Diffusion-based approach to deploying wireless sensor networks

R. Machado*

Department of Computer Science,
New Jersey Institute of Technology,
Newark, NJ 07102, USA
Email: rmm23@njit.edu
*Corresponding author

S. Tekinay

Department of Electrical and Electronics Engineering,
Ozyegin University,
Istanbul, Turkey
Email: sirin.tekinay@ozyegin.edu.tr

W. Zhang

Department of Computer Science,
Iowa State University,
Ames, IA 50011, USA
Email: wzhang@cs.ia.edu

G. Wang

Department of Computer Science,
New Jersey Institute of Technology,
Newark, NJ 07102, USA
Email: gwang@njit.edu

Abstract: An important objective of Wireless Sensor Networks (WSNs) is to reliably sense data about the environment in which they are deployed. Reliability in WSNs has been widely studied in terms of providing reliable routing protocols for message dissemination and reliability of communication from sink to sensors. In this work, we define a reliability metric by the amount of data sensed by the network. In order to satisfy this reliability constraint, we propose a diffusion-based approach for a deployment pattern for the sensor nodes. We show that this deployment pattern achieves sufficient coverage and connectivity and requires lesser number of sensors than popular regular deployment patterns. We further obtain the bounds on establishing connectivity between nodes in the WSN and extend this analysis for heterogeneous WSNs.

Keywords: deployment; diffusion; connectivity; WSNs; wireless sensor networks.

Reference to this paper should be made as follows: Machado, R., Tekinay, S., Zhang, W. and Wang, G. (2010) 'Diffusion-based approach to deploying wireless sensor networks', *Int. J. Sensor Networks*, Vol. 8, Nos. 3/4, pp.222–232.

Biographical notes: R. Machado received her PhD in Electrical Engineering from New Jersey Insitute of Technology in 2009. Her research interests are in the areas of wireless network topologies, wireless sensor networks and wireless communications.

S. Tekinay is the Vice-Rector for Research and Technology at Ozyegin University Turkey. She is also the Associate Professor of Electrical Engineering at New Jersey Institute of Technology and has served as the Program Director at the NSF. She received her PhD from George Mason University. Her research interests are in network science, complex dynamic systems and networks.

W. Zhang is an Assistant Professor of Computer Science at Iowa State University. He received his PhD in Computer Science from The Pennsylvania State University. His research interests are wireless networks design, integration and applications; applied cryptography and distributed system privacy/security.

G. Wang is an Assistant Professor of Computer Science at New Jersey Institute of Technology. She received her PhD in Computer Science from The Pennsylvania State University. Her research interests are network and systems security, wireless ad-hoc/sensor networks, vehicular networks, mobile computing.

1 Introduction

The placement of wireless sensor nodes in a deployment region resulting from various deployment strategies impacts the coverage in WSNs. The primary objective of deployment in a WSN is coverage of the entire sensing region for the duration of deployment. However, since sensor nodes are prone to failure due to running out of battery energy or failure due to device unreliability, it is desirable to have every point of the deployment region covered by more than one sensor at all times. The problem of coverage in the deployment region has led to research in both deterministic and random deployment patterns. While randomly scattering nodes in a deployment region is suitable for remote regions and eliminates the need for planning the locations of sensor nodes, sensor nodes deployed deterministically optimise available resources in terms of number of nodes and achieve other objectives such as rapid convergence of localisation algorithms and efficient communication between nodes and sink.

In this paper, we study the following fundamental deployment problem: Given that we have N homogenous sensor nodes each with a certain sensing radius R_s that cover a deployment region of area A, what is a deployment pattern such that we can achieve (a) k-redundancy (where k-redundancy refers to each point in the deployment region being covered by $k > 1$ sensors) in the deployment region and (b) reliable network operation for at least T time units? This paper addresses the problem of deployment of sensors from the point of view of data obtained from the sensing operation. We define a measure of reliability of the sensing operation by the total amount of data obtained by the Wireless Sensor Network (WSNs) at the end of the minimum time T for which it was deployed to be operational. Assuming that a sensor obtains an amount of data I bytes in unit time, a WSN of y number of nodes in the network yields yIT bytes of data at the end of T time units.

We propose a diffusion-based deployment pattern that achieves this measure of reliability by modelling the total data obtained as a hypothetical, non-stationary, singular source that diffuses through the deployment region. At every time instant when a node is deployed, the data source is assumed to have diffused by an amount equal to the data sensed by a single sensor for duration T.

At the end of the diffusion process, the deployment is complete and all the sensors have been deployed across the region with k-redundancy. We show that the resultant deployment pattern is advantageous in terms of use of resources when compared to current regular deployment patterns. Its advantage of requiring fewer number of sensors holds for any desired level of redundancy in the network, given that the number of nodes that we have to start with is at least A/A_s, where A_s is the area sensed by a single node, and A is the area of the entire deployment region. The following list is a summary of differences between our proposed diffusion-based deployment pattern and regular deployment patterns:

- We describe a novel diffusion-based pattern for node deployment in WSNs. The novelty of this pattern lies in its potential for use in applications where deterministic or random deployment of nodes is not feasible. An example would be the deployment of a sensor network for monitoring toxic chemical leakages in a wide area, or monitoring the spread of an epidemic in a large geographical region. Our proposed diffusion-based deployment pattern assumes that all nodes are initially located at a single point, and are then diffused (placed at locations) throughout the sensing region incrementally. This diffusion may be performed by a mobile robot for unmanned applications. We design the diffusion-based deployment algorithm such that the distribution of the nodes at the end of the deployment is uniform.
- The diffusion-based deployment pattern is designed to provide network operation for at least a pre-specified time T. This is in contrast to regular deployment patterns, where the topology of the network is determined by application-specific parameters or is random in random deployments.

We also study the entropy of establishing connectivity, E_c, between nodes in a WSN. Establishing connectivity between nodes in WSN is key to ensuring reliable network operation. Connectivity guarantees that links in the WSN exist for timely transfer of data from nodes to the base station and for dissemination of queries and broadcast messages from the base station to nodes in the network. With the help of an analytical framework, we show the dependence of E_c on the densities of nodes and the role of the fading environment on E_c. We illustrate this framework by presenting bounds on E_c for a two-tier distribution of computationally powerful nodes and regular wireless sensor nodes in the network. These results can be used to design WSNs for robust network operation even in the case of sparse distribution of computationally powerful nodes in the network and thus highlight the importance of optimal network design in WSNs.

The following are the contributions of this work:

- We design a diffusion-based deployment pattern that is guaranteed to ensure network operation for at least time *T*. This constraint is suitable for WSN applications, where the utility is fixed and application-specific (Ahn and Krishnamachari, 2006). An example of this would be a WSN deployed to determine the presence of intruders in a fixed time interval, after which the utility of the application is unaffected by the presence of intruders. We show that our diffusion-based deployment pattern requires lesser number of sensors than regular deterministic deployment patterns.
- We establish the bounds on the entropy to establish connectivity between nodes in a WSN. These bounds can be easily extended to the case of heterogeneous networks, where one tier of nodes is more powerful than the next lower tier of nodes. The knowledge of the entropy to establish connectivity is important in the design of such heterogeneous networks, since the density of nodes in each tier determines cluster sizes. For the simple case of a hierarchical two-tier network design of powerful and regular nodes, the cluster size affects network performance parameters such as data aggregation efficiency and latency of data transfer between nodes and base station. Knowing the entropy of establishing connectivity can help us obtain optimal cluster sizes for efficient routing and data aggregation.
- The analytical framework of our diffusion-based deployment pattern can be readily modified to find the entropy of establishing connectivity in random as well as deterministic deployment patterns. This can be done by obtaining the expected distance between locations of nodes and using them in the equations to find the entropy.

The rest of the paper is organised as follows. Section 2 presents related work in placement models for WSNs. Section 3 describes the diffusion-based deployment approach, presents the diffusion-based deployment algorithm for placement of nodes by overlay of planar graphs, and shows that the deployment pattern satisfies the constraint of for WSNs to establish connectivity. In Section 4, we present simulation results to obtain the number of nodes required to cover a deployment region in the diffusion-based as well as other regular deployment patterns. Section 5 presents the entropy of connectivity between nodes for the diffusion based model of node placement. Finally Section 6 concludes the paper and presents directions for future research.

2 Related work

Many placement models for the deployment of sensor nodes have been developed (Hsin and Liu, 2004; Aldosari and Moura, 2006) according to various deterministic and stochastic processes. Node placement in mobile WSNs has been studied by Popa et al. (2006). Hsin and Liu (2004) describe algorithms to improve coverage by using random and coordinated sleep schedules in a network of static sensors deployed according to the Poisson point process. A simple coverage model is described by Wang et al. (2003), where the authors assume the sensing circle of a node as the boundary of the node's coverage region. A point is said to be covered if there is a node whose Euclidean distance to itself is less than the node's sensing range R_s. Wang et al. (2004, 2005) discuss information-theoretic approaches for sensor placement and selection for target localisation and tracking. Dulman et al. (2006) study the relationship between the number of hops and physical distance between nodes in a random deployment of nodes. They show that information about hop counts and node distances obtained from the deployment strategy can be used in designing efficient localisation protocols for WSNs. Krause et al. (2005a) model the sensed spatial phenomena as Gaussian processes. Knowledge of the Gaussian process allows for representation of uncertainty about the sensed field. Using mutual-information as an optimisation criterion, the authors describe an approximation algorithm to find sensor placements that provide most information about un-sensed locations. Koutsougeras et al. (2008) study *event-driven* deployment that deploys nodes based on the concentration of events in the deployment region. With the help of self-organising maps, the authors study the event-driven sensor coverage problem and show that the proposed *event-driven* deployment results in better distribution of sensors in the deployment region. Krause et al. (2005b) optimise sensor placement using mutual information deduced from modelling the sensed phenomenon as a Gaussian process and taking communication costs and sensing quality into consideration. Our work differs in that we study the deployment problem with threefold objectives: placement of nodes to satisfy coverage, connectivity and reliability. While the objective of coverage and connectivity has been widely studied, we study the problem with an additional constraint of reliability. To this end, we present an algorithm for statistically similar placement of static power-limited nodes in homogenous WSNs. We use the general disc-based sensing and coverage model for sensor nodes. The network is deployed for event-driven sensing, where the nodes continuously sense the environment for a given parameter and report to the base station if the parameter deviates from some pre-defined threshold condition, e.g. high temperature. There are two widely used measures of network lifetime, one in which lifetime is defined as the time until the first node runs out of battery energy. The other measure of lifetime is the alpha-lifetime (Zhang and Hou, 2004), which is defined as the entire interval in which at least alpha portion of the region *R* is covered by at least one sensor node. Existing literature in WSNs seeks to improve network lifetime by focusing on parameters such as improvement in coverage (Zhang and Hou, 2004), MAC protocol (Srinivasan et al., 2001) and operating system performance (Gummadi et al., 2005). It is also important to study the problem where we would want a network to run for at least *T* time units with certain reliability metric. Protocols designed for such operation would result in robust real-life deployments, since the sensor network would reliably operate for the desired time interval without being subject to loss of information due to node failure from battery depletion.

3 Diffusion-based deployment approach

We consider the deployment of a WSN of homogenous, static sensor nodes over a deployment region for monitoring applications. We assume that the network of homogenous static sensor nodes is deployed to run for at least T time units, such that every point in the deployment region is continuously sensed for the interval T. We assume the disc model of sensing coverage, and denote the area sensed by a node as A_s, where A_s is equal to πR_s^2, and R_s is the sensing radius of a sensor node. Let the amount of data sensed by a node n_i in unit time interval be I bytes. We also assume k sensors sensing every point in the region at any point of time. Hence, the total amount of data obtained from the sensing region at the end of the time interval T is $kITA/A_s$. We also assume error-free communication links between the nodes and the base station. At the end of the interval T, the base station would have obtained $kITA/A_s$ data from the network if the sensing region were completely covered by at least kA/A_s sensor nodes for the entire interval T. This is the lower bound on the data that can be obtained from the region with the given number and sensing range of sensor nodes. The problem statement is as follows: Given that the minimum total data to be gathered is $kITA/A_s$, where k is the redundancy in the network, I is the data in bytes sensed by a single node with sensing area R_s, T is the desired interval of network operation and A is the deployment area, we need to find a placement strategy that achieves the desired amount of data. We assume the total data is aggregated into a single entity I_T centred at a point p, where point p is an arbitrary location of placement of the first sensor node. We view I_T as a diffusing data source that diffuses across the deployment region. At the end of the deployment process, the total data is spread across the sensing region by the individual nodes that individually sense ITA/A_s amount of data. The diffusion equation for the diffusing source I_T can be written as:

$$\frac{\partial I_T}{\partial t} = D\frac{\partial^2 I_T}{\partial x^2}, \tag{1}$$

where D is the diffusion coefficient. Let x be the Euclidean distance between any two locations of sensor nodes. Our goal is to find the length scale function $L(t)$ (communication range), of the independent variable x such that for any two times t_1 and t_2, there are two locations x_1 and x_2 where the solution appears to be the same. In other words, when distance x is properly scaled, the ratios

$$\frac{I_T(x_1,t_1)}{T(t_1)} = \frac{I_T(x_2,t_2)}{T(t_2)} \tag{2}$$

must have the same numerical value for all pairs of values of time t_1 and t_2 (Bai et al., 2006). There are two possible criticisms of the use of this equations that we will address:

- Equation (1) is the one-dimensional diffusion equation, and since we are studying the deployment pattern in a two-dimensional deployment region, this equation is not appropriate. However, regardless of dimensionality, the diffusion equation can be split into uncoupled dimensionally independent equations (Ursell, 2007). Hence for ease of calculation, we use the one-dimensional equation in two iterations to determine the x and y coordinates in each iteration of the diffusion.
- The data distribution model is a singular, mobile source that is moving around in the sensing field to determine the locations of sensors. Even though WSNs are deployed to sense unpredictable events in varying bit rates, since we assume that there is no data aggregation, the total amount of data bytes accumulated at a central base station is the sum of individual node data at the end of the time duration T. Thus deployment according to the diffusion-based deployment pattern is a reverse approach to designing the network based on anticipated network capacity.

We wish to find the length scale L that will satisfy the diffusion equation (1). The significance of determining the length scale is that it will provide us with the diffusion length. We use the diffusion length to find the distance between nodes. The diffusion equation (1) can be satisfied if it has the form

$$I_T(x,t) = T(t)F(\eta) \tag{3}$$

with $\eta = x/L(t)$. Thus, the solution is the product of a pure function T of time and a function F that depends, at most, on a similarity variable η. In order to find the solution to equation (1) with the given initial and boundary conditions, we compute the partial derivatives of equation (3) and inserting them into equation (1), we obtain,

$$D\frac{\partial^2 F}{\partial \eta^2} + \left(L\frac{\partial L}{\partial t}\right)\eta\frac{\partial F}{\partial \eta} - \left(\frac{L^2\frac{\partial T}{\partial t}}{T}\right)F = 0. \tag{4}$$

This equation can be viewed as an ordinary differential equation in $F(\eta)$, if the coefficients in parentheses are independent of time. Hence, we impose the additional constraints,

$$L\frac{\partial L}{\partial t} = 2D\alpha \tag{5}$$

and

$$\frac{L^2\frac{\partial T}{\partial t}}{T} = D\beta. \tag{6}$$

Equation (4) becomes

$$\frac{\partial^2 F}{\partial \eta^2} + 2\alpha\pi\frac{\partial F}{\partial \eta} - \beta F = 0, \tag{7}$$

where α and β are constants, and the diffusion coefficient D is inserted for simplicity. To find the length scale L and the concentration scale T, integrate equation (5) with respect to t,

$$L^2(t) = L_0^2 + 4\alpha Dt, \tag{8}$$

where L_0 is some initial value. Substituting this value of L from equation (8) in equation (6), we get

$$T(t) = \gamma\left(L_0^2 + 4\alpha Dt\right)^{\beta/4\alpha}, \tag{9}$$

where the multiplicative constant γ can be chosen arbitrarily because the linear diffusion equation is homogenous in I_T, and hence in T. Equations (8) and (9) represent the functional form of the length and concentration scales respectively. Our scale of interest is the length scale L, since it determines the distances between sensor nodes. Hence, α must be real and non-negative, or else from equation (8), L becomes imaginary. Accordingly we solve the diffusion equation for the case $\alpha = 0$. In this case L_0 must be non-zero or the length scale equation (8) ceases to exist. Let the arbitrary constant $\gamma = |L_0|^{\frac{-\beta}{2\alpha}}$ so that equation (9) becomes,

$$T(t) = \lim_{\alpha \to 0}\left(1 + \frac{4\alpha Dt}{L_0^2}\right)^{\frac{\beta}{4\alpha}} = e^{\frac{\beta Dt}{L_0^2}}. \tag{10}$$

Equation (7) becomes

$$\frac{\partial^2 F}{\partial \eta^2} - \beta F(\eta) = 0, \tag{11}$$

where, the similarity variable η is

$$\eta = \frac{x}{L} = \frac{x}{(\pm |L_0|)}. \tag{12}$$

For a bounded solution to the length and concentration scale equations, the exponent in equation (10) should be negative. Let $\beta = -\lambda^2$. Thus the solution to equation (11) takes the form

$$F(\eta) \propto e^{\frac{\pm i\lambda x}{|L_0|}}. \tag{13}$$

Collecting the partial results equations (10) and (13) and substituting them in equation (3), we get

$$I_T(x,t) \propto e^{\frac{-\lambda^2 Dt}{L_0^2}} e^{\frac{\pm i\lambda x}{|L_0|}}, \tag{14}$$

where we need to solve for the constant length scale L_0 to obtain the Euclidean distances between sensor positions in the placement problem. Consider a circular deployment region of radius R that contains the aggregate data source I_T. The deployment process causes the aggregate source to diffuse at a rate proportional to the number of sensors and the distance from the initial position of the source. We define equilibrium of the diffusion process as the time instant t where all the sensors have been placed in the sensing region. Let the equilibrium data concentration be a constant value I_e. We need to solve the diffusion equation throughout the sensing region with the initial and boundary conditions,

$$I_T(r,0) = I_\infty, 0 < r < R \tag{15}$$

and

$$I_T(r,t) = I_e, t > 0. \tag{16}$$

The one-dimensional diffusion equation being a linear equation, any constant is a trivial solution of the diffusion equation as is any first-degree polynomial in x. We use a function $u(r,t) = r(I + k_1)$ (where k_1 is a constant) as a solution of the diffusion-equation of the data source such that

$$I_T(r,t) = I_e + (I_\infty - I_e)\frac{u(r,t)}{r}, \tag{17}$$

which satisfies the initial and boundary conditions.

$$u(r,0) = r, 0 < r < R, \tag{18}$$

$$u(r,t) = 0, t > 0, \tag{19}$$

and

$$\lim_{r \to 0} u/r < \infty. \tag{20}$$

We use the solution developed in equation (14) such that the initial conditions of the unknown function u are satisfied. The data source diffuses along the radius R of the sensing region. The radius R of the deployment region is a natural length scale L_0 and we write the second exponent in expanded trigonometric form,

$$u(r,t) \propto e^{\frac{-\lambda^2 Dt}{L_0^2}}\left[A\sin(\lambda r/R) + B\cos(\lambda r/R)\right]. \tag{21}$$

Thus the complete solution of the data source diffusion-equation from Ghez (2001) is

$$I_T = I_e + 2(I_\infty - I_e)\sum_{n=1}^{\infty}\frac{(-1)^{n+1} e^{\frac{-n^2\pi^2 Dt}{R^2}}\sin(nr\pi/R)}{(nr\pi/R)}. \tag{22}$$

Next, we compute the rate of data source diffusion J_0 across the sensing region.

$$J_0 = -D\frac{\partial I_T}{\partial r} = \frac{2D(I_\infty - I_e)}{R} * \sum_{n=1}^{\infty}\frac{e^{-\left(n\pi\sqrt{Dt}/R\right)^2}\left\{n\pi r/R\cos(n\pi r/R) - \sin(n\pi r/R)\right\}}{(n\pi r/R)^2}. \tag{23}$$

We evaluate the rate of data diffusion at a small distance ε from the initial position of the source $r = 0$, where $\varepsilon \approx 0$. Using the small-angle approximation in equation (23), we get,

$$J_0 = 2D(I_\infty - I_e)\left(n\pi\sqrt{\varepsilon}/\sqrt{2}R\right)\sum_{n=1}^{\infty} e^{\left(-n\pi\sqrt{Dt}/R\right)^2}. \tag{24}$$

We denote the quantity $\sqrt{Dt}$ as the diffusion length. If the diffusion length is small compared to the radius of the sensing region, the exponent in equation (24) varies slowly with n and we can replace the sum by an integral over the continuous variable

$$\xi = n\pi\sqrt{Dt}/R.$$

Hence, we get,

$$J_0 = \left(I_\infty - I_e\right)* \left(n\pi\sqrt{\varepsilon}/\sqrt{2}R\right)R\sqrt{Dt}/\sqrt{\pi}erfc\left(\pi\sqrt{Dt}/R\right). \tag{25}$$

Denote $d_l = \sqrt{Dt}$ as the diffusion length and re-arranging the terms we get,

$$\frac{J_0}{m} = d_l erfc\left(\frac{\pi d_l}{R}\right), \tag{26}$$

where m is a constant given by

$$m = \left(I_\infty - I_e\right)\left(n\pi\sqrt{\varepsilon}/\sqrt{2}R\right)R/\sqrt{\pi}. \tag{27}$$

The problem of finding the distances between sensor nodes thus reduces to determining d_l at time instants t. Since the diffusion-length d_l is small compared to the dimensions of the sensing region, the argument of the complementary error function, $\frac{\pi d_l}{R} \to 0$.

$$d_l = \frac{J_0}{m}t, \tag{28}$$

J_0, the rate of data diffusion, can be viewed as the diffusion of an amount equal to ITA/A_s from the data source at every time instant when a sensor is deployed at a distance d_l from the position of the source at the previous time instant. Conserving the dimensionality of J_0,

$$J_0 = \frac{ITA/A_s}{d_l^2}. \tag{29}$$

Therefore the diffusion length d_l , i.e., the Euclidean distance between sensor nodes for a diffusion-based placement of sensor nodes across the sensing region is given by

$$d_l = \sqrt[3]{\frac{1}{m}\frac{ITA}{A_s}t}. \tag{30}$$

3.1 Deployment algorithm for sensor nodes according to the diffusion-based deployment algorithm

Since we know the Euclidean distance d_l between sensor nodes for the placement problem, we now define the graph G, which is obtained by tracing the motion of the mobile data source along the sensing region.

3.1.1 Simulation of the diffusion-based deployment pattern

We denote the point where a sensor is deployed as the vertex of the graph G, and the distance between sensors deployed at time instants t and $t + 1$ as an edge of G. The length of this edge is d_l. Our aim is to draw planar graphs iteratively in the plane given by the sensing region. The algorithm proceeds as follows:

Step 1: Starting from the center of the plane, we draw a planar graph G_1 over the plane of the sensing region that traces the motion of the mobile data source, i.e. distance between sensors deployed at time instants t and $t + 1$ is d_l. In this step, we deploy A/A_s number of sensors in the region. This step deploys the minimum number of sensors required to cover the region.

Step 2: Repeat step one to increase the redundancy of coverage. This procedure continues for k steps, where k is the known value of redundancy that we wish to achieve in the network. At the end of k iterations, the plane of the sensing region can be viewed as covered by overlay of planar graphs, each of which covers the sensing region completely. The overlaying of planar graphs G_i, where $i = 1, 2, ...k$, completely covers the sensing region with redundancy k.

3.2 Relationship between the diffusion length and the sensing radius

Since the communication range R_c has been shown to be at least twice the sensing radius R_s (Zhang and Hou, 2004) for connectivity, we check if our deployment satisfies this condition of connectivity. Re-writing equation (30) with $A_s = \pi R_s^2$, we get,

$$d_l = \sqrt[3]{\frac{1}{m}\frac{ITA}{\pi R_s^2}t} = jR_s^{-2/3}, \tag{31}$$

where j is a constant given by $j = \sqrt[3]{\frac{1}{m}\frac{ITA}{\pi}t}$ for a constant t. Since the area of the sensing region A is large and the time for which the network is deployed T is large, $j >> 2$,

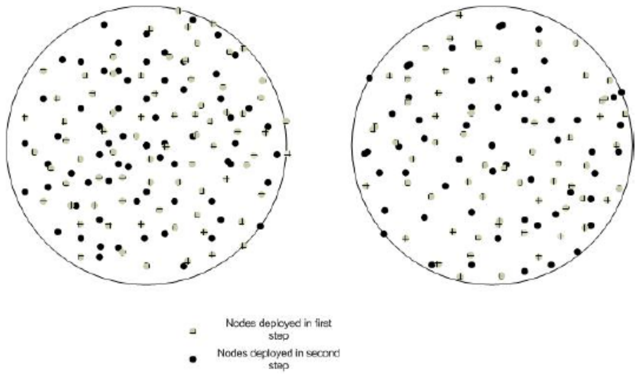
$$d_l > 2R_s^{-2/3}. \tag{32}$$

For the relationship between d_l and sensing radius R_s, as $R_s \to \infty$

$$d_l = \lim_{R_s \to 0} 2R_s^{-2/3} = 0, \tag{33}$$

This equation holds because, as $R_s \to \infty$, the entire sensing region can be covered by only one sensor and there is no requirement for other sensors to be placed at distances equal to the diffusion length from the position of the previous sensor. For finite values of R_s, the diffusion length is always greater than twice the sensing radius and satisfies connectivity between adjacent nodes. Thus, our deployment problem satisfies the conditions for coverage and connectivity in the sensing region. Figure 1 shows the topology resulting from diffusion-based deployment pattern for two different ratios of communication radius to the sensing radius, i.e. r_c/r_s .

Figure 1 Diffusion-based deployment for varying r_c/r_s. The figure on the left depicts $r_c/r_s = 2$ and the figure on the right depicts $r_c/r_s = 5$ (see online version for colours)



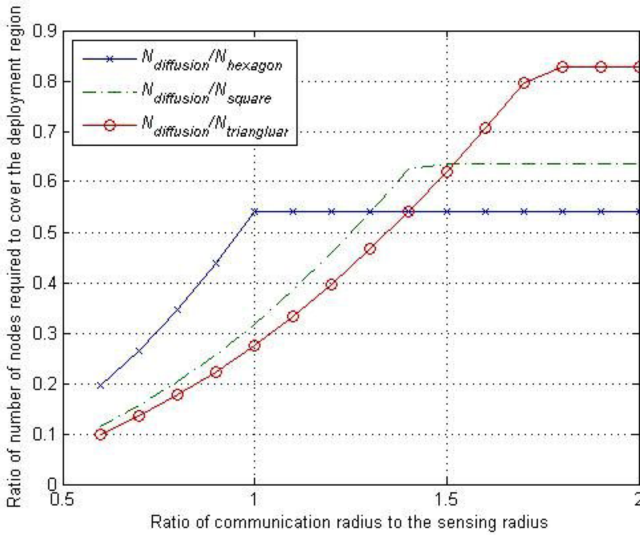
4 Results

We now show how the diffusion-based sensor deployment pattern results in higher reliability while requiring less number of sensors than regular deployment patterns. From Section 3.2, we see that the diffusion-length d_l is the distance between sensors deployed at successive time instants. When there is no redundancy in the network, d_l is also the minimum distance required for nearest sensor neighbours to communicate. Hence, the ratio of the diffusion length to the sensing radius d_l/R_s is also the widely studied coverage and connectivity metric r_c/r_s, which is the ratio of the communication range to the sensing range.

4.1 Number of sensors required to sense the entire deployment region

In this section, we show that the diffusion-based deployment pattern satisfies not only reliable operation by sensing I_T amount of data; it requires less number of sensors than regular deployment patterns with the required level of *k*-redundancy in the network. To show this, we use the results obtained by Bai et al. (2006) where they define a parameter called the Area Per Node (APN) to show that out of the four regular patterns of deployment, namely square, triangular lattice, rhombus and hexagonal patterns, no pattern is best for all values of r_c/r_s. The APN denotes the area of the Voronoi polygon of all points in a Euclidean plane *S*. The number of nodes needed to cover the entire sensing region for a given regular deployment pattern is obtained by dividing the area of the deployment region by the maximum APN. We use R_s = 30 m and a circular deployment region of radius 1000 m. The communication range is varied between 15 m and 30 m. Figure 2 shows the ratio of the number of sensors needed to cover the deployment region using the diffusion-based approach to the number of sensors needed for each of the regular deployment patterns. The diffusion-based approach requires 63% ($N_{diffusion}/N_{square}$), 54% ($N_{diffusion}/N_{hexagon}$) and 82% ($N_{diffusion}/N_{triangular}$) less number of sensors than the square, hexagon and triangular-based deployment patterns.

Figure 2 Ratio of number of nodes needed in the diffusion-based approach to that in the square, hexagonal and triangular patterns (see online version for colours)



4.2 Reliability of WSN as a function of the deployment pattern of nodes in the network

In this section, we derive the reliability of the data obtained from a wireless sensing network as a function of the amount of data sensed by the nodes in the network. Since every node senses a quantity of data *I* in unit time, the total amount of data sensed by the network of sensor nodes for duration *T* of deployment is given by $I_T = kITA/A_s$. But,

$$d_l = r_c = \sqrt[3]{\frac{ITA}{A_s}t} = \sqrt[3]{\frac{ITN}{m}t}, \tag{34}$$

where $N = A/A_s$ is the number of nodes needed to cover the region completely according to the disc model for sensing. Solving for I_T and substituting for $N_{diffusion}$, N_{square}, $N_{hexagon}$ and $N_{triangular}$, we compare the amount of data sensed by the diffusion-based pattern with that in other regular deployment patterns namely, $I_T - diffusion$, $I_T - square$, $I_T - triangular$ and $I_T - hexagon$ and plot these ratios in Figure 3. The diffusion-based pattern has higher reliability as shown by improvement in amount of data bytes sensed. As seen from Figure 5, the improvement in reliability from the diffusion-based deployment pattern is by a constant factor equal to 1.9 (hexagonal), 1.7 (square) and 1.3 (triangular lattice) patterns. The transition behaviour of the curves in Figures 2 and 3 at the values 1, $\sqrt{2}$ and $\sqrt{3}$ is due to the bounds of the APN and has been derived by Bai et al. (2006) for the hexagonal, square and triangular patterns, respectively. We direct the reader to Bai et al. (2006) for a detailed derivation of the APN for various regular deployment patterns. As seen in Figure 4, the greater efficiency of the diffusion-based deployment algorithm is because the coverage from nodes placed at the same point is treated differently in the diffusion-based deployment pattern and the regular deployment patterns studied by Bai et al. (2006). In the diffusion-based deployment pattern, if $n > 1$

sensors are deployed at a point, the coverage provided by these sensors is considered to be n times the coverage provided by a single sensor. Defining vacancy as the area in the deployment region that is not within the sensing range of any node, the diffusion-based deployment pattern results in a lower vacancy by virtue of the redundancy coverage which is factored into the estimation of coverage in the network. This is illustrated in Figure 4, where two sensors n_i and n_j are deployed at adjacent locations. For the diffusion-based deployment, the computation of coverage provided by n_i and n_j is represented by the dotted circles. A more realistic coverage computation is shown for regular deployment patterns, where the coverage provided by n_i and n_j is the union of the coverage provided by n_i and n_j. These differing computations of coverage result in the diffusion-based deployment pattern requiring lesser number of sensors and providing more data yield than that obtained in regular deployment patterns.

Figure 3 Data sensed as a function of the coverage in regular and diffusion-based patterns (see online version for colours)

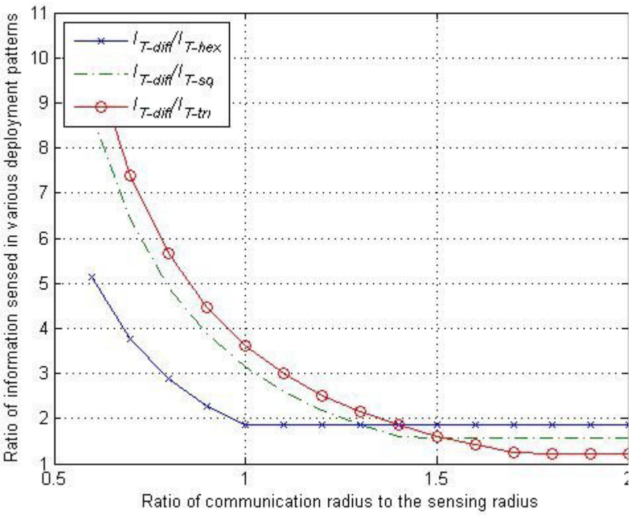
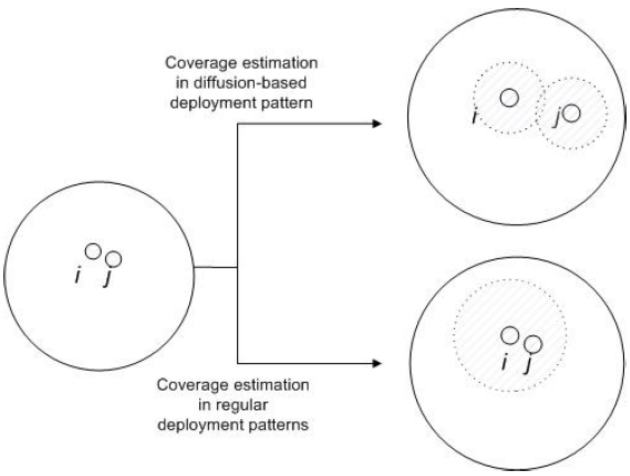


Figure 4 Ratio of the amount of data sensed by the WSN during time T as a function of r_c/r_s (see online version for colours)



5 Entropy of establishing connectivity between nodes

In this section, we study the entropy of establishing connectivity between nodes in WSNs. We use the results of diffusion length from Section 3 to obtain the entropy of establishing connectivity from nodes to a central base station. We then establish the bounds on this entropy for a heterogeneous WSN of two types of nodes: computationally powerful nodes like cluster-heads (CHs) that act as local data processing centres and points of communication from nodes to base station (BS) and regular nodes that gather data about the phenomenon and transmit it to a local CH. Here we extend our analysis of connectivity entropy to heterogeneous networks and establish the dependence of this entropy on the density of nodes in each tier of the network. In the rest of this paper, we will refer to the powerful nodes in the second tier as CHs, though this framework can be extended to different kinds of heterogeneous networks.

In this section, we study the following problem: What is the entropy of establishing connectivity E_c to a BS from any node in the WSN? Nodes can connect to a BS through multiple nodes that act as relays. We provide an analytical framework for deriving the dependence between E_c and the distributions of nodes in the deployment region. Since the randomness of the wireless channel plays an important role alongside node locations in determining the connectivity graph of the network, we model the wireless channel with randomness introduced by a lognormal shadow fading environment. This implies that a link between two nodes n_i and n_j separated by a Euclidean distance $s(i,j) = d_l$ exists only if the signal attenuation between the nodes does not exceed the threshold attenuation ratio for communication between the nodes. Knowing the Euclidean distance between the nodes n_i and n_j, we denote the probability of forming a link between the nodes by $P\left(\Lambda(i,j) \mid s(i,j)\right)$. In a shadow fading environment, the expression for $P\left(\Lambda(i,j) \mid s(i,j)\right)$ is given by Bettstetter (2002).

$$P\left(\Lambda(i,j) \mid s(i,j)\right) = \frac{1}{2} - \frac{1}{2} erf\left(\frac{10\alpha}{\sqrt{2}\sigma} \log_{10} \frac{s(i,j)}{r_0} dB\right), \quad (35)$$

where α is the path loss exponent due to the deterministic geometric component of attenuation in a shadow fading environment. σ denotes the standard deviation of the stochastic component chosen from a normal probability density function of attenuation. r_0 is the normalisation term denoting the maximum distance granting a link in the absence of shadow fading. For threshold attenuation β, r_0 is given by

$$r_0 = 10^{\left(\frac{\beta_{th}}{\alpha.10db}\right)}.$$

Further, we model the entropy to establish connectivity as a random sequence of vertices that a node n_i traverses as it proceeds to establish connectivity with the base station. The probability of choosing another node n_j as the next hop node is proportional to the weight of the edge linking nodes n_i and n_j and is given by P_{ij}. The weight of an edge is given by the probability of forming a link between the nodes n_i and n_j subject to the shadow fading environment. P_{ij} can be expressed as the ratio of the weight of edge W_{ij} between nodes n_i and n_j, to the total weights of all edges W_i emanating from node n_i. Thus P_{ij} is given by

$$P_{ij} = W_{ij}/W_i. \tag{36}$$

Additionally, since nodes are assumed to be randomly assigned the *sleep* or *awake* states and since a node in the *sleep* state cannot form links with other nodes, the weight of an edge linking node n_i to another node n_j in the *sleep* state is taken to be zero. We also assume that node n_i knows the number of neighbours it has, by counting the number of links it can form. The node n_j with the highest edge weight (lowest signal attenuation) is chosen as the next hop node for node n_i. The sequence of nodes that a node n_i uses to reach the base station is modelled as the sequence of random states $\{S_i\}$, where the maximum value of n is given by the number of nodes in the network. The sequence of states can be expressed as

$$S_0 = 0. \tag{37}$$

and $S_1 = P_r$ (randomly choosing any node in the network as the starting vertex).

$$S_{n+1} = S_n * P_{ij}. \tag{38}$$

Thus the entropy to establish connectivity E_c is the entropy of the random sequence of states $\{S_i\}$ and is given by,

$$H(S_1, S_2, \ldots S_n) = \sum_{i=1}^{n} H(S_i \mid S^{i-1}) \tag{39}$$

$$= H(S_1) + \sum_{i=2}^{n} H(S_i \mid S^{i-1}). \tag{40}$$

where equation (39) holds because of the chain rule of entropies and equation (40) is due to property of Markov chains. Let l be the number of edges emanating from node n_i, i.e. the probability of forming a link from node n_i to l other nodes exists. Thus, the second term of the expression can be expressed as,

$$\sum_{i=2}^{n}\sum_{j=1}^{l} H\left(W_{ij}/W_i\right). \tag{41}$$

where H depends on the (probability of forming a link from node n_i to node n_j)/ sum of probabilities of forming a link from node n_i to all l nodes. l is the number of neighbour links that a node has in presence of shadow fading environment. The entropy of starting the graph at any node is given by

$$H(S_1) = -1/n \log 1/n, \tag{42}$$

where n = number of nodes in the network. To obtain the second term of the expression for E_c, we note that $P\left(\Lambda(i,j) \mid s(i,j)\right)$ is given by equation (35) and W_i is obtained as the sum of the probabilities $P\left(\Lambda(i,j) \mid s(i,j)\right)$ on all links emanating from node n_i.

5.1 Bounds on E_c

To study the entropy to establish connectivity between tiers of nodes in a heterogeneous network, we consider a WSN of two types of nodes: a homogeneous set of wireless sensor nodes that perform sensing and another homogeneous set of CHs, where each CH aggregates data from the nodes belonging to its cluster. The deployment region is assumed to be a two-dimensional Euclidean space of area A in which the wireless sensor nodes are distributed as a Boolean model formed by a Poisson-distributed sequence of random sets. These random sets are the disk-shaped circles of coverage of individual nodes with radius r. The centers of coverage circles are assumed to form a stationary Poisson process of intensity λ_1. We define another similar stationary Poisson process of intensity λ_2 for the CHs. The CHs are assumed to have greater processing power to enable inter-cluster communication capabilities and for data aggregation from nodes belonging to its cluster. Clustering is a realistic scenario for large-scale, dense WSNs deployed according to a random topology for sensing in remote, hostile deployment regions.

We model the links between nodes and the CHs as a connected random graph, where the nodes comprise the vertices of the random graph, and the links between nodes comprise the edges of the random graph. The links between nodes represent the connectivity between nodes and CHs and the number of nodes attached to a CH represent the cluster size. We assume that a node can be either in *sleep* or *awake* states. In the *sleep* state, a node's transceiver is turned off and hence cannot form links with any other nodes. In this section, we provide upper and lower bounds on E_c and show its dependence on the intensity of CH distribution as well as on the probability of connectivity between nodes, $P\left(\Lambda(u,v) \mid s(u,v)\right)$.

If the WSN is designed for operation such that all wireless sensor nodes as well as CHs are awake at all times, the probabilistic nature of a node being in the 'awake' or 'sleep' states is eliminated. The sequence of states $\{S_i\}$ thus depends only on the probability of forming links between nodes, i.e. when the signal attenuation stays below the threshold required for communication. The upper bound on the entropy E_{c-max} is obtained when all the nodes are awake and every node i maintains links to all its neighbors who lie within the one-hop communication range. Additionally, the assumption of the log-normal distribution allows for forming links with nodes located greater than distance d_l away from a node i. The sequence of states

$\{S_i\}$ can then be modelled by a random walk on a connected graph with a stationary Markov chain, and its entropy E_{c-max} is given by

$$\begin{aligned} E_{c-\max} &= H(S_1, S_2, ..., S_n) = H(S_2 \mid S_1), \\ &= \sum_{i=1}^{n} \mu_i H(S_2 \mid S_1 = i), \end{aligned} \quad (43)$$

where μ_i is the stationary distribution of the Markov chain. The lower bound on the entropy E_{c-min} can be similarly obtained. When the attenuation due to fading results in zero probability of connectivity for any node pair, the graph formed by the wireless sensor nodes and CHs does not contain any edges. Thus the lower bound E_{c-min} is just the entropy of choosing any given node as the starting vertex to analyse the entropy of connectivity to the CH.

$$E_{c-\min} = H(S_1). \quad (44)$$

While the above analysis holds for a given intensity λ_2 of distribution of CHs, we can also present bounds on E_c for varying levels of CH distribution. The bounds in this case are readily obtained. If the intensity of distribution of CHs, λ_2, is increased to be equal to the intensity λ_1 of wireless sensor nodes while keeping the area of the deployment region constant, the distribution of CHs relative to that of the nodes becomes a high-intensity distribution and the expected number of nodes per cluster is one. In this case the entropy of the sequence of states, assuming connectivity between nodes and CHs, is just the entropy of choosing one out of *n* nodes in the WSN and provides the lower bound on entropy for equal distributions of CHs and wireless sensor nodes. Thus,

$$E_{c-\min} s.t.(\lambda_2 = \lambda_1) = H(S_1) = -1/n \log 1/n. \quad (45)$$

Similarly, when the distribution of CHs is much less than that of nodes, i.e. $\lambda_2 = \lambda_1$ the expected cluster size per CH increases. The largest cluster size is obtained when the intensity of CH distribution λ_2 equals zero. This results in a WSN with only one CH for all $n = kA/A_s$ wireless nodes, and the entropy E_c is the highest. Thus the upper bound on E_c, assuming connectivity among nodes and CHs is given by

$$\mathrm{E}_{c-\max} \mathrm{s.t.}(\lambda_2 = 0) = H(S_1, S_2, ..., S_n). \quad (46)$$

6 Concluding remarks

In this paper, we addressed the problem of reliability in WSNs from the point of view of data obtained from the sensing operation in the deployment region. We proposed a diffusion-based approach to the placement of nodes in the sensing region by modelling the total data content of the network as hypothetical, non-stationary, singular data source. At every time instant when a node is deployed, the source diffuses an amount of data equal to the data content sensed by a single node over the duration of deployment. The solution to the diffusion-based deployment problem is a function of the communication radius of the sensor node. We also investigated the ratio between the number of sensors required to cover the sensing region using the diffusion-based approach to those required using regular (square, hexagon and triangular-lattice) deployment patterns and found that the diffusion-based pattern requires lesser number of sensors than the aforementioned regular patterns. We introduced a measure of reliability of network operation in terms of the total amount of data sensed by the network and showed that the reliability of network operation is a function of the deployment pattern of nodes in the network. Finally, we measured the reliability of WSNs deployed using regular deployment patterns and showed that the diffusion-based approach to deployment of sensor nodes results in reliable network operation compared to other regular deployments. We also provided an analytical framework to determine the entropy of establishing connectivity in heterogeneous networks in a shadow fading environment. The sequence of nodes used to determine E_c may not be the shortest path to the powerful node, but is more robust since at every time instant, a node n_i chooses the next-hop node that has the least attenuation among all nodes linked to node n_i. Our results draw attention to the need for sensitivity analysis of distributions of CHs relative to the number of nodes in a WSN. We also present bounds on the entropy to establish connectivity and show its dependence on the density of CHs as well as on the attenuation caused to the randomness in the wireless channel.

Future work in this direction includes incorporating edge effects by providing efficient coverage in the boundary of the deployment region. More work in this area can be investigation of the problem of stochastic sources of error in sensed data due to device unreliability and sink-initiated error detection and recovery problem for deviant sensors in WSNs.

References

Ahn, J. and Krishnamachari, B. (2006) 'Fundamental scaling laws for energy-efficient storage and querying in wireless sensor networks', *7th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Florence, Italy, pp.334–343.

Aldosari, S. and Moura, J. (2006) 'Topology of sensor networks in distributed detection', *IEEE International Conference on Signal Processing*.

Bai, X., Kumar, S., Xuan, D., Yun, Z. and Lai, T. (2006) 'Deploying wireless sensors to achieve both coverage and connectivity', *ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Vol. 1, No. 1, pp.131–142.

Bettstetter, C. (2002) 'On the connectivity of wireless multihop networks with homogeneous and inhomogeneous range assignment', *Proceedings of the 56th Vehicular Technology Conference*, pp.1706–1710.

Dulman, S., Rossi, M., Havinga, P. and Zorzi, M. (2006) 'On the hop count statistics for randomly deployed wireless sensor networks', *International Journal of Sensor Networks*, Vol. 1, No. 1, pp.89–102.

Ghez, R. (2001) *Diffusion Phenomena: Cases and Studies*, Springer.

Gummadi, R., Gnawali, O. and Govindan, R. (2005) 'Macro-programming wireless sensor networks using Kairos', *Proceedings of the International Conference on Distributed Computing in Sensor Systems*.

Hsin, C-F., and Liu, M. (2004) 'Network coverage using low duty-cycled sensors random and coordinated sleep algorithms', *Information Processing in Sensor Networks*, pp.433–442.

Koutsougeras, C., Liu, Y. and Zheng, R. (2008) 'Event-driven sensor deployment using self organizing maps', *International Journal of Sensor Networks*, Vol. 3, No. 3, pp.142–151.

Krause, A., Guestrin, C. and Kleinberg, J. (2005a) 'Near-optimal sensor placements: maximizing information while minimizing communication cost', *Proceedings of 5th International Conference on Information Processing in Sensor Networks*, Nashville, Tennessee, USA.

Krause, A., Singh, A. and Williams, C. (2005b) 'Near-optimal sensor placements in Gaussian processes', *Proceedings of the International Conference on Machine Learning*.

Popa, D., Mysorewala, M. and Lewis, F. (2006) 'Deployment algorithms and in-door experimental vehicles for studying mobile wireless sensor networks', *International Journal of Sensor Networks*, Vol. 6, No. 1, pp.28–43.

Srinivasan, V., Nuggehalli, P. and Rao, R. (2001) 'Design of optimal energy aware protocols for wireless sensor networks', *Proceedings of the 53rd Vehicular Technology Conference*, Vol. 4, pp.2494–2498.

Ursell, T. (2007) *The Diffusion Equation: A Multi-dimensional Tutorial*, Department of Physics, California Institute of Technology.

Wang, H., Yao, K. and Estrin, D. (2005) 'Information theoretic approaches for sensor selection and placement in sensor networks for target localization and tracking', *Journal of Communications and Networks*, Vol. 7, pp.438–449.

Wang, H., Yao, K., Pottie, G. and Estrin, D. (2004) 'Entropy-based sensor selection heuristic for target localization', *Proceedings of 3rd International Symposium on Information Processing in Sensor Networks*, pp.36–45.

Wang, X., Xing, G., Zhang, Y., Lu, C., Pless, R. and Gill, C. (2003) 'Integrated coverage and connectivity configuration in wireless sensor networks', *ACM Conference on Embedded Sensor Network Systems*, pp.28–39.

Zhang, H. and Hou, J. (2004) 'On deriving the upper bounds of alpha lifetime for large sensor networks', *5th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp.121–132.