Management of Orphaned-Nodes in Wireless Sensor Networks for Smart Irrigation Systems

Suhinthan Maheswararajah, Saman K. Halgamuge, *Member, IEEE*, Kithsiri B. Dassanayake, and David Chapman

***Abstract*—Wireless sensor networks may have unconnected nodes known as orphaned nodes due to their failure in obtaining a network address from a router-capable parent node in the network initialization process. The existence of a large number of orphaned nodes adversely affects the performance of wireless sensor network applications such as real time sensor-based automated irrigation control systems. We investigated the optimal management of orphaned nodes in a sensor network deployed in an automated irrigation system. In practice, it is unavoidable that sensor measurements contain random noise. The presence of orphan nodes adds to the effect of measurement noise further reducing the precision of irrigation management. However, reconnecting/restoring orphaned nodes to the sensor network may require some compromises to be made since the parent nodes are restricted in the maximum number of children they can possess. Optimal restoration can be achieved by finding the optimal parent node for each orphaned node that improves irrigation management. We propose two algorithms to suboptimally restore the orphaned nodes to the network, satisfying network constraints for small and large areas of farming. Numerical examples are presented to demonstrate the performance of the proposed methods.**

***Index Terms*—Kalman filters, optimization, sensor fusion, state estimation, wireless sensor networks (WSNs).**

I. Introduction

RECENT developments in integrating microelectromechanical systems and wireless technology have resulted in smaller and better performing wireless sensor nodes. An interconnected network of such devices, with the capacity to sense, process and transmit information with limitations, is called a wireless sensor network (WSN). Due to their cost-effective nature and deployment flexibility, WSNs have been successfully used in many real-world applications [1]–[4]. While WSNs have numerous advantages over other traditional approaches, operational parameters of the network need to be optimized to get the best performance. Moreover, these inexpensive sensor nodes are severely constrained in resources: processing speed, memory and especially in energy availability. For instance, reducing the communication range of the sensor nodes may result in degradation of the whole network connectivity. The network's life span will also be reduced, due to large amount of energy dissipation required for a long distance transmission when the communication range is set too high.

A sensor node delivers its measurements by establishing a connection to the base station (or coordinator) through single or multiple hop communication. In order to establish a network connection, it is required to obtain a network address (network ID) from a router-capable parent node [5]. In practice, this might not occur if the parent node is running out of network addresses due to inefficient address assignment or to a very low signal level received at the node, due to obstacles and undesirable atmospheric conditions. When a sensor node has lost its connectivity with other nodes or is unable to obtain a valid network ID, it becomes an orphan node. An orphan node may keep attempting to reconnect with its previous parent node immediately after the disconnection from the network, thus wasting energy [6]. Many strategies have been proposed to alleviate the orphaned node problem and its undesirable effects [5], [7]–[10]. In [5], a network formation heuristic algorithm was proposed to minimize the number of orphan nodes in a network for a given set of parameters and in [8], an optimized orphan node algorithm was proposed to alleviate associated problems such as delays due to the orphan process and power consumption. When a router-capable node becomes orphaned, its all descendants become orphaned too. This orphan propagation was controlled by using a scheme called enhanced self configuration [11], in which the orphan node can spread its orphan state to all its descendants immediately, allowing its descendants to take appropriate action. Minimizing the number of orphan nodes while maximizing the throughput of a network with bandwidth limitation was studied in [12]. An adaptive clustering technique was introduced in [9] and [13] to control the cluster size while limiting the number of orphan nodes. The salient feature of all these reported methods is to minimize the number of orphan nodes in the network. However, we approach the problem by maximizing the performance of the network by restoring the orphan node, although perhaps sacrificing some connected nodes.

This work is highly motivated by our experience with a Smarter Irrigation Project that studied implementing an automated irrigation system at Dookie College, University of Melbourne, Victoria, using a WSN [2]. The deployed sensor nodes capture a set of measurements around their physical location about soil moisture levels at different depths, atmospheric temperature and water inflow rate and transmit to the base station in a multi-hop fashion. The decision, whether activating an irrigation event or not, is based on gathered measurements from the WSN as well as various other environmental factors.

Manuscript received September 27, 2010; revised February 07, 2011 and May 23, 2011; accepted May 30, 2011. Date of publication June 23, 2011; date of current version September 14, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Anna Scaglione. This work was funded by the Victorian State Government, the University of Melbourne, and Goulburn-Murray Water through the Science and Technology Innovation (STI) Initiative and G08/DAAD Grant on Wireless Sensor Networks.

S. Maheswararajah, K. B. Dassanayake, and D. Chapman are with the Department of Agriculture and Food Systems, Melbourne School of Land and Environment, University of Melbourne, Victoria 3010, Australia (e-mail: smah@unimelb.edu.au; kbd@unimelb.edu.au; d.chapman@unimelb.edu.au).

S. K. Halgamuge is with the Department of Mechanical Engineering and MERIT theme Sustainable Systems and Energy, Melbourne School of Engineering, University of Melbourne, Victoria 3010, Australia (e-mail: saman@unimelb.edu.au).

Digital Object Identifier 10.1109/TSP.2011.2160258

Thus, the gathered measurements play a key role in the decision making process. However, in the Dookie experimental setup, it was observed that, on average, nearly half of the total sensor nodes are connected to the network at a given time during the test period, meaning that the other half are orphaned nodes. An orphaned sensor node may not be discarded if it is critical for activating an irrigation actuator or if its measurement significantly influences on decision making process of irrigation. In this paper, we focus on restoring connections of selected orphaned nodes back to the network despite sacrificing some of the connected nodes to improve the performance of the WSN. Therefore, in our approach the WSN's performance ultimately takes priority over simply decreasing the number of orphan nodes in the network.

In this study, we assume that the gathered sensor measurements are corrupted by white Gaussian noise and the state of the observed environmental phenomenon varies with time according to nonlinear Gaussian models. The estimation of the state of the environmental phenomenon is carried out at the base station using the extended Kalman filtering techniques. We assumed that there is a cost associated with the estimation error which is a strictly increasing function of estimated error. For instance, an irrigation actuator may be triggered when it is not required due to poor estimation of an environmental phenomenon and *vice versa* which leads to unnecessary cost. In this paper, we assumed that the presence of an orphan node can only be due either to all adjacent router-capable nodes (within the orphan node's communication range) already possessing the maximum number of children or to the connection link quality is not strong enough at the initial network establishment. Our objective is to minimize the total cost due to fault triggering of irrigation caused by poor estimation error. We further assume that the energy consumption rate for the sensor network is limited at any given time. For instance, the energy required for the sensor node's operation is provided by an external source such as solar power. We formulate the problem as a combinatorial optimization, which minimizes the total estimation error subject to the energy consumption rate imposed on the sensor network. We provide the computational complexity involved in using the explicit enumeration technique and propose an implicit enumeration technique instead. We study the problem when the environmental phenomenon varies with time as well as with space and time. We linearize our objective function and optimally solve the linearized problem using branch-and-bound (BB) method when the environmental phenomenon varies only with time. A heuristic hill-climbing technique is used to solve the problem when the environmental phenomenon varies with both time and space.

The remainder of this paper is organized as follows. Section II describes the orphan node problem in a Zigbee wireless platform and the models and estimation technique used in this study. In Section III, the formulation of the problem and the assumptions made for this study are presented. Moreover, we define the cost function with energy constraints and provide the maximum bound of the computational cost required to solve the problem. The proposed methods for orphan node management are presented in Section IV and their merits and demerits are discussed. Section V describes a numerical example which depicts the orphan node problem in a Zigbee WSN platform and results from the proposed methods are presented. Finally, Section VI concludes the research with a summary of the work presented in this paper.

II. Problem Description and System Models

A Zigbee network consists of a base station (or coordinator), routers and end-devices, where each end-device can be connected to a router or base station and each router can be connected to another router or base station [14]. The base station initializes the network establishment by advertising child-request messages. The same procedure is repeated by all router-capable nodes after they obtain their parents. Fig. 1 shows two different snapshots of established network topologies in a Zigbee WSN deployed in a farm field site in northern Victoria, Australia. The purpose of the network is to collect field measurements (temperature and soil moisture levels) from the deployed sensor nodes to perform automated irrigation, scheduling and control. The network shown in Fig. 1(a) has all nodes connected to the coordinator, whereas in Fig. 1(b), there are orphaned nodes. A sensor node becomes orphaned when it does not have a parent node and consequently the area covered by the node becomes unobservable. The orphaned node creation is mainly due either to its potential adjacent parent nodes already holding maximum number of children or to the poor quality of the communication link to its parent node during child-request advertising process. The presence of orphaned nodes may degrade the performance of the WSN, as the data gathered by the orphaned node may otherwise help to improve the precision of the automated irrigation system. We further explore the problem of restoring the orphaned node using the following scenario.

Fig. 2 shows a network where the parent node is allowed to have a maximum of three children. Let us assume that the measurement from orphaned node o_1 is less corrupted than measurements obtained from connected sensor nodes at a given time instant. As shown later in (11), fusing this measurement with the remaining measurements enhances the estimation of the environmental phenomenon and reduces faulty triggering of irrigation. The connection of o_1 to the network will contribute to the reduction of false triggering of irrigation.

The procedure of reconnecting involves the coordinator carefully examining which sensor node should sacrifice its connectivity to restore o_1. In this example, s_2 and s_3 are the adjacent parent nodes to o_1 in terms of communication range. Since we assume that a router-capable sensor node will not sacrifice its connectivity to restore an orphaned node, either s_6, s_7, s_8, s_9, s_{10}, or s_{11} can be sacrificed to restore the orphaned node o_1. Thus, the coordinator runs an algorithm and, for instance, selects s_7 to sacrifice its connectivity to restore o_1 as illustrated in Fig. 2(b). The selection criteria is based on the defined objective function and will be explained later in Section IV.

A. System Model

As we mentioned before, each sensor node gathers the atmospheric temperature and soil moisture level surrounding its location and transmits this information to the base station. The states of the phenomenon are modeled using a spatial-temporal stochastic system. Thus, the area of the farm is divided into g number of subgroups, denoted by $\theta \in \{a_1, a_2, \ldots .a_g\}$. Let the column vector $\mathrm{x}(\theta, k) = [t(\theta, k) \;\; m(\theta_1, k) \ldots m(\theta_l, k)]$ represents the true state of phenomena at time step k in the spatial location of θ, where $t(\theta, k)$ and $m(\theta_l, k)$ denote the temperature

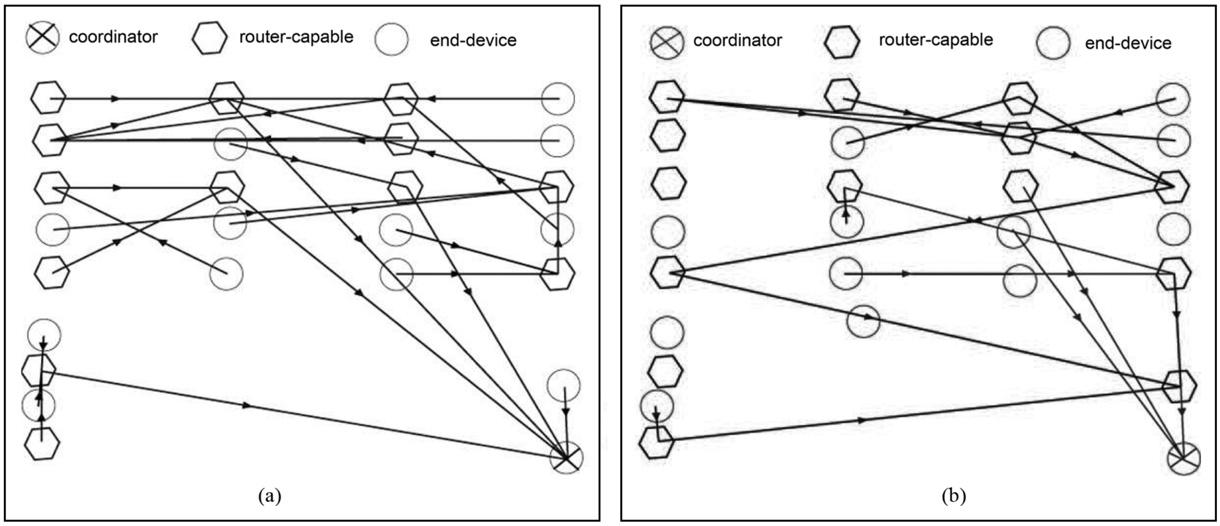


Fig. 1. Snap shot of two different topologies of a Zigbee wireless sensor network deployed at Dookie campus farm (a) Successful network establishment (b) Unsuccessful network establishment which produced orphaned nodes.

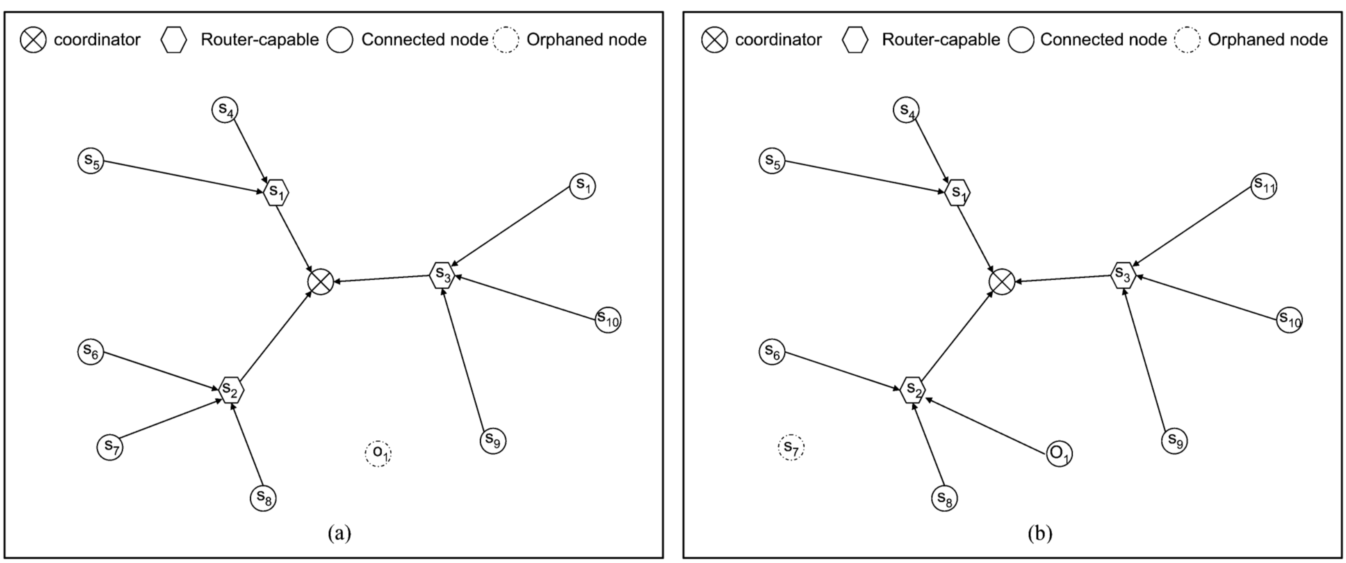


Fig. 2. Restoring an orphan node in WSN where a parent node is allowed to have a maximum of three children: (a) Network with orphaned nodes and (b) network after restoring the orphaned nodes.

and the soil moisture level at lth depth respectively. Gaussian stochastic model has been successfully used in indoor and outdoor environmental phenomena modeling [15]–[19]. We also assume that the states of the atmospheric temperature and soil moisture level are given by the nonlinear stochastic difference equation

$$\mathrm{X}(\theta, k) = f(\mathrm{X}(\theta, k-1)) + w(\theta, k-1) \tag{1}$$

where $f(.)$ is a nonlinear function relates the state at the previous time step $k-1$ to the state at the current time step k and $w(\theta, k)$ is assumed as the white Gaussian process noise, which is independent in space and time, with the known error covariance matrix $\mathrm{Q}(\theta, k)$, such that $w(\theta, k) \sim N(0, \mathrm{Q}(\theta, k))$. Let S denote the set of sensor nodes deployed in the farm and N denote the total number of sensor nodes in S. Each subarea, θ, has n_i number of sensor nodes, such that $\sum_{i=1}^{g} n_i = \mathrm{N}$. The connected sensor nodes in the ith subarea at time step k is denoted by $\mathrm{S}(i, k) = \left\{s_1^i, s_2^i, \ldots, s_{n(i,k)}^i\right\}$, thus, $n_i - n(i, k)$ number of sensor nodes became orphaned at time step k represented as $\mathrm{O}(i, k) = \left\{o_1^i, o_2^i, \ldots, o_{n_i - n(i,k)}^i\right\}$. Here, s_p^i and o_p^i represent the pth connected and orphaned node respectively in the ith subarea. Therefore, $\mathrm{X}(a_i, k)$ is observed by $n(i, k)$ number of sensor nodes, with the upper bound to n_i, at time step k. The measurement from the s_p^ith sensor node is given by

$$z_{s_p^i}(a_i, k) = h_{s_p^i}(\mathrm{X}(a_i, k)) + v_{s_p^i}(a_i, k) \tag{2}$$

where $h_{s_p^i}(.)$ is a nonlinear function relates the state $\mathrm{X}(a_i, k)$ to the measurement $z_{s_p^i}(a_i, k)$ and $v_{s_p^i}(a_i, k)$ is assumed to be an i.i.d. measurement noise, such that $v_{s_p^i}(a_i, k) \sim N(0, \mathrm{R}_{s_p^i}(a_i, k))$. We assume that the nonlinear function $h_{s_p^i}(.)$ and the error covariances are known *a priori* for each sensor node. Multi-hop transmission is used for the communication between the sensor node and the base station.

Each sensor node consumes energy in sensing, processing and transmitting the data. The energy consumption of the s_p^ith sensor node at time step k is given by

$$\psi(s_p^i,k) = \begin{cases} \psi_1 + \psi_{tx}(s_p^i,k), & \text{if transmitting} \\ \psi_1, & \text{otherwise} \end{cases} \tag{3}$$

where ψ_1 represents the energy consumed by sensing and processing data and $\psi_{tx}^i(p,k)$ represents the energy consumed by transmitting the data. The transmission power is calculated as $\psi_{tx}(s_p^i,k) = (\alpha_1 + \alpha_2 d_{i_p,j_q}^2)r$, where r denotes the data rate, α_1 denotes the electronic energy required to transmit one bit of data and α_2 is a constant related to the radio energy. d_{i_p,j_q} is the Euclidean distance between sensor node s_p^i and its parent node s_q^j.

B. Estimation

Since the implementation of the optimal filter for nonlinear state estimation is challenging due to huge memory space requirement, suboptimal algorithms, such as extended Kalman filtering (EKF), particle filter, unscented Kalman filter, etc., have been developed [20]. As the use of EKF is straight forward and the computational cost requirement is less than particle filter and unscented Kalman filter, we used EKF to estimate the states of the phenomena and its error covariance. In practice, optimal estimation methods for a nonlinear Gaussian model are computationally expensive and the EKF performs well, if the initial error and the noises are not too large [20]. The linearized equations of (1) and (2) using first-order Taylor series expansion method are given as follows:

$$\begin{aligned} \mathrm{x}(a_i,k) &\approx f(\hat{\mathrm{x}}(a_i,k-1)) + \mathrm{F}_{a_i,k} * (\mathrm{x}(a_i,k) \\ &\quad - \hat{\mathrm{x}}(a_i,k-1)) + w(a_i,k-1) \quad (4) \\ z_{s_p^i}(a_i,k) &\approx h_{s_p^i}(f(\hat{\mathrm{x}}(a_i,k-1))) + \mathrm{H}_{s_p^i,k} * (\mathrm{x}(a_i,k) \\ &\quad - f(\hat{\mathrm{x}}(a_i,k-1))) + v_{s_p^i}(a_i,k) \quad (5) \end{aligned}$$

where $\hat{\mathrm{x}}(a_i,k-1)$ is *a posteriori* estimate of the state $\mathrm{x}(a_i,k-1)$ at time step $k-1$ and, $\mathrm{F}_{a_i,k}$ and $\mathrm{H}_{s_p^i,k}$ are the Jacobian matrices of partial derivatives of $f(.)$ and $h(.)$ with respect to $\mathrm{x}(a_i,k)$ respectively and given as follows:

$$\mathrm{F}_{a_i,k} = \left.\frac{\partial f(\mathrm{x})}{\partial \mathrm{x}}\right|_{\mathrm{x}=\hat{\mathrm{x}}(a_i,k-1)} \tag{6}$$

$$\mathrm{H}_{s_p^i,k} = \left.\frac{\partial h_{s_p^i}(\mathrm{x})}{\partial \mathrm{x}}\right|_{\mathrm{x}=f(\hat{\mathrm{x}}(a_i,k-1))}. \tag{7}$$

At time step k, the stacked sensor measurements from the ith subarea can be described as

$$\begin{aligned} \mathbb{Z}(a_i,k) &= h_{a_i}(f(\hat{\mathrm{x}}(a_i,k-1))) + \mathrm{H}(a_i,k) * (\mathrm{x}(a_i,k) \\ &\quad - f(\hat{\mathrm{x}}(a_i,k-1))) + v(a_i,k) \end{aligned}$$

where

$$\mathbb{Z}(a_i,k) = \left[z'_{s_1^i}(a_i,k), z'_{s_2^i}(a_i,k), \ldots, z'_{s^i_{n(i,k)}}(a_i,k)\right]'$$

$$h_{a_i}(.) = \left[h'_{s_1^i}(.), h'_{s_2^i}(.), \ldots, h'_{s^i_{n(i,k)}}(.)\right]'$$

$$\mathrm{H}(a_i,k) = \left[\mathrm{H}'_{s_1^i,k}, \mathrm{H}'_{s_2^i,k}, \ldots, \mathrm{H}'_{s^i_{n(i,k)},k}\right]'$$

$$v(a_i,k) \sim N(0, \mathrm{R}(a_i,k)),$$

$$\mathrm{R}(a_i,k) = \begin{bmatrix} \mathrm{R}_{s_1^i}(a_i,k) & 0 & \cdots & 0 \\ 0 & \mathrm{R}_{s_2^i}(a_i,k) & \cdots & 0 \\ 0 & 0 & \cdots & \mathrm{R}_{s^i_{n(i,p)}}(a_i,k) \end{bmatrix}.$$

For a given set of stacked measurements $\mathbb{Z}(a_i,k) = \{\mathrm{z}(a_i,1), \mathrm{z}(a_i,2), \ldots, \mathrm{z}(a_i,k)\}$ from the ith subarea, the estimated state of $\mathrm{x}(a_i,k)$ and its error covariance are denoted as $\hat{\mathrm{x}}(a_i,k) = \mathbb{E}\{\mathrm{x}(a_i,k)|\mathbb{Z}(a_i,k)\}$ and $\hat{\mathrm{P}}(a_i,k) = \mathbb{E}\{(\mathrm{x}(a_i,k) - \hat{\mathrm{x}}(a_i,k))(\mathrm{x}(a_i,k) - \hat{\mathrm{x}}(a_i,k))'|\mathbb{Z}(a_i,k)\}$ respectively. At time step k, the estimated state is given by

$$\begin{aligned} \hat{\mathrm{x}}(a_i,k) &= \hat{\mathrm{x}}^-(a_i,k) + \mathrm{K}(a_i,k) * (\mathrm{z}(a_i,k) \\ &\quad - \mathrm{H}(a_i,k) * \hat{\mathrm{x}}^-(a_i,k)) \end{aligned} \tag{8}$$

where $\mathrm{K}(a_i,k)$ is the Kalman gain represented as $\mathrm{K}(a_i,k) = \hat{\mathrm{P}}(a_i,k) * \mathrm{H}(a_i,k)' * \mathrm{R}(a_i,k)^{-1}$ and $\hat{\mathrm{x}}^-(a_i,k)$ is the predicted state of $\mathrm{x}(a_i,k)$. The error covariance of the estimated state is an important part of our objective function and is given by

$$\hat{\mathrm{P}}(a_i,k)^{-1} = \hat{\mathrm{P}}^-(a_i,k)^{-1} + \mathrm{H}(a_i,k)' * \mathrm{R}(a_i,k)^{-1} * \mathrm{H}(a_i,k) \tag{9}$$

where

$$\hat{\mathrm{P}}^-(a_i,k) = \mathrm{F}_{a_i,k} * \hat{\mathrm{P}}(a_i,k-1) * \mathrm{F}'_{a_i,k} + \mathrm{Q}(a_i,k-1)$$

$$\hat{\mathrm{x}}^-(a_i,k) = f(\hat{\mathrm{x}}(a_i,k-1)).$$

Since we assume that sensor nodes' noises are cross-independent, part of (9) can be simplified as suggested in [21], thus

$$\mathrm{H}(a_i,k)'\mathrm{R}(a_i,k)^{-1}\mathrm{H}(a_i,k) = \underbrace{\sum_{p=1}^{n(i,k)} \mathrm{H}'_{s_p^i,k} * \mathrm{R}_{s_p^i}(a_i,k)^{-1} * \mathrm{H}_{s_p^i,k}}_{\hbar(\mathrm{s}(i,k))}. \tag{10}$$

We can rewrite the error covariance of the estimated state given in (9) by substituting (10) as

$$\hat{\mathrm{P}}(a_i,k)^{-1} = \hat{\mathrm{P}}^-(a_i,k)^{-1} + \hbar(\mathrm{s}(i,k)). \tag{11}$$

It can be inferred from (11) that the accuracy of the estimated state depends on the sensor nodes to be selected at time step k and the sensor nodes selected at time step $k-1$. Therefore, $\hat{\mathrm{P}}(a_i,k)$ produced by EKF cannot be calculated off line as the case in using Kalman filtering technique. However, it can be calculated one step prior to the measurement in hand. Selecting the optimal set of sensor nodes increases $\hbar(\mathrm{s}(i,k))$ in (11), and leads to higher accuracy of estimation.

III. Objective Function Formulation

Fig. 3 illustrates a sample irrigation strategy, where the irrigation is performed when the state of the phenomenon is in the shaded area. Let us consider two different instances where the true state of phenomena are denoted by $\mathrm{x}_1(\theta,k)$ and $\mathrm{x}_2(\theta,k)$,

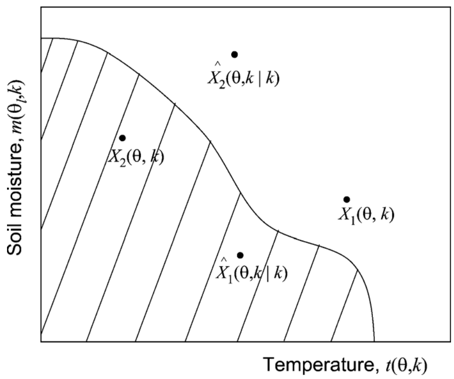


Fig. 3. Unsuccessful estimation of measurements leading to fault triggering of irrigation in a sample irrigation strategy.

and their estimates are denoted by $\hat{\mathrm{x}}_1(\theta,k)$ and $\hat{\mathrm{x}}_2(\theta,k)$. In the first instance, $\mathrm{x}_1(\theta,k)$ is in an unshaded area representing a relatively wetter area, not requiring irrigation. However, its estimated value $\hat{\mathrm{x}}_1(\theta,k)$ falls in the shaded area resulting in an unwanted irrigation event which may damage crops due to excessive water. In contrast, the second instance $\mathrm{x}_2(\theta,k)$ has caused damage to the crops due to shortage of water. In both cases, there is a cost in terms of water value, and possible damage to the crops due to the inaccurate estimation. Therefore, it is important to reduce the error in estimation.

Let $\mathrm{s}_k = \{\mathrm{s}(1,k), \mathrm{s}(2,k), \ldots, \mathrm{s}(g,k)\}$ and $\mathrm{o}_k = \{\mathrm{o}(1,k), \mathrm{o}(2,k), \ldots, \mathrm{o}(g,k)\}$ denote the set of connected sensor nodes and orphaned nodes in the network at time instant k, respectively. At any time instant, $\mathrm{s}_k \cup \mathrm{o}_k = \mathrm{s}$ and $|\mathrm{s}_k| + |\mathrm{o}_k| = \mathrm{N} = \sum_{i=1}^{g} n_i$. The set of connected sensor nodes s_k consists of M_k number of router-capable nodes at time step k denoted as $\mathrm{I}_k = \{r_1, r_2, \ldots, r_{\mathrm{M}_k}\}$, such that $\mathrm{I}_k \subseteq \mathrm{s}_k$ and $|\mathrm{I}_k| = \mathrm{M}_k$ (since the router-capable sensor nodes in the orphaned node set is not considered for this study, $\mathrm{I}_k \cap \mathrm{o}_k = \emptyset$). We assume that each router-capable sensor node can have a maximum of $c_{\max}$ number of children at any time instant k. The children of the mth router-capable sensor node, which has D_m number of children, is denoted as $\{c_1^m, c_2^m, \ldots, c_{\mathrm{D}_m}^m\}$, where c_d^m is the dth child node of the mth router and $c_d^m \in \mathrm{s}_k,\ \forall\, d = 1, 2, \ldots, \mathrm{D}_m$. Therefore, each router-capable node $r_m \in \mathrm{I}_k$ should satisfy the following condition:

$$\mathrm{D}_m \le c_{\max}, \quad \forall\, m = 1, 2, \ldots, \mathrm{M}_k. \tag{12}$$

We assume that each sensor node is within the communication ranges of the other sensor nodes. As mentioned in Section I, the energy consumption rate is limited to the network at each time instant k and defined as

$$\sum_{i=1}^{g} \sum_{p=1}^{n(i,k)} \psi(s_p^i, k) \le \mathrm{E}_k \tag{13}$$

where E_k denotes the total energy budget available for the whole network at time step k. It is possible that orphaned nodes may have a higher or lower accuracy of measurement than those of connected nodes. In the former case, it is of value to connect the orphaned nodes to the network by sacrificing some of the already connected nodes (refer to the assumption in Section II that a router-capable sensor node will not sacrifice its connectivity) in order to increase the accuracy of irrigation management. Minimizing the total cost due to false triggering of irrigation is equivalent to minimizing the total estimation error as we assume that it is strictly increasing with the increase of total estimation error. We need to define a scalar performance measure based on the calculated estimation error. There are several possible methods to define a scalar based on a given matrix such as trace, determinant or maximum eigenvalue and finding the trace of a matrix is less expensive than the rest in terms of computational cost. Therefore, we define our objective function as a sum of trace of estimated error covariances of the estimated states of each subarea given by

$$\mathrm{J}(\mathrm{s}_k) = \sum_{i=1}^{g} \mathrm{trace}(\hat{\mathrm{P}}(a_i, k)) \tag{14}$$

where $\mathrm{trace}(\hat{\mathrm{P}}(a_i,k))$ represents the trace of $\hat{\mathrm{P}}(a_i,k)$ which is the sum of the diagonal elements of $\hat{\mathrm{P}}(a_i,k)$. Our objective is to find out the optimal set of connected sensor nodes, which minimizes (14) subject to satisfying the constraints in (12) and (13). The optimal set of connected sensor nodes $\mathrm{s}_k^* = \{\mathrm{s}^*(1,k), \mathrm{s}^*(2,k), \ldots, \mathrm{s}^*(g,k)\}$ is obtained from

$$\mathrm{s}_k^* = \arg \min_{\forall s(i,k)} \sum_{i=1}^{g} \mathrm{trace}(\hat{\mathrm{P}}(a_i, k)) \tag{15}$$

subject to

$$\mathrm{D}_m \le c_{\max}$$
$$\sum_{i=1}^{g} \sum_{p=1}^{n(i,k)} \psi(s_p^i, k) \le \mathrm{E}_k.$$

Any exhaustive search technique can find the optimal solution by examining all possible solutions to the problem.

Let us consider a scenario where $g = 1$, i.e., θ is a single undivided area. We further consider three non-router-capable sensor nodes connected to the coordinator. Let $\mathrm{s}_k = \{s_1, s_2, s_3\}$, $\mathrm{o}_k = \{o_1, o_2\}$ and $\mathrm{I}_k = \emptyset$. The number of connected nodes, orphaned nodes and router-capable nodes are 3, 2, and 0, respectively. The set of possible solutions to the problem are as follows:

$$\{o_1, s_2, s_3\},\ \{s_1, o_1, s_3\},\ \{s_1, s_2, o_1\},\ \{o_2, s_2, s_3\},$$
$$\{s_1, o_2, s_3\},\ \{s_1, s_2, o_2\},\ \{o_1, o_2, s_3\},\ \{o_2, o_1, s_3\},$$
$$\{s_1, o_1, o_2\},\ \{s_1, o_2, o_1\},\ \{o_1, s_2, o_2\},\ \{o_2, s_2, o_1\},$$
$$\{s_1, s_2, s_3\}$$

For example, $\{o_1, s_2, s_3\}$ represents the solution where the connected node s_1 has been restored by the orphaned node o_1. Since some of the above solutions may not be feasible due to the constraints depicted in (13), the maximum possible solutions are 13 for this problem. It is possible that the computational cost needed to obtain the optimal solution may increase as the number of sensor nodes in the network increases. In the above example none of the connected sensor nodes are router-capable. If $\mathrm{M}_k > 0$, then the number of available restoring nodes reduces from $|\mathrm{s}_k|$ to $|\mathrm{s}_k| - \mathrm{M}_k$ as any of the router-capable sensor nodes will not sacrifice its connectivity. However, we should consider the number of children belonging to each router-capable node. Since each router-capable node r_m can have a maximum of $c_{\max}$ number of children, r_m can restore $c_{\max} - \mathrm{D}_m$ number of orphaned nodes without sacrificing any of the already connected nodes. Considering all possible permutations and combinations, the maximum number of possible solutions to this problem is given as follows (detailed derivation is presented in the Appendix):

$$\sum_{u=0}^{\min\{|\mathrm{O}_k|,f\}} |\mathrm{O}_k|_{\mathrm{C}_u} *^{f}_{\mathrm{P}_u}$$
$$\text{where } f = |\mathrm{S}_k| - \mathrm{M}_k + \sum_{m=1}^{\mathrm{M}_k}(c_{\max} - \mathrm{D}_m)$$
$$^{|\mathrm{O}_k|}\mathrm{C}_u = \frac{|\mathrm{O}_k|!}{(|\mathrm{O}_k| - u)! * u!}$$
$$^{f}\mathrm{P}_u = \frac{f!}{(f-u)!}. \tag{16}$$

For large $|\mathrm{O}_k|$ and f values, the number of possible solutions would be huge and exhaustive search methods would take a long time to find the optimal solution. Therefore, as discussed in the next section, we proposed new methods to solve this problem in an acceptable time.

IV. Orphaned Node Management

In this section, we analyze the problem for two different scenarios:

1) the farm area is small enough to be considered as a single homogeneous farming area: the state of phenomenon varies only with time but not with space. Thus, θ represents the whole farm and $g = 1$;
2) the farm area is large and therefore to be considered as multiple heterogeneous farming areas: the state of phenomenon varies with time and space. Thus, the farm area is divided into many subareas and $g > 1$.

We formulate the first problem as a binary integer programming and optimally solve by branch-and-bound algorithm. Since the second problem is more complex than the first one, we use a suboptimal method called random-bit-climbing (RBC) technique to solve the problem.

The assumptions made in this study are as follows:

1) each sensor node is within the transmission ranges of other sensor nodes in the network;
2) orphaned-nodes are created only because the received signal strengths are below an accepted level either at the parent or child node during the child request process of the network initialization;
3) a router-capable sensor node will not sacrifice its connectivity for restoring an orphaned-node;
4) none of the orphaned-nodes are exhausted nodes.

A. Single Homogeneous Farming Area

This scenario is suitable for a small farming area with the single irrigation actuator where the state of phenomenon is uniformly distributed over the space. For this single homogeneous farming area, the stochastic equations depicted in (1) and (2) are reduced to

$$\mathrm{X}_k = f(\mathrm{X}_{k-1}) + w_{k-1} \tag{17}$$
$$z_k^{s_p} = h_{s_p}(\mathrm{X}_k) + v_k^{s_p}. \tag{18}$$

For simplicity, the following notations will be used in this subsection:

- $\mathrm{S}_k = \{s_1, s_2, \ldots, s_{n_k}\}$ denotes the connected nodes at time step k;
- $\mathrm{O}_k = \mathrm{O}(i,k) = \{o_1, o_2, \ldots, o_{\mathrm{N}-n_k}\}$ denotes the orphaned nodes at time step k;
- $\mathrm{I}_k = \{r_1, r_2, \ldots, r_{\mathrm{M}_k}\} \subseteq \mathrm{S}_k$ denotes the router-capable nodes which are connected at time step k;
- $\mathrm{H}_{s_p^i,k} \longrightarrow \mathrm{H}_k^{s_p}$ denotes the Jacobian matrix of the pth connected node;
- $\mathrm{H}_{o_q^i,k} \longrightarrow \mathrm{H}_k^{o_q}$ denotes the Jacobian matrix of the qth orphaned node;
- $\mathrm{R}^{s_p}(\theta,k) \longrightarrow \mathrm{R}_k^{s_p}$ denotes the error covariance of the pth connected node;
- $\mathrm{R}^{o_q}(\theta,k) \longrightarrow \mathrm{R}_k^{o_q}$ denotes the error covariance of the qth orphaned node;
- $\hat{\mathrm{P}}(\theta,k) \longrightarrow \hat{\mathrm{P}}_k$ denotes the error covariance of the estimated States of the phenomenon;
- $\psi(s_p^i,k) \longrightarrow \psi_k^{s_p}$ denotes the energy consumption of the pth connected node;
- $\psi_k^{o_q}$ denotes the energy consumption of the qth orphaned node.

We can rewrite the objective function in (14) for the single area problem:

$$\mathrm{J}(\hbar(\mathrm{S}_k)) = \mathrm{trace}(\hat{\mathrm{P}}_k) = \mathrm{trace}\left(\left[\hat{\mathrm{P}}_k^{-^{-1}} + \hbar(\mathrm{S}_k)\right]^{-1}\right),$$
$$\text{where } \hbar(\mathrm{S}_k) = \sum_{p=1}^{n_k} \mathrm{H}_k^{s_p'} \mathrm{R}_k^{s_p^{-1}} \mathrm{H}_k^{s_p}. \tag{19}$$

It can be inferred from (19) that $\mathrm{J}(\hbar(\mathrm{S}_k))$ is a nonlinear function of $\hbar(\mathrm{S}_k)$ and finding the optimal solution $\hbar(\mathrm{S}_k^*)$ using exhaustive search methods may take a long time if the number of possible solutions obtained from (16) is large. Therefore, we first approximately linearize the problem and then optimally solve the linearized problem using implicit enumeration technique.

We use the first-order Taylor series expansion method to linearize our objective function defined in (19) and given by

$$\begin{aligned}\mathrm{J}(\hbar(\mathrm{S}_k)) &\approx \mathrm{trace}\Bigg(\left[\hat{\mathrm{P}}_k^{-^{-1}} + \hat{\hbar}(\mathrm{S}_k)\right]^{-1} \\ &\quad + \mathrm{J}'(\hat{\hbar}(\mathrm{S}_k)) * \left[\hbar(\mathrm{S}_k) - \hat{\hbar}(\mathrm{S}_k)\right]\Bigg), \\ &\approx \mathrm{J}(\hat{\hbar}(\mathrm{S}_k)) + \mathrm{trace}\left(\mathrm{J}'(\hat{\hbar}(\mathrm{S}_k)) * \left[\hbar(\mathrm{S}_k) - \hat{\hbar}(\mathrm{S}_k)\right]\right).\end{aligned} \tag{20}$$

where $\hat{\hbar}(\mathrm{S}_k)$ is calculated using the connected sensor nodes before the restoring process takes place and $\mathrm{J}'(\hat{\hbar}(\mathrm{S}_k))$ is the first-order differentiation of $\mathrm{J}(\hbar(\mathrm{S}_k))$, at $\hbar(\mathrm{S}_k) = \hat{\hbar}(\mathrm{S}_k)$, with respect to $\hbar(\mathrm{S}_k)$ and given by

$$\mathrm{J}'(\hat{\hbar}(\mathrm{S}_k)) = \left.\frac{d\mathrm{J}(\hbar(\mathrm{S}_k))}{d\hbar(\mathrm{S}_k)}\right|_{\hbar(\mathrm{S}_k)=\hat{\hbar}(\mathrm{S}_k)}.$$

By using the matrix differentiation [22, p. 372], $\mathrm{J}'(\hbar(\mathrm{S}_k))$ can be given as

$$\mathrm{J}'(\hbar(\mathrm{S}_k)) = -\left[\hat{\mathrm{P}}_k^{-^{-1}} + \hbar(\mathrm{S}_k)\right]^{-1} * \frac{d\left[\hat{\mathrm{P}}_k^{-^{-1}} + \hbar(\mathrm{S}_k)\right]}{d\hbar(\mathrm{S}_k)}$$

$$
\begin{aligned}
&* \left[\hat{\mathrm{P}}_k^{-^{-1}} + \hbar(\mathrm{s}_k) \right]^{-1} \\
&= - \left[\hat{\mathrm{P}}_k^{-^{-1}} + \hbar(\mathrm{s}_k) \right]^{-2}.
\end{aligned}
$$

For simplicity, we rewrite (20) as follows:

$$
\begin{aligned}
\mathrm{J}(\hbar(\mathrm{s}_k)) &\approx C + \mathrm{trace}(M * \hbar(\mathrm{s}_k)) \\
\text{where } M &= \mathrm{J}'(\hat{\hbar}(\mathrm{s}_k)) \\
C &= \mathrm{J}(\hat{\hbar}(\mathrm{s}_k)) - \mathrm{trace}(\hat{\hbar}(\mathrm{s}_k)).
\end{aligned} \tag{21}
$$

Now we can apply a linear optimization technique on our approximated objective function defined in (21). Let $p \in \{1, 2, \ldots, n_k\}$, $q \in \{1, 2, \ldots, \mathrm{N} - n_k\}$ and $m \in \{1, 2, \ldots, \mathrm{M}_k\}$. We define A_q^p and B_q^m as binary variables, where $\mathrm{A}_q^p = 1$ means that the connected sensor node s_p sacrifices its connectivity to restore the orphaned-node o_q, otherwise it remains connected with the network and $\mathrm{B}_q^m = 1$ means that the orphaned-node o_q become a child node to the mth router-capable node without disconnecting any of its (mth router) children, otherwise it will not. Since a router-capable node will not sacrifice its connectivity, $\mathrm{A}_q^p = 0 \; \forall \; s_p \in i_k$. Minimizing our problem (20) can be achieved by minimizing the following function, which has been derived from (20) with the binary variables A_q^p and B_q^m:

$$
\begin{aligned}
J = C + \mathrm{trace} \Bigg\{ & M \sum_{p=1}^{n_k} \mathrm{H}_k^{s'_p} \mathrm{R}_k^{s_p^{-1}} \mathrm{H}_k^{s_p} \\
& + M \sum_{q=1}^{\mathrm{N}-n_k} \sum_{p=1}^{n_k} \mathrm{A}_q^p \left(\mathrm{H}_k^{o'_q} \mathrm{R}_k^{o_q^{-1}} \mathrm{H}_k^{o_q} - \mathrm{H}_k^{s'_p} \mathrm{R}_k^{s_p^{-1}} \mathrm{H}_k^{s_p} \right) \\
& + M \sum_{q=1}^{\mathrm{N}-n_k} \sum_{m=1}^{\mathrm{M}_k} \mathrm{B}_q^m \mathrm{H}_k^{o'_q} \mathrm{R}_k^{o_q^{-1}} \mathrm{H}_k^{o_q} \Bigg\}
\end{aligned} \tag{22}
$$

subject to

$$
\sum_{p=1}^{n_k} \mathrm{A}_q^p + \sum_{m=1}^{\mathrm{M}_k} \mathrm{B}_q^m \leq 1 \quad \forall \, q = 1, 2, \ldots, (\mathrm{N} - n_k) \tag{23}
$$

$$
\sum_{q=1}^{(\mathrm{N}-n_k)} \mathrm{A}_q^p \leq 1 \quad \forall \, p = 1, 2, \ldots, n_k \tag{24}
$$

$$
\sum_{q=1}^{(\mathrm{N}-n_k)} \mathrm{B}_q^m \leq c_{\max} - \mathrm{D}_m \quad \forall \, m = 1, 2, \ldots, \mathrm{M}_k \tag{25}
$$

$$
\begin{aligned}
&\sum_{\forall \mathrm{S}_k} \psi_k^i + \sum_{q=1}^{\mathrm{N}-n_k} \sum_{\forall s_p \in \mathrm{S}_k - \mathrm{I}_k} \mathrm{A}_q^p (\psi_k^{o_q} - \psi_k^{s_p}) \\
&\quad + \sum_{q=1}^{\mathrm{N}-n_k} \sum_{m=1}^{\mathrm{M}_k} \mathrm{B}_q^m \psi_k^{o_q} \leq \mathrm{E}_k
\end{aligned} \tag{26}
$$

where

$$
\begin{aligned}
\mathrm{A}_q^p &\in \begin{cases} 0, & \text{if } s_p \text{ is router-capable node} \\ \{0, 1\}, & \text{if otherwise} \end{cases} \\
\mathrm{B}_q^m &\in \{0, 1\} \; \forall q.
\end{aligned}
$$

Finding the optimal set of connected nodes S_k^* obtained from (19) is equivalent to finding the optimal set of variables A^*

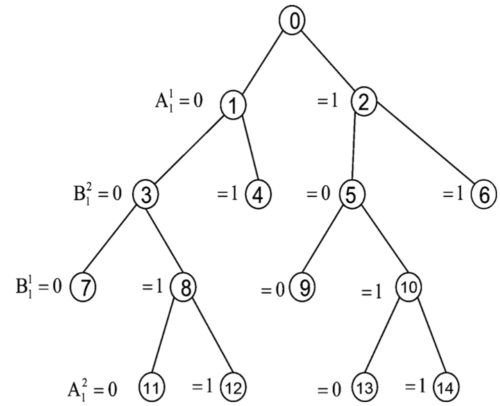


Fig. 4. Systematic enumeration in branch-and-bound method.

and B^*, where $\mathrm{A}^* = \left\{ \mathrm{A}_1^{1^*}, \mathrm{A}_1^{2^*}, \ldots, \mathrm{A}_1^{n_k^*}, \mathrm{A}_2^{1^*}, \ldots, \mathrm{A}_{\mathrm{N}-n_k}^{n_k^*} \right\}$ and $\mathrm{B}^* = \left\{ \mathrm{B}_1^{1^*}, \mathrm{B}_1^{2^*}, \ldots, \mathrm{B}_1^{\mathrm{M}_k}, \mathrm{B}_2^{1^*}, \ldots, \mathrm{B}_{\mathrm{N}-n_k}^{\mathrm{M}_k} \right\}$.

It can be observed from (22) that J is a linear function of the binary variables A_q^p and B_q^m with a set of linear inequality constraints. The linear constraints given from (23) to (26) are explained as follows:

- constraint-(23): to make sure that an orphaned node will be restored with only one router-capable node;
- constraint-(24): to make sure that a connected node sacrifices its connectivity to restore only one orphaned node;
- constraint-(25): to make sure that a router-capable node satisfies its constraint on the maximum possible number of children;
- constraint-(26): to depict the energy budget for the whole network.

The BB method is a non heuristic technique, widely used for solving discrete and combinatorial optimization problems. Fig. 4 illustrates BB's systematic enumeration for a simple problem. Two main actions branching and bounding, are involved in its searching procedure. A node represents a partial or complete solution to the problem. Branching takes place when a node is selected for further growth and the next generation of that node is created. Selecting a node and the variable for the next growth is decided by a predefined selection policy. The bounding function is an optimistic estimator, such that it underestimates the actual best achievable objective value for a minimization problem and overestimates for a maximization problem. Thus, BB only examines a small fraction of the full enumeration. For instance, nodes 4, 6,7, and 9 have been discarded for further growth, as the value obtained from the bounding function is not promising. Therefore, the main intention in BB is to minimize computational cost by discarding the infeasible and nonpromising solutions while considering all feasible solutions. As shown in Fig. 4, BB systematically grows only the most promising nodes at any stage using node selection and variable selection policies.

The linear relaxation of the original problem is used as the bounding function denoted as $\tilde{J}$, such that the binary constraints are relaxed to be between 0 and 1, and the following condition holds:

$$
\arg \min_{\tilde{\mathrm{A}}_q^p, \tilde{\mathrm{B}}_q^m} \tilde{J} \leq \arg \min_{\mathrm{A}_q^p, \mathrm{B}_q^m} J \tag{27}
$$

where $\tilde{\mathrm{A}}_q^p$ and $\tilde{\mathrm{B}}_q^m$ denote the relaxed constraints of A_q^p and B_q^m respectively. It can be inferred from (27) that the optimal solution to the relaxed problem $\tilde{J}$ is an upper-bound of the original problem J. Let, $\{\tilde{\mathrm{A}}_t^*, \tilde{\mathrm{B}}_t^*\}$ denote the partial solution to J and the optimal solution to $\tilde{J}$ at tth node. The main steps involved in BB are given as follows.

- Step 1) Initially, the linear relaxation problem is considered and $\{\tilde{\mathrm{A}}_0^*, \tilde{\mathrm{B}}_0^*\}$ is stored at $t = 0$. If $\{\tilde{\mathrm{A}}_0^*, \tilde{\mathrm{B}}_0^*\}$ is a binary solution, then the solution to J is $\{\tilde{\mathrm{A}}_0^*, \tilde{\mathrm{B}}_0^*\}$, otherwise generate two children nodes as explained in Step 2).
- Step 2) Select a variable from $\{\tilde{\mathrm{A}}_t^*, \tilde{\mathrm{B}}_t^*\}$ which is closer to 0 or 1 but not equal to them. Branching on the selected variable and generate two children nodes $t+1$ and $t+2$ by making the selected variable 0 and 1 respectively. Note that we use the variable selection strategy as selecting a variable closer but not equal to 0 or 1.
- Step 3) Solve the linear relaxation of both children nodes $t+1$ and $t+2$ with frozen values of previously selected variables. If $\{\tilde{\mathrm{A}}_{t+1}^*, \tilde{\mathrm{B}}_{t+1}^*\}$ is a binary solution, then the $t+1$th node is the candidate solution, and set the value of $\tilde{J}_{t+1}$ as the updated lower-bound.
- Step 4) If any unbranched nodes have values of $\tilde{J}$, which are less than the current lower-bound, they will be pruned and further growth is stopped.
- Step 5) If there are no nodes to branch for further growth, the solution to the original problem is the current candidate node. Otherwise select an unbranched node with the largest value of $\tilde{J}$ and go to Step 2). Note that we use the node selection strategy to select the node with the largest value.

For our problem, the number of elements in A_q^p and B_q^m are $(\mathrm{N} - n_k)n_k$ and $\mathrm{M}_k(\mathrm{N} - n_k)$ respectively. However, $(\mathrm{N} - n_k)(\mathrm{M}_k)$ number of elements in A_q^p are zeros, as $\mathrm{A}_q^p = 0\,\forall\, s_p \in i_k$. Thus, the total number of variables for this problem is given as $(\mathrm{N} - n_k)(n_k - \mathrm{M}_k) + \mathrm{M}_k(\mathrm{N} - n_k) = n_k(\mathrm{N} - n_k)$. Therefore, an exhaustive method needs to search a maximum of $2^{n_k(\mathrm{N}-n_k)}$ possible solutions to find the optimal solution. However, using branching and bounding strategies as explained in Step 5) brings the computational cost below the total number of enumerations.

B. Multiple Heterogeneous Farming Areas

In this section, we assume that each state of phenomena varies with time as well as space and its stochastic equation is represented in (1). This scenario is more suitable for a big farm area, which has multiple irrigating actuators. The objective function of the multiple area problem is the same as in (14), subject to satisfying (12) and (13) and given by

$$\mathrm{J}(\mathrm{S}_k) = \sum_{i=1}^{g} \mathrm{trace}\left(\left[\hat{\mathrm{P}}^-(a_i,k)^{-1} + \bar{h}(\mathrm{S}(i,k))\right]^{-1}\right). \tag{28}$$

The objective function for a multiple area problem, defined in (28), is a nonlinear function with respect to $\mathrm{R}_{s_p^i}(a_i,k)$ and therefore, a nonlinear optimization method should be used to solve the problem. We used the random-bit-climbing (RBC) method to solve this problem.

The RBC, a suboptimal method, was first introduced in [23]. It begins with a randomly generated single parent solution that may be replaced by an offspring only if it can produce a better performing offspring by mutating one of its (randomly selected) dimensions. In the original version of RBC, the parent solution is instantaneously replaced by an improved offspring as opposed to generating several offspring and selecting the best to replace the parent. If every dimension in the parent solution has been checked and no improvement is found before the termination conditions reached, the procedure is restarted from a new random parent solution. Since RBC only mutates a single dimension at a time, the mutated solution always satisfies the children constraint as stipulated in (12). This feature of RBC makes it applicable to our problem, unlike PSO/GA (citation required) where changes occur in more than one dimension simultaneously. In order to satisfy the energy constraint (13), we need to define a probability for each orphaned node based on its energy consumption. Let $\mathrm{O}_k(\tau)$ and $\mathrm{S}_k(\tau)$ denote the set of orphaned and connected sensor nodes at τth iteration respectively. The probability $\gamma_\tau(o_q, s_p^i)$ of selecting the orphaned node $o_q \in \mathrm{O}_k(\tau)$ to replace the connected node $s_p^i \in \mathrm{S}_k(\tau)$ at τth iteration is defined as follows:

$$\gamma_\tau(o_q, s_p^i) = \begin{cases} \dfrac{\mathrm{G}(o_q,s_p^i)}{\sum\limits_{\forall o_q \in \mathrm{O}_k(\tau)} \mathrm{G}(o_q,s_p^i)}, & \text{if } s_p^i \notin \mathrm{I}_k, \\ 0, & \text{if otherwise} \end{cases} \tag{29}$$

where

$$\mathrm{G}(o_q, s_p^i) = \begin{cases} 1, & \text{if } \lambda(o_q, s_p^i) \le \mathrm{E}_k \\ 0, & \text{if otherwise} \end{cases}$$

$$\lambda(o_q, s_p^i) = \sum_{j=1}^{g} \sum_{\forall s_r^j \in \mathrm{S}_k(\tau)} \psi(s_r^j, k) + \psi(o_q, k) - \psi(s_p^i, k).$$

Similarly, the probability $\gamma_\tau(o_q, c_d^m)$ of selecting the orphaned node $o_q \in \mathrm{O}_k(\tau)$ and connecting to the mth router-capable node as its dth child at τth iteration, where $d \ge \mathrm{D}_m$, is defined as

$$\gamma_\tau(o_q, c_d^m) = \frac{\mathrm{G}(o_q, c_d^m)}{\sum\limits_{\forall o_q \in \mathrm{O}_k(\tau)} \mathrm{G}(o_q, c_d^m)} \tag{30}$$

where

$$\mathrm{G}(o_q, c_d^m) = \begin{cases} 1, & \text{if } \lambda(o_q, m) \le \mathrm{E}_k \\ 0, & \text{if } \lambda(o_q, m) > \mathrm{E}_k \text{ or } d > c_{\max} \end{cases} \tag{31}$$

$$\lambda(o_q, m) = \sum_{i=1}^{g} \sum_{\forall s_r^i \in \mathrm{S}_k(\tau)} \psi(s_r^i, k) + \psi(o_q^i, k). \tag{32}$$

Zero probability of selecting an orphaned node indicates that the restoring process will violates either the constraint on energy or on the maximum number of children and therefore the particular orphaned node will not be selected for the restoring process. Let us define the Roulette-wheel selection function $\Pi(cdf(\gamma_\tau(o_q, s_p^j)))$ to select an orphaned node from $\mathrm{O}_k(\tau)$ to replace the connected sensor node s_p^j using $\gamma_\tau(o_q, s_p^j)$, where cdf denotes the cumulative distribution function. Similarly, $\Pi(cdf(\gamma_\tau(o_q, c_d^m)))$ selects an orphaned node from $\mathrm{O}_k(\tau)$ to restore at the mth router-capable node as its dth child. The solution vector has two parts:

1) Part 1: set of connected sensor nodes;
2) Part 2: number of unassigned children slots available at each router-capable nodes.

RBC uses $\Pi(cdf(\gamma_\tau(o_q, s_p^j)))$ and $\Pi(cdf(\gamma_\tau(o_q, c_d^m)))$ to mutate Part 1 and Part 2 of the solution respectively. Initially none of the orphaned nodes are restored to the network and therefore the initial solution vector is the same as the set of connected sensor nodes below:

$$\mathrm{sol}_0 = \left\{ \underbrace{\mathrm{s}(1,k), \mathrm{s}(2,k), \ldots, \mathrm{s}(g,k)}_{\text{Part 1}} \; \underbrace{\{\emptyset\}}_{\text{Part 2}} \right\} \tag{33}$$

where $\{\emptyset\}$ in Part 2 denotes the empty set meaning that in the initial solution none of the orphaned nodes have been restored to the unassigned children slots available at router-capable nodes. However, the maximum number of elements that can reside in Part 2 is the total number of unassigned children slots available given by $\sum_{m=1}^{\mathrm{M}_k}(c_{\max} - \mathrm{D}_m)$. Therefore, the maximum number of dimensions of the solution vector, denoted as Λ, is calculated as the sum of the number of connected nodes and total number of unassigned children slots available in each router-capable node, such that $\Lambda = |\mathrm{S}_k| + \sum_{m=1}^{\mathrm{M}_k}(c_{\max} - \mathrm{D}_m)$.

We used RBC with a minor modification in its parent replacement, such that we generate a certain number of offspring from sol_τ by mutating each dimension separately and by replacing the parent solution by a better performing offspring. It can be inferred from (29) that $\gamma_\tau(o_q, s_p^i) = 0$, if s_p^i is a router-capable node, leading to only $|\mathrm{S}_k| - \mathrm{M}_k$ number of connected sensor nodes can be switching to orphaned nodes in Part 1 of the solution vector. Therefore, we can only generate $\Lambda - \mathrm{M}_k$ number of offspring ignoring the router-capable nodes. We use the maximum number of iterations as the termination condition for RBC denoted as MAXITER. The main steps involved with RBC for our problem are given as follows.

- Step 1) At iteration $\tau = 0$, set $\mathrm{sol}_0 = \{\mathrm{s}(1,k), \mathrm{s}(2,k), \ldots, \mathrm{s}(g,k), \{\emptyset\}\}$.
- Step 2) At the τth iteration, generate $\Lambda - \mathrm{M}_k$ number of offspring from sol_τ by switching each dimension separately using $\gamma_\tau(o_q, s_p^j)$ and $\Pi(cdf(\gamma_\tau(o_q, c_d^m)))$.
- Step 3) Calculate fitness for each offspring using (28) and select the offspring with the lowest value out of $\Lambda - \mathrm{M}_k$ number of mutated offspring.
- Step 4) Replace the current parent solution by the selected offspring in Step 3), if its fitness value is less than that of parent.
- Step 5) Update $\mathrm{S}_k(\tau)$, $\mathrm{O}_k(\tau)$ and the number of unassigned children slots available in each router-capable node.
- Step 6) if τ equals MAXITER then go to Step 7), else increment τ by one and go to Step 2).
- Step 7) The suboptimal solution to our problem is given by sol_τ.

To illustrate the algorithm better, we describe the procedure of RBC algorithm for a simple problem as illustrated in Fig. 5, where a farm is divided into two separate areas, such that $g = 2$. We consider the maximum number of children constraint as $c_{\max} = 3$. The following statements are valid for this specific problem:

$$\begin{array}{ll}
\mathrm{s}(1,k) = \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1\} & \mathrm{s}(2,k) = \{s_1^2, s_2^2, s_3^2\} \\
n(1,k) = 5 & n(2,k) = 3 \\
\mathrm{o}(1,k) = \{o_1^1, o_2^1\} & \mathrm{o}(2,k) = \{o_1^2, o_2^2, o_3^2\} \\
\mathrm{S}_k = \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2\} & \mathrm{O}_k = \{o_1^1, o_2^1, o_1^2, o_2^2, o_3^2\} \\
\mathrm{I}_k = \{s_1^1, s_4^1, s_1^2\} & \mathrm{D}_1 = 1,\ \mathrm{D}_2 = 2,\ \mathrm{D}_3 = 2.
\end{array}$$

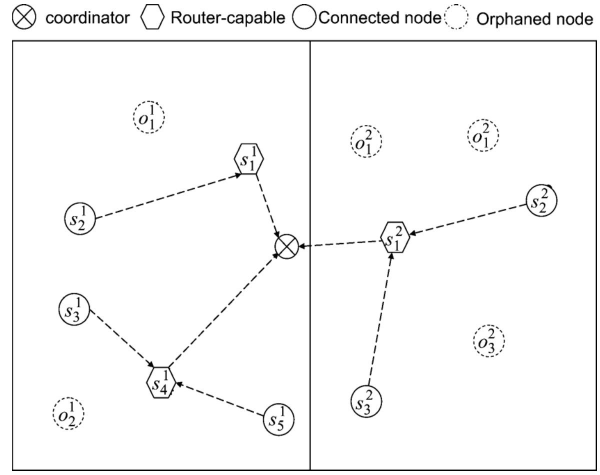


Fig. 5. A snapshot of a network connection.

The initial solution to this problem will be $\mathrm{sol}_0 = \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, \{\emptyset\}\}$. It can be seen in Fig. 5 that $c_{\max} = 3$, thus the router-capable nodes s_1^1, s_4^1 and s_1^2 have 2, 1, and 1 number of unassigned children slots respectively. The main steps involved in RBC at $\tau = 1$ are described as follows.

- Step 1) Calculate $\gamma_1(o_q, s_p^i)$ and $\gamma_1(o_q, c_d^m)$.
- Step 2) At $\tau = 1$, generate $\Lambda - \mathrm{M}_k = (9)$ number of offspring from the initial solution $\mathrm{sol}_0 = \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, \{\emptyset\}\}$. Let us assume them to be

$$\begin{aligned}
\mathrm{Temp}_1 &= \{s_1^1, o_2^2, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, \{\emptyset\}\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, s_2^1))) = o_2^2 \\
\mathrm{Temp}_2 &= \{s_1^1, s_2^1, o_1^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, \{\emptyset\}\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, s_3^1))) = o_1^1 \\
\mathrm{Temp}_3 &= \{s_1^1, s_2^1, s_3^1, s_4^1, o_1^2, s_1^2, s_2^2, s_3^2, \{\emptyset\}\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, s_2^1))) = o_1^2 \\
\mathrm{Temp}_4 &= \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, o_1^1, s_3^2, \{\emptyset\}\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, s_2^2))) = o_1^1 \\
\mathrm{Temp}_5 &= \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, o_3^2, \{\emptyset\}\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, s_3^2))) = o_3^2 \\
\mathrm{Temp}_6 &= \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, o_2^2\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, c_2^1))) = o_2^2 \\
\mathrm{Temp}_7 &= \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, o_2^1\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, c_3^1))) = o_2^1 \\
\mathrm{Temp}_8 &= \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, o_1^2\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, c_3^2))) = o_1^2 \\
\mathrm{Temp}_9 &= \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, o_2^1\}, \text{ where } \\
&\quad \Pi(cdf(\gamma_1(o_q, c_3^3))) = o_2^1.
\end{aligned}$$

- Step 3) Calculate each offspring fitness value using (28) and select the offspring that has the lowest value out of Λ mutated offsprings. Let $\mathrm{Temp}_8 = \arg\min_{\forall k=1,\ldots,9}\{\mathrm{J}(\mathrm{Temp}_k)\}$.
- Step 4) Replace the current parent solution by the selected offspring in Step 3), if its fitness value is less than that of parent. Let $\mathrm{J}(\mathrm{Temp}_8) < \mathrm{J}(\mathrm{sol}_0)$ then, $\mathrm{sol}_1 = \{s_1^1, s_2^1, s_3^1, s_4^1, s_5^1, s_1^2, s_2^2, s_3^2, o_1^2\}$.

- Step 5) Update $\mathrm{S}_k(1), \mathrm{O}_k(1)$ and the number of unallocated children available in each router-capable node.
- Step 6) If τ equals MAXITER then go to Step 7), else increment τ by one and go to Step 2).
- Step 7): The suboptimal solution to our problem is given by sol_τ.

At each iteration, sol_τ is replaced if the selected offspring is better than the parent solution. Therefore, the solution gets refined as the algorithm proceeds. At each iteration, RBC requires $\Lambda - \mathrm{M}_k$ number of fitness evaluations, and therefore, a total of $\Lambda - \mathrm{M}_k \times$ MAXITER number of fitness evaluations are required. The main reason for opting RBC for this problem is that the constraints can be easily taken into account since RBC produces offspring by mutating each single dimension separately. As we can see, we need to consider only the energy constraint not the constraint on the maximum number of children during the mutation. Furthermore, RBC's computational time is controllable by using the termination condition.

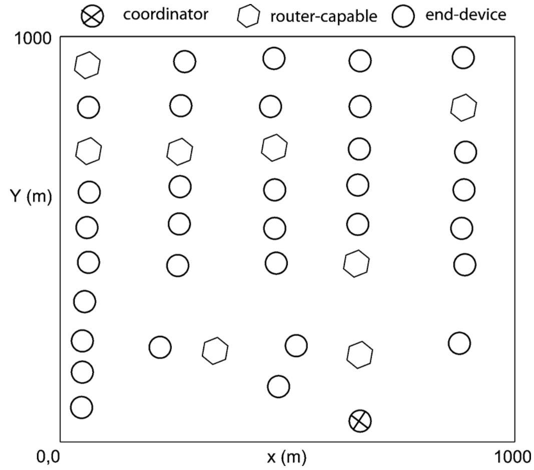


Fig. 6. Characteristics and locations of sensor nodes deployed in the simulated network.

V. Results and Discussion

In this section, we present a numerical problem to demonstrate the orphaned node management in a wireless sensor network. We consider a simulated sensor network of 40 sensor nodes deployed in a vineyard farm of 1000 m $\times$ 1000 m, where sensor nodes continuously produce measurements of atmospheric temperature and soil moisture levels to the base station (or coordinator) by multi-hop communication. The schematic diagram of the simulated sensor network is illustrated in Fig. 6, where 9 nodes are router-capable nodes. We assume that the each sensor node is within the transmission range of other nodes. For clarity, only a single moisture level is observed by the sensor nodes. Thus, $\mathrm{X}(a_i, k) = [t(a_i, k)\ m(a_{i_1}, k)]$ is a 2×1 column vector. Even though our proposed algorithms can be applied on any linear/nonlinear Gaussian models, the following models are considered only for this simulation purpose.

$$\mathrm{X}(a_i, k) = \begin{bmatrix} t(a_i, k-1) + 0.5\sqrt{|t(\theta_1, k-1) - 80|} \\ 0.7m(a_{i_1}, k-1) + 0.7\sqrt{m(a_{i_1}, k-1)} \end{bmatrix} + w(a_i, k-1) \quad (34)$$

$$z_{s_p^i}(a_i, k) = \begin{bmatrix} \sqrt{t(a_i, k)} \\ \sqrt{m(a_{i_1}, k)} \end{bmatrix} + v_{s_p^i}(a_i, k). \quad (35)$$

If the noise matrices $w(a_i, k-1)$ and $v_{s_p^i}(a_i, k)$ are ignored, for given $t(a_i, 0) > 0$ and $m(a_{i_1}, 0) > 0$, (34) is nonlinearly increasing while (35) is nonlinearly decreasing. In this simulation, we use the error covariances of the 40 sensor nodes are different from each other and not listed due to page limitations. We use a measurement data size of 1 MB and a data rate $r = 8$ Mb/s. The energy model parameters in $\psi_{tx}(s_p^i, k)$ are set $\alpha_1 = 100\ \frac{nJ}{b}$ and $\alpha_2 = 1\ \frac{pJ}{(bm^2)}$. For simulation purposes, we consider only the energy consumed by transmission, since $\psi_{tx}(s_p^i, k) \gg \psi_1$. Measurements are collected at the base station at every 30 minutes for a total of 10 hours. Therefore, the base station has to manage the orphaned nodes for the time steps $k = 1, 2, \ldots, 20$. The data latency is neglected for the purpose of simulation, thus if a sensor node connected to the base station, the measurement will be transmitted immediately after the observed measurement is collected.

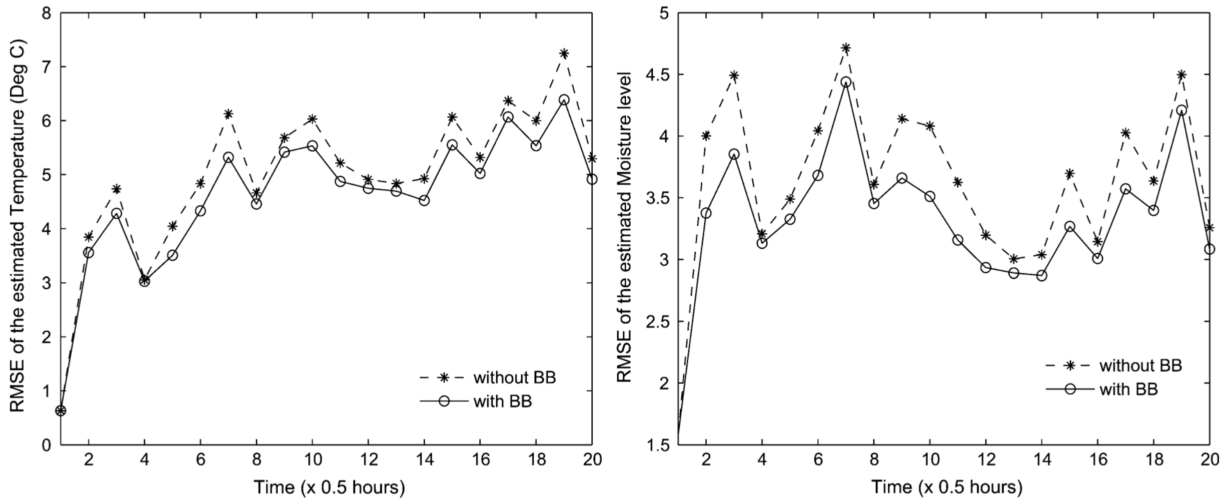
For the simulation purpose, we set the initial state $\mathrm{X}(a_i, 0)$ and the initial error covariance $\mathrm{P}(a_i, 0 \mid 0)\ \forall\ a_i \in \theta$ as $\mathrm{X}(a_i, 0) = \begin{bmatrix} 30 \\ 60 \end{bmatrix}, \mathrm{P}(\theta, 0|0) = \begin{bmatrix} 0.4 & 0.0 \\ 0.0 & 2.5 \end{bmatrix}$.

First, we present the results for a single area problem. We assume that the state of temperature and soil moisture level vary only with time, thus $g = 1$ for this simulation. The process error covariance is assumed to be Gaussian and given by $\mathrm{Q}(a_1, k) = \begin{bmatrix} 50.0 & 0.0 \\ 0.0 & 20.0 \end{bmatrix}$.

Fig. 7 illustrates the variation of RMSE of the estimated state of phenomena with the energy constraint $\mathrm{E}_k = 30$ JJ and $c_{\max} = 6$ for the time steps $k = 1, 2, \ldots, 20$. As we can see in Fig. 7, the orphaned node management brings down the RMSE (root mean square error) of the estimated states of the phenomenon. Furthermore, the energy constraint imposed on the network produces significant changes in restoring the orphaned nodes. As we reduced the energy budget from 30 J to 20 J for a given network setup, it can be seen in Fig. 8 that the RMSE of the estimated state is larger than in the previous case.

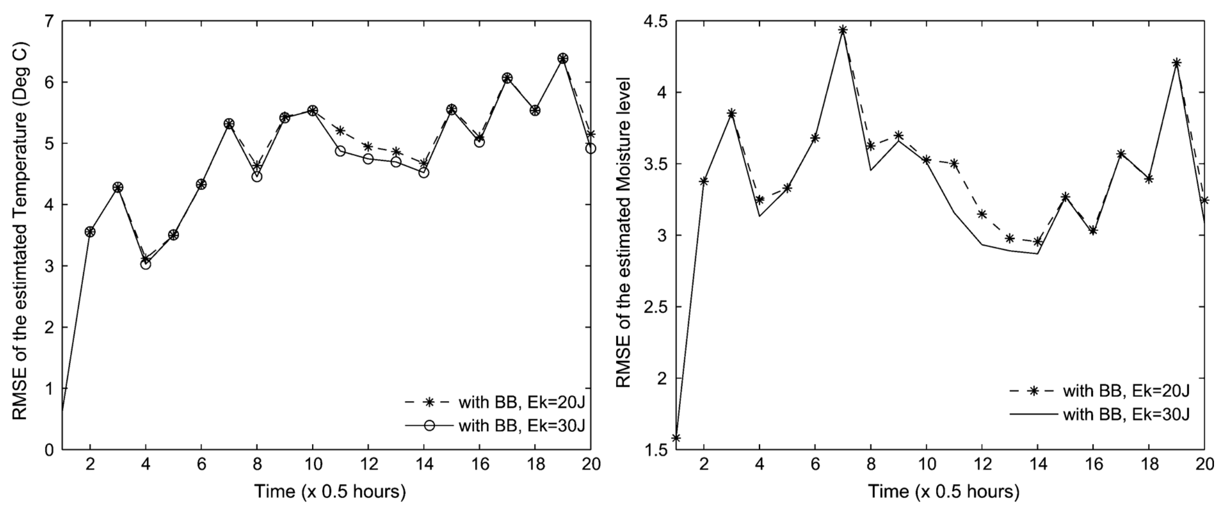
Even though the branch-and-bound method is optimal for the linearized single-area problem, it may not produce the optimal solution within an acceptable time period when either the constraints are very tight or the approximation between the original problem and the linearized problem is not within an acceptable limit. If a feasible solution is not found, it can keep branching resulting in a huge computational cost. Therefore, we also apply RBC to a single-area problem to obtain a solution at a lower computational cost. We analyze the performance of the algorithms for several energy budgets for two different $c_{\max}$ as presented in Tables I and II. We used MAXITER = 50 for the RBC technique.

The results shown in Tables I and II were obtained by executing each algorithm for 5 minutes. As we can see, branch-and-bound algorithm produce the best solutions for most cases. However, it was unable to find a better solution within the allocated time for some cases, as shown in Tables I and II. In our case, the measurements are collected every 30 minutes, therefore, we can allocate more time to the algorithms' execution time. We observed that as we increase the algorithms' execution



(a) (b)

Fig. 7. Variation of the RMSE of the estimated state of phenomena with and without performing orphaned node management using branch-and-bound algorithm for $E_k = 30$ J: (a) RMSE of the estimated temperature and (b) RMSE of the estimated soil moisture level.



(a) (b)

Fig. 8. Variation of the RMSE of the estimated States of the phenomenon with performing orphaned node management using branch-and-bound algorithm for $E_k = 30$ J and $E_k = 20$ J: (a) RMSE of the estimated temperature and (b) RMSE of the estimated soil moisture level.

time, branch-and-bound provides the optimal solution. However, branch-and-bound may not be suitable when the time period between two consecutive collected measurements is small. On the other hand, RBC always produce a solution even the allocated time period is small or the constraints are very tight as opposed to branch and bound method. The results presented in Tables I and II are the RMSE of the estimated states only for a single time step and are small in terms of numerical value. However, interpretation of the cumulative error in terms of water cost and damage to the crops may be significant as explained in Section III. As it can be seen in Table II that the restoring method did not produce significant reduction in the RMSE of the estimated states. This is due to the number of available children slots for restoring orphaned nodes is low as the result of reduction in maximum number of children constraints for the router-capable nodes.

Second, we present the results for the multi-area problem by using the RBC technique. The farm area is divided into 4 subgroups denoted by $\theta = \{a_1, a_2, a_3, a_4\}$ as shown in Fig. 9. The error covariance of the process noise, $Q(\theta, k)$, considered for this simulation is as follows:

$$Q(a_1,k) = \begin{bmatrix} 50.0 & 0.0 \\ 0.0 & 15.0 \end{bmatrix} \; Q(a_2,k) = \begin{bmatrix} 25.0 & 0.0 \\ 0.0 & 20.0 \end{bmatrix}$$
$$Q(a_3,k) = \begin{bmatrix} 30.0 & 0.0 \\ 0.0 & 10.0 \end{bmatrix} \; Q(a_4,k) = \begin{bmatrix} 50.0 & 0.0 \\ 0.0 & 20.0 \end{bmatrix}$$

Fig. 10(a) shows the variation of sum of RMSE of the estimated states with and without the orphaned node management. As expected, orphaned node management using the RBC technique produces improvement in estimation. We run RBC for two different energy budgets with the same network setup. Fig. 10(b) illustrates the cumulative RMSE of the estimated states for two different energy budgets. Since RBC is a stochastic optimization technique, the presented results are the mean value of 30 independent runs. We set MAXITER=50 for RBC technique for

TABLE I
RMSE OF THE ESTIMATED STATE OF PHENOMENA OBTAINED BY BRANCH-AND-BOUND AND RANDOM-BIT-CLIMBER FOR DIFFERENT ENERGY BUDGETS AND NETWORK TOPOLOGIES WITH $c_{\max} = 6$

No. of orphaned nodes ($\lvert \mathrm{O}_k \rvert$)	Energy budget (E_k)	without restoring technique	branch-and-bound method	random-bit-climber method
3	26.0	3.7441	3.6375	3.6750 ± 0.036
3	27.0	3.7441	3.6197	3.6345 ± 0.012
4	33.0	3.8100	3.6617	3.6781 ± 0.024
10	18.0	4.0564	3.7654	3.7721 ± 0.011
13	19.0	4.1908	3.8272	3.8301 ± 0.024
14	19.0	4.1860	3.9281	3.8420 ± 0.035
14	20.0	4.1860	3.7430	3.7525 ± 0.005
23	10.0	4.8840	4.3236	4.3940 ± 0.028
23	13.0	4.8840	4.3236	4.3300 ± 0.001
29	5.0	5.5460	5.5460	5.1358 ± 0.061
29	7.0	5.5460	5.5460	4.9342 ± 0.029
29	10.0	5.5460	4.9020	4.8890 ± 0.018

TABLE II
RMSE OF THE ESTIMATED STATE OF PHENOMENA OBTAINED BY BRANCH-AND-BOUND AND RANDOM-BIT-CLIMBER FOR DIFFERENT ENERGY BUDGETS AND NETWORK TOPOLOGIES WITH $c_{\max} = 4$

No. of orphaned nodes ($\lvert \mathrm{O}_k \rvert$)	Energy budget (E_k)	without restoring technique	branch-and-bound method	random-bit-climber method
7	33.0	3.8713	3.7969	3.812 ± 0.029
9	21.0	3.9601	3.7969	3.8345 ± 0.011
13	19.0	4.1535	4.1002	3.9801 ± 0.014
13	21.0	4.1535	3.9564	3.9531 ± 0.004
19	16.0	4.6096	4.3021	4.3120 ± 0.001
19	18.0	4.6096	4.2753	4.2923 ± 0.041
21	18.0	4.6504	4.3384	3.4331 ± 0.016
35	16.0	6.4399	6.4399	6.2841 ± 0.055
35	18.0	6.4399	6.1879	6.1987 ± 0.005

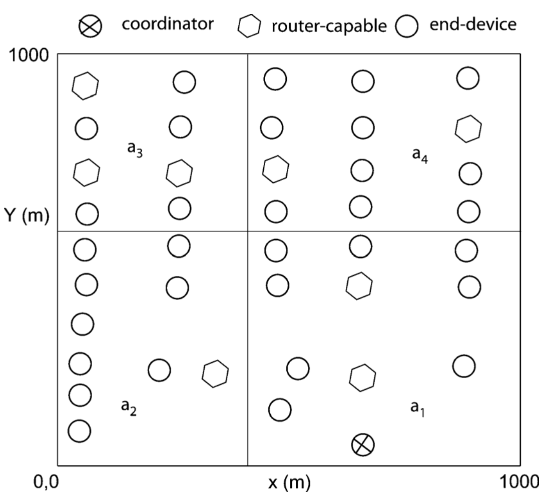


Fig. 9. Characteristics and locations of sensor nodes deployed in the network where the whole area consists of multiple heterogeneous farming areas.

each run. The advantage of using RBC technique is that it easily handles both linear and nonlinear objective functions with constraints. It is also possible that we can give priority to a particular area over the rest by introducing weighting factors as in (28). Moreover, we reduce the maximum number of children constraints $c_{\max}$ from 6 to 3 and Fig. 11 illustrates the cumulative sum of RMSE of the states with and without using RBC for $c_{\max} = 6$ and 3. We observed that number of orphaned nodes present in the network with $c_{\max} = 3$ is higher than that of with $c_{\max} = 6$. Since $c_{\max}$ is reduced, number of available children slots for restoring orphaned nodes also reduced. Due to large number of orphaned node presence and less number of available children slots, the RMSE of the estimated states with $c_{\max} = 3$ is higher than that of with $c_{\max} = 6$ as it can be seen in Fig. 11.

VI. CONCLUSION

In this paper, we analyzed the problem of orphaned node management in a wireless sensor network, which is deployed in a harsh environment such as an agricultural farm area for automated irrigation scheduling and control. We assumed that the measurement from each sensor node consists of temperature and soil moisture level at its location with white Gaussian noise. Unlike the exhausted (or dead) nodes, the orphaned-nodes can still receive and transmit messages and thus it is possible to restore them to the network. We analyzed the problem for two different cases, where the farm area is considered as single and multiple. We assumed that the total cost due to false triggering of irrigation increases as the total estimation error increases. We demonstrated that restoring the orphaned nodes improves the estimation of necessary field parameters, ultimately minimizing the costs involved in irrigation. We linearized the problem of single homogeneous farming area and optimally solved by using branch-and-bound. Although branch-and-bound method is optimal for the linearized problem, it may be suboptimal to the original problem due to the approximation involved in linearizing the original objective function. On the other hand, the random-bit-climbing technique is a suboptimal method and it has the flexibility of handling linear and nonlinear objective functions. In this work, we used the state of phenomena and the sensor measurement models as nonlinear Gaussian systems, and used extended Kalman filtering technique for the estimation. It is possible that the noises considered in the models

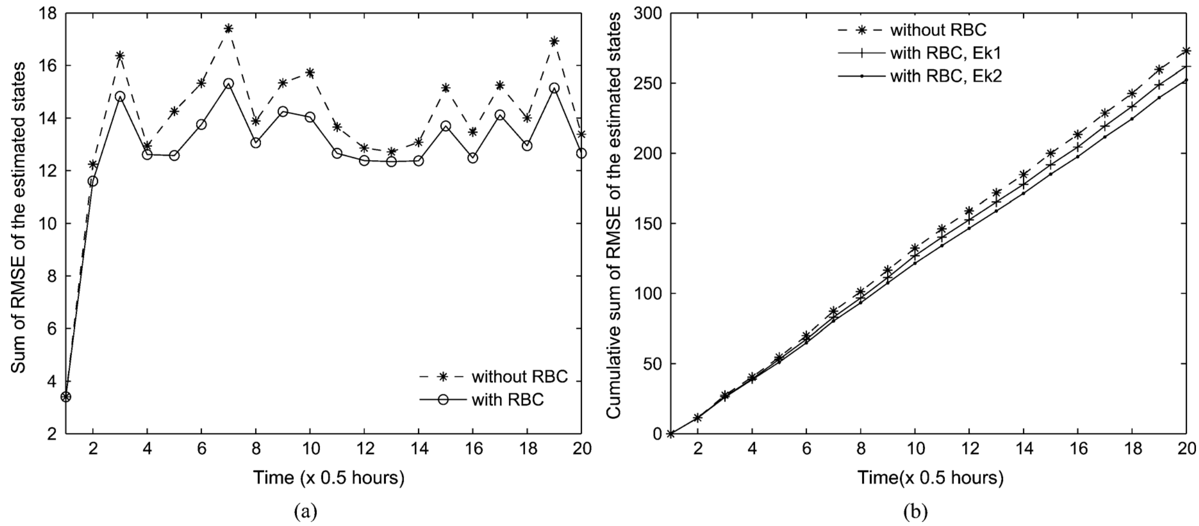


Fig. 10. (a) Sum of RMSE of the estimated states with and without using RBC (b) Variation of the cumulative sum of RMSE of the estimated states for different E_k.

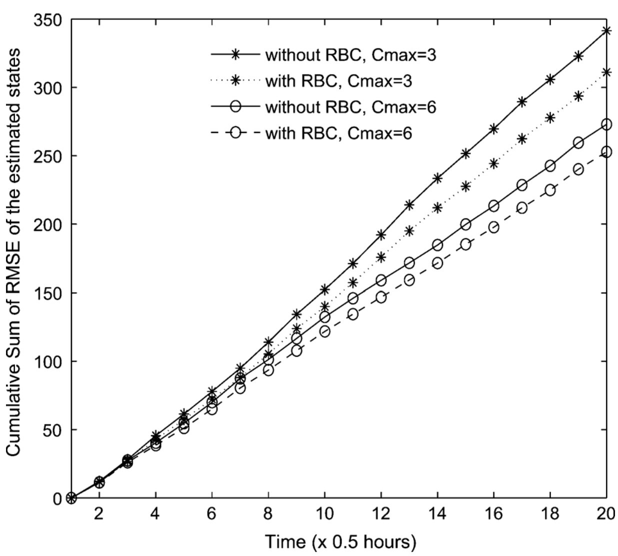


Fig. 11. Variation of the cumulative sum of RMSE of the estimated states with and without using RBC for $c_{\max} = 3$ and $c_{\max} = 6$.

may not be Gaussian in some cases and the formulation of the objective function should be redefined. Our future research will focus on developing and refining the algorithms to handle nonlinear non-Gaussian models.

APPENDIX

As defined in Section III, $\mathrm{S}_k = \{s(1,k), s(2,k), \ldots, s(g,k)\}$, $\mathrm{O}_k = \{o(1,k), o(2,k), \ldots, o(g,k)\}$ and $\mathrm{I}_k = \{r_1, r_2, \ \ldots, r_{\mathrm{M}_k}\}$ denote the set of connected sensor nodes, orphaned nodes and connected router-capable nodes in the network respectively at time instant k. The objective is to minimize the total estimated error by restoring the orphan nodes to the network, subject to satisfying (12) and (13). The maximum number of solutions to this problem is calculated with the following concerns:

- relaxing the energy constraint;
- assuming that the solution when connecting o_q at the dth unassigned child slot of the mth router-capable node is different from the solution when connecting at any other unassigned children slots of the mth router-capable node.

The maximum number of orphan nodes which can be restored with the network is $\min\{|\mathrm{O}_k|, f\}$, where $f = |\mathrm{S}_k| - \mathrm{M}_k + \sum_{m=1}^{\mathrm{M}_k}(c_{\max} - \mathrm{D}_m)$. If we select u number of orphan nodes to restore to the network, the number of possible permutations and combinations can be calculated as: (select any u number of orphan nodes from $|\mathrm{O}_k|$ number of available orphan nodes) and (restore them at any u number of possible places available, where $u \leq f) = ({}^{|\mathrm{O}_k|}\mathrm{C}_u) * ({}^{f}\mathrm{P}_u)$.

We can select 0, 1,. . . or $\min\{|\mathrm{O}_k|, f\}$ number of orphan nodes to restore to the network, thus the maximum number of possible solutions to this problem is given as

$$\sum_{u=0}^{\min\{|\mathrm{O}_k|, f\}} {}^{|\mathrm{O}_k|}\mathrm{C}_u * {}^{f}\,\mathrm{P}_u.$$

ACKNOWLEDGMENT

The authors greatly acknowledge the Farms Rivers and Markets project of the University of Melbourne and its funders for providing funds to employ the first author that enabled us complete this paper. The authors also wish to record their gratitude to K. Steer of Department of Mechanical Engineering, University of Melbourne, Dr. K. Saleem, and Dr. M. Pham of the Victorian Laboratory of National Information and Communication Australia (NICTA) for their comments and supports and G08/DAAD grant on Wireless Sensor Networks for its support.

REFERENCES

[1] P. Bauer, M. Sichitiu, R. Istepanian, and K. Premaratne, “The mobile patient: Wireless distributed sensor networks for patient monitoring and care,” in *Proc. IEEE EMBS Int. Conf. Inf. Technol. Appl. Biomed.*, 2000, pp. 17–21.

[2] S. Ooi, I. Mareels, N. Cooley, G. Dunn, and G. Thomas, “A systems engineering approach to viticulture on-farm irrigation,” in *Proc. 17th IFAC World Congr.*, 2008, pp. 9569–9574.

[3] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless sensor networks for habitat monitoring,” in *Proc. 1st ACM Int. Workshop Wireless Sensor Netw. Appl.*, 2002, pp. 88–97.

[4] J. Yick, B. Mukherjee, and D. Ghosal, “Wireless sensor network survey,” *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, 2008.

[5] M.-S. Pan, C.-H. Tsai, and Y.-C. Tseng, "The orphan problem in Zigbee wireless networks," *IEEE Trans. Mobile Comput.*, vol. 8, pp. 1573–1584, Nov. 2009.

[6] M. Herlihy, N. Lynch, M. Merritt, and W. Weihl, "On the correctness of orphan management algorithms," *J. ACM*, vol. 39, pp. 881–930, Oct. 1992.

[7] F. A. Silva, L. B. Ruiz, T. R. M. Braga, J. M. S. Nogueira, and A. A. F. Loureiro, "Designing a self-organizing wireless sensor network," in *Proc. MATA*, 2004, pp. 186–195.

[8] A.-J. Garcia-Sanchez, F. Garcia-Sanchez, and J. Garcia-Haro, "Enhancements in the orphan process for wireless personal area networks: Real implementation scenarios," in *Proc. 4th Int. Conf. Softw. Eng. Adv. (ICSEA)*, Sep. 2009, pp. 371–377.

[9] Y. He, H. Liu, and F. Zhu, "Improved schemes for node clustering in decentralized peer-to-peer networks," in *Proc. 3rd Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Dec. 2007, pp. 461–467.

[10] E. Curley, B. Ravindran, J. Anderson, and E. D. Jensen, "Recovering from distributable thread failures in distributed real-time Java," *ACM Trans. Embed. Comput. Syst.*, vol. 10, pp. 8:1–8:37, Aug. 2010.

[11] K.-I. Hwang, B.-J. Choi, and S.-H. Kang, "Enhanced self-configuration scheme for a robust Zigbee-based home automation," *IEEE Trans. Consum. Electron.*, vol. 56, pp. 583–590, May 2010.

[12] C. D'Amour, R. Life, G. Elmasry, and B. Welsh, "Determining network topology using governing dynamics based on nodal and network requirements," in *Proc. IEEE Conf. Military Commun. Conf. (IEEE MILCOM)*, 2004, vol. 3, pp. 1579–1583.

[13] Y. Li, S. Verma, L. Lao, and J.-H. Cui, "SACA: SCM-based adaptive clustering algorithm," in *Proc. 13th IEEE Int. Symp. Modeling, Anal., Simulat. Comput., Telecommun. Syst.*, Sep. 2005, pp. 271–279.

[14] S. B. Attia, A. Cunha, A. Koubaa, and M. Alves, "Fault-tolerance mechanisms for Zigbee wireless sensor networks," in *Proc. 19th Euromicro Conf. Real-Time Systems (ECRTS)*, 2007 [Online]. Available: http://www.open-zb.net/publications/ECRTS07_WiP_camera%20ready.pdf

[15] J. A. Benth, F. E. Benth, and P. Jalinskas, "A spatial-temporal model for temperature with seasonal variance," *J. Appl. Stat.*, vol. 34, pp. 823–841, Sep. 2007.

[16] T. L. Yee, V. Sehgal, and H. S. Hamid, Sensoclean: Handling noisy and incomplete data in sensor networks using modeling Univ. of Maryland, College Park, 2005.

[17] T. Chung, V. Gupta, B. Hassibi, J. Burdick, and R. Murray, "Scheduling for distributed sensor networks with single sensor measurement per time step," in *Proc. IEEE Conf. Robot. Automation*, Apr. 2004, vol. 1, pp. 187–192.

[18] R. Tharmarasa, T. Kirubarajan, M. Hernandez, and A. Sinha, "Pcrlb-based multisensor array management for multitarget tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 2, pp. 539–555, 2007.

[19] S. Maheswararajah, S. Halgamuge, and M. Premaratne, "Sensor scheduling for target tracking by suboptimal algorithms," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 1467–1479, Mar. 2009.

[20] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. New York: Wiley, 2001.

[21] E. Song, Y. Zhu, and J. Zhou, "The optimality of Kalman filtering fusion with cross-correlated sensor noises," in *Proc. CDC. 43rd IEEE Conf. Dec. Control*, Dec. 2004, vol. 5, pp. 4637–4642.

[22] G. A. F. Seber, *A Matrix Handbook for Statisticians*, 1st ed. New York: Wiley-Interscience, 2007.

[23] L. Davis, "Bit-climbing, representational bias, and test suite design," in *Proc. 4th Int. Conf. Genetic Algorithms*, Los Altos, 1991, pp. 18–23.

Suhinthan Maheswararajah received the B.Sc. degree in electrical engineering from the University of Moratuwa, Moratuwa, Sri Lanka, in 2004 and the Ph.D. degree from the University of Melbourne, Australia, in 2009.

He is currently a Research Scientist in the Melbourne School of Land Environment, the University of Melbourne, Australia. His research interests include state estimation algorithms, optimization, and sensor management applications in wireless sensor networks.

Saman K. Halgamuge (M'85) received the B.Sc.Eng. degree from the University of Moratuwa, Sri Lanka, in 1984 and the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technical University of Darmstadt, Germany, in 1990 and 1995, respectively.

He is a Professor in the Department of Mechanical Engineering, Melbourne School of Engineering, The University of Melbourne. He is also working within Sustainable Systems and Energy and Biomedical Engineering themes of the Melbourne Engineering Research Institute. He leads a group of postdoctoral fellows and Ph.D. students working on pattern recognition and optimization with applications in the broad areas of bioengineering and mechatronics. He is the coauthor of over 60 journal papers, 150 conference papers (h-index 21), 8 books, and 15 book chapters. His web page is at http://www.mech.unimelb.edu.au/people/staff/saman_halgamuge.html

Kithsiri B. Dassanayake received the B.Sc. (Agric) degree in soil and water engineering and the Ph.D. degree in ecological physiology from the University of Peradeniya, Sri Lanka, in 1986 and 1997, respectively.

He is a soil and water scientist. He worked at the Coconut Research Institute of Sri Lanka as a Senior Scientist for 15 years before migrating to Australia in mid-2001. He is currently a Research Fellow attached to the Melbourne School of Land Environment, the University of Melbourne. His expertise includes landscape hydrology and agricultural water management, micrometeorology, farming systems modeling and climate change impacts, and adaptation in agriculture. He successfully delivered the outcomes of the Smarter Irrigation Systems research project under the Science and Technology Innovation Initiative program funded by the Victorian State Government. This project successfully demonstrated the economic and environmental benefits of wireless sensor based automation of irrigation scheduling and operations. Currently, as a key member of Farms & Rivers and Markets (FRM) project, he currently leads four subprojects. He modeled impacts of future climates on various farming systems in Northern Victoria and designed and implemented large-scale field trials under the FRM project.

David Chapman received the Ph.D. degree on white clover physiology at Grassland Research Institute and the University of Reading, U.K., in 1989.

He joined DSIR Grasslands (Later AgResearch Grasslands) as a Scientist in 1979, working initially on the ecology of white clover in hill pastures and subsequently in plant and pasture improvement strategies for hill country. He was leader of the AgResearch Grasslands Forage Ecology and Management group from 1993 to 1996. He is currently the Principal Scientist, Feed and Farm Systems, DairyNZ, Lincoln. He was Foundation Chair of Pasture Science in the Faculty of Land and Food Resources, University of Melbourne, from April 1996 to April 2010. He led four major research projects while in Victoria: Sustainable Grazing Systems, western Victoria national experiment site (Meat & Livestock Australia funding)—grazing management for perennial pasture productivity and increased water use in the high-rainfall zone; development of "DairyMod" (Dairy Australia funding)—biophysical model of the dairy grazing system; pasture persistence (Dairy Australia and ARC funding)—more persistent and productive pastures for nonirrigated southern Australia dairy regions; and Project 3030 (Dairy Australia, DPI-Victoria, University of Melbourne funding)—lifting home-grown forage consumption and profit by 30% in nonirrigated dairy systems in southern Australia.