
[6] T. H. Cormen, C. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 2001.
[7] W. L. Winston, *Operations Research, Applications and Algorithms*, 3rd ed. Belmont California: Wadsworth Publishing Company, 1994.
[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocols for wireless microsensor networks," in *Proc. 33rd Hawaii Int. Conf. Syst. Sciences*, Jan. 2000, vol. 8, pp. 8020–8029.
[9] V. Srinivasan, C.-F. Chiasserini, P. S. Nuggehalli, and R. R. Rao, "Op-

timal rate allocation for energy-efficient multipath routing in wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 891–899, May 2004.

[10] J. Tang, B. Hao, and A. Sen, "Relay node placement in large scale wireless sensor networks," *Computer Commun.*, vol. 29, no. 4, pp. 490–501, Feb. 2006.

[11] S. Toumpis and L. Tassiulas, "Packetostatics: Deployment of massively dense sensor networks as an electrostactics problem," in *Proc. IEEE INFOCOM*, Mar. 2005, vol. 4, pp. 2290–2301.

[12] M. Patel, R. Chandrasekaran, and S. Venkatesan, "Energy efficient sensor, relay and base station placements for coverage, connectivity and routing," in *Proc. IEEE IPCCC*, Apr. 2005, pp. 581–586.

[13] M. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, and S. Singh, "Exploiting heterogeneity in sensor networks," in *Proc. IEEE INFOCOM*, Mar. 2005, vol. 2, pp. 878–890.

[14] V. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff, "A minimum cost heterogeneous sensor network with a lifetime constraint," *IEEE Trans. Mobile Comput.*, vol. 4, no. 1, pp. 4–15, Jan. 2005.

[15] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1992.

[16] C. V. Ramamoorthy and B. W. Wah, "The isomorphism of simple file allocation," *IEEE Trans. Comput.*, vol. 32, no. 3, pp. 221–231, Mar. 1983.

[17] G. C. Armour and E. S. Buffa, "A heuristic algorithm and simulation approach to relative location in facilities," *Manag. Sci.*, vol. 9, no. 1, pp. 294–309, Jan. 1963.

[18] A. S. Manne, "Plant location under economies of scale decentralization and computation," *Manag. Sci.*, vol. 11, pp. 213–235, Nov. 1964.

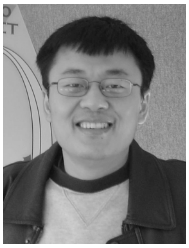
[19] K. Xu, Q. Wang, H. Hassanein, and G. Takahara, "Optimal wireless sensor networks (WSNS) deployment: Minimum cost with lifetime constraint," in *Proc. IEEE WiMob*, Aug. 2005, pp. 454–461.

[20] Q. Wang, K. Xu, H. Hassanein, and G. Takahara, "Minimum cost guaranteed lifetime design for heterogeneous wireless sensor networks (WSNS)," in *Proc. IEEE IPCCC*, Apr. 2005, pp. 599–604.

[21] T. X. Brown, H. N. Gabow, and Q. Zhang, "Maximum flow-life curve for a wireless ad hoc network," in *Proc. MOBIHOC*, Oct. 2001, pp. 128–136.

[22] Q. Wang, K. Xu, G. Takahara, and H. Hassanein, "Locally optimal relay node placement in heterogeneous wireless sensor networks," in *Proc. IEEE GLOBECOM*, Nov.–Dec. 2005, vol. 6, pp. 3549–3553.

Quanhong Wang (S'03-M'07) received her B.Eng and M.Sc.(Eng.) degrees in Electrical Engineering from Harbin Institute of Technology and Beihang University, China, respectively. She received her Ph.D. degree in the Department of Electrical and Computer Engineering at Queen's University, Canada, in 2006 and is currently working as a Post-Doctoral Fellow in the area of Sensor Networks at the Royal Military College of Canada in Kingston, Canada. She has authored and co-authored a number of papers in international scientific journals, book chapters and conferences in the area of wireless communication and networks. Her research interests are in Wireless Ad Hoc and Sensor Networks.



Kenan Xu (S'03) received the B.Eng degree in computer engineering from University of Electronics Science and Technology of China (UESTC) in 1997, and the M.Sc. degree from Queen's University, Kingston, Canada, in 2003. He is currently pursuing the Ph.D. degree in Computer Engineering at Queen's University. He has authored and/or co-authored more than 10 journal papers, book chapters and international technical conference papers. He has been awarded the NSERC graduate scholarship since 2004.

Glen Takahara (M'96) received the B.A. degree in mathematics from the University of British Columbia, Vancouver, B.C., in 1988, and the M.S. and Ph.D. degrees in statistics from Carnegie Mellon University in 1990, and 1994, respectively. Currently, he is an associate professor in Mathematics and Engineering in the Department of Mathematics and Statistics at Queen's University in Kingston, Ontario, Canada. His research interests are in applied probability modeling in communications networks, sensor networks, wireless ad hoc networks, optical networks, modulation, queueing theory, and statistics. He is a member of the IEEE Communications Society.

Hossam Hassanein (S'85-M'90-SM'05) is a leading researcher in the School of Computing at Queen's University in the areas of broadband, wireless and variable topology networks architecture, protocols, control and performance evaluation. Before joining Queen's University in 1999, he worked at the department of Mathematics and Computer Science at Kuwait University (1993-1999) and the department of Electrical and Computer Engineering at the University of Waterloo (1991-1993). Dr. Hassanein obtained his Ph.D. in Computing Science from the University of Alberta in 1990. He is the founder and director of the Telecommunication Research (TR) Lab http://www.cs.queensu.ca/ trl in the School of Computing at Queen's. Dr. Hassanein has more than 250 publications in reputable journals, conferences and workshops in the areas of computer networks and performance evaluation. Dr. Hassanein has organized and served on the program committee of a number international conferences and workshops. He is a senior member of the IEEE and serves as the Secretary of the IEEE Communication Society Technical Committee on Ad hoc and Sensor Networks (TC AHSN). Dr. Hassanein is the recipient of Communications and Information Technology Ontario (CITO) Champions of Innovation Research award in 2003.