

Sizes of Minimum Connected Dominating Sets of a Class of Wireless Sensor Networks

Ji Li¹, Chuan Heng Foh², Lachlan L. H. Andrew³, Moshe Zukerman¹

¹ *ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN), EEE Dept., The University of Melbourne, Victoria 3010, Australia, {jili, m.zukerman}@ee.unimelb.edu.au*

² *School of Computer Engineering, Nanyang Technological University, Singapore 639798, aschfoh@ntu.edu.sg*

³ *Department of Computer Science, California Institute of Technology, lachlan AT caltech.edu*

Abstract— We consider an important performance measure of wireless sensor networks, namely, the least number of nodes, N , required to facilitate routing between any pair of nodes, allowing other nodes to remain in sleep mode in order to conserve energy. We derive the expected value and the distribution of N for single dimensional dense networks.

I. INTRODUCTION

To prolong network lifetime, energy conservation is a major concern for wireless sensor networks due to the difficulty of battery replacement. To achieve this goal, short-hop-communication are generally considered as a better choice compared to its long-hop counterpart, mainly because it requires, in principle, less power than longer hops. However, as sensors often lack power control, power consumption depends primarily on the number of hops taken. More importantly, making fewer, longer hops allows nodes to spend longer in a low-power “sleep” mode, prolonging the overall network lifetime. Thus it is important that the number of nodes involved in routing is kept small.

To traverse least number of hops in routing, we need to construct a wireless backbone, and minimize its size. Only nodes in that backbone set are involved in routing. One promising approach to construct such a backbone is based on Connected Dominating Set [1].

A **Connected Dominating Set (CDS)** of a sensor network is defined as a subset of nodes, called ‘relay nodes’ which form a connected network, such that any node in the original network is either a member of the CDS or is within the transmission range r of at least one node in the CDS. The CDS with minimum cardinality is known as **Minimum Connected Dominating Set (MCDS)**. Finding the MCDS in a connected network was proved to be NP-complete [2]. Publications related to the utilization of CDS in ad hoc network routing include [1], [3], [4], and references therein. Those studies are mainly concentrate on developing heuristic algorithms to compute approximated MCDS.

In contrast with previous works, we do not aim at providing algorithms to obtain the sub-optimal CDS for a given network instance. We consider a single dimension stochastic network, and derive the probabilistic properties of the cardinality of MCDS instead of finding MCDS explicitly. Indeed, wireless sensor networks are often deployed in inaccessible terrain with a large number of sensor nodes, which prevent them from

being placed deterministically. Therefore, random deployment is a widely accepted sensor network implementation scheme and motivates this research.

It should be noted that 1-dimensional (1-D) sensor networks have practical importance. For example, there have been proposals for networking cars on main roads [5] [6] for purposes such as reporting traffic disruptions, and sensor networks monitoring rivers, or deployed along a mountain ridge, etc., which can be approximated as 1-D networks. More importantly, recent research shows that the optimal node placement pattern to achieve both coverage and connectivity in 2-D networks is a strip-based pattern [7], which implies that single dimension results are useful when considering MCDS asymptotically in 2-D networks.

Our main contributions are new approaches to computing the distribution and the mean of cardinality of MCDS. These results are important because they can be adopted to evaluate the performance of sensor network routing protocols in terms of the least number of nodes involved in routing. Moreover, they also allude to the critical performance measure, network lifetime, in sensor networks. Though our results are for single dimension networks, the approaches to obtain these results do suggest a potential methodology to evaluate MCDS in two dimensional networks.

The rest of this paper is organized as follow. In Section II, we define our problem and notation first, and then introduce a 1-D network model; in Section III, we provide a simple approach to obtain the mean of N ; a numerical method for the distribution of N is proposed in Section IV. Simulation and numerical results in Section V show the validity of our evaluation approaches. We discuss 2-D cases briefly in Section VI and conclude this work in Section VII.

II. PROBLEM STATEMENT AND NETWORK MODEL

For a 1-D sensor network with n nodes randomly and uniformly distributed in an interval, we study the minimum cardinality of its CDS, which is also the necessary number of nodes involved in the Route Request (RREQ) broadcasting in routing to achieve complete RREQ delivery in the entire network.

In the network we are considering, let the coordinates of the n sensors be $X_1 \leq X_2 \dots \leq X_n$. In order to make CDS minimum in 1-D topology, **greedy** routing must be applied.

Greedy routing is defined as a routing scheme in which every node involved in data transmission forwards packets to the neighbor that is closest to the destination [8]. In 1-D topology networks, the resulting hops will have longest length in the direction toward the destination among all possible hops, so we also call it longest hop routing in this paper.

This paper uses the following notation:

- n — total number of nodes in the network.
- d — total network distance, which is the length of interval that all sensor nodes uniformly distributed in.
- r — radio transmission range, which is identical for all the sensor nodes.
- N_{CDS} — the cardinality of a particular CDS.
- $N = N = \min_{CDS's} (N_{CDS})$.
- $P(N)$ — probability mass function of N . We will denote $P(N = k)$ the probability that N takes a particular value k .
- C, \bar{C} — C denotes event that the network is connected, i.e., there exists at least one path between any pair of sensor nodes; and \bar{C} is complement event of C , i.e., network is disconnected.
- L — the network span, $L = X_n - X_1$.
- $F_L(\cdot), F_L(l|C)$ — cumulative distribution function (CDF) of L and the CDF of L conditioning on the network connectivity.
- $f_L(\cdot)$ — probability density function (pdf) of L .
- W_i — W_i is the length of the i th hop in greedy routing, in which $i = 1, 2, \dots, N$.
- T_i — the residual of i th hop, i.e., $T_i = r - W_i$.
- D_k — sum of length of first k hops, $D_k = \sum_{i=1}^k W_i$.

Variables N, L, W_i, T_i are all random, dependent on the random placement of sensors.

The standard assumption that direct links only exist between any two nodes with straight distance no more than a predefined threshold r is adopted here. This may not be the case in some practical situations. However, in addition to its analytical tractability, this model implies a bound if we select r as the smallest value of transmission range. Without loss of generality, we normalized all the distance parameters with transmission radius r , therefore, $r = 1$.

In a sensor network, when the number of nodes n is small, it is highly possible that the network will be disconnected. Therefore it is meaningless to discuss the number of relay nodes in this situation, and it is not the case that we are interested in. We assume that n is large such that the network is connected with high probability. From [9] pages 23 - 25, we know that a collection of n Independent and Identically Distributed (IID) uniform random points on $[0, d]$ has identical exponentially distributed distance with parameter d/n between any pair of adjacent points in the limit as n goes to infinity, i.e., $(X_{i+1} - X_i) \sim \text{Exp}(d/n)$, $i = 1, 2, \dots, n-1$. In this work, since we are considering dense sensor network, it is reasonable to adopt this approximation and we approximate the uniform point process as a Poisson process. In [10], the expected number of hops with longest hop routing until a 1-D Poisson distributed network is getting disconnected is studied. Our work differs from it as we assume the network with volume d is connected and obtain the cardinality of MCDS. As

n is large, W_i 's, for $i = 1, 2, \dots, N$ are considered to be IID random variables roughly. Furthermore, each individual W_i 's is approximately independent of N . These approximations are supported by simulation results in Section VII. The network model is shown in Figure 1. As the following analysis is based on the Poisson approximation, our results will not only apply to uniformly distributed network with large value of n , but are also applicable to networks with all nodes forming a Poisson process, which increases the applicability of our results.

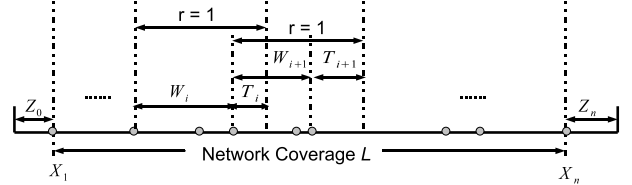


Fig. 1. Network Model for Poisson Approximation

III. MEAN OF N

In order to evaluate the performance of a given routing protocol in terms of the necessary number of relays to be involved, N , we derive its expectation directly, not via its distribution, which is done numerically in Section IV.

In a system of many nodes that provides a connected network, after each hop there will usually be another node “near” the radio range, and so all hops will be only slightly below r . Moreover, these hop lengths will be approximately independent. (In the Poisson limit, the length $r - W_i$ is an exponential truncated to W_{i-1} , with the dependence between W_i and W_{i-1} only through this truncation.) Thus

$$E(N) \approx E(L|C)/E(W) \approx E(L)/r. \quad (1)$$

The rest of this section quantifies how much below r the average hop is, refining the second approximation above, and obtains $E(N)$ by studying $E(L|C)$ and $E(W)$. We first derive the distribution and density functions of the network coverage L .

Lemma 1: Consider sensor nodes randomly deployed in 1-D space, forming a Poisson point process with parameter λ . Let n be the number of nodes falling in $[0, d]$. The expected value of the network coverage L is:

$$E(L) = d - \frac{2}{\lambda} + O(e^{-\lambda d}). \quad (2)$$

Proof: Let $Z_0 = X_1$ and $Z_n = d - X_n$. Note they are truncated exponentials, with pdf

$$f_Z(x) = \frac{\lambda \exp(-\lambda x)}{1 - \exp(-\lambda x)}, \quad x \in [0, d]$$

and 0 otherwise, and hence have mean

$$E(Z) = \frac{e^{\lambda d} - \lambda d - 1}{\lambda(e^{\lambda d} - 1)} = \frac{1}{\lambda} - \frac{d}{e^{\lambda d} - 1} = \frac{1}{\lambda} + O(e^{-\lambda d}) \quad (3)$$

for large λ and fixed d . Combining (3) with the fact that the network coverage is $L = d - (Z_0 + Z_n)$ gives (2). ■

Equation (2) shows that as λ goes to infinity, $E(L) \approx d - 2/\lambda$, which indicates that L is approximately the difference between d and the sum of two exponential RV's. Based on this observation, we can approximate the probability density function (pdf) and cumulative distribution function (CDF) of the network coverage L , $f_L(l)$ and $F_L(l)$ as:

$$f_L(l) = \lambda^2(d-l)e^{-\lambda(d-l)}, \quad l \in [0, d] \quad (4)$$

and

$$F_L(l) = e^{-\lambda(d-l)}(1 + \lambda(d-l)), \quad l \in [0, d] \quad (5)$$

Indeed, we approximate Z_0 and Z_n as IID exponential RV's, resulting in (4) and (5).

The following lemma shows that $E(L|C)$ is a good approximation to $E(L)$ for dense networks; both of them tend to d , but the difference between $E(L|C)$ and $E(L)$ is much smaller than the difference between either of them and d .

Lemma 2: Consider sensor nodes randomly distributed in 1-D space according to a Poisson point process with parameter λ . Nodes within the required network terrain, which is an interval $[0, d]$, forming an sensor network. As $\lambda \rightarrow \infty$, the cumulative distribution function (CDF) of the network coverage, L , and the CDF of L conditioned on the network being connected satisfy

$$F_L(l|C) - F_L(l) = O(e^{-\lambda r/2}) = o(1/\lambda), \quad \lambda \rightarrow \infty. \quad (6)$$

In particular, the expected L and conditional expectation of L given the connectivity of the network satisfy

$$E(L|C) - E(L) = o(d - E(L)), \quad \lambda \rightarrow \infty. \quad (7)$$

Proof: To show (6), note that

$$F_L(l|C) - F_L(l) \quad (8)$$

$$= F_L(l|C)P(C) + F_L(l|C)P(\bar{C}) - (F_L(l|C)P(C) + F_L(l|\bar{C})P(\bar{C}))E(L|C). \quad (9)$$

The factor in brackets is bounded above by $1 = O(1)$. To bound \bar{C} , partition $[0, d]$ into $2d/r$ intervals, each of length at most $r/2$. A sufficient condition for the network to be connected is that each interval has at least one node, which occurs with probability at least $(1 - e^{-\lambda r/2})^{2d/r}$. Thus $P(\bar{C}) = O(e^{-\lambda r/2})$ for large λ and fixed r, d .

Also, note that $d - E(L) = 2/\lambda + o(1/\lambda)$ from Lemma 1, giving (7). ■

To derive the distribution of W 's, we consider a sensor network with sensor nodes forming a Poisson point process with parameter λ in 1-D space. Let W_0 be the random hop length in the longest hop routing scheme, for any but the last hop. The CDF and pdf of W_0 can be approximated by

$$F_{W_0}(t) \approx 1 - \frac{1 - \exp(-\lambda(r-t))}{1 - \exp(-\lambda r)}, \quad 0 < t \leq r, \quad (10)$$

$$f_{W_0}(t) \approx \frac{\lambda \exp(-\lambda(r-t))}{1 - \exp(-\lambda r)}, \quad 0 < t \leq r, \quad (11)$$

and the expectation of W_0 is:

$$\begin{aligned} E(W_0) &\approx \frac{1}{1 - \exp(-\lambda r)} \left(r - \frac{1}{\lambda} + \frac{1}{\lambda} \exp(-\lambda r) \right) \\ &= \frac{r}{1 - \exp(-\lambda r)} - \frac{1}{\lambda} \end{aligned} \quad (12)$$

Equation (10) is only strictly true for the first hop W_1 , which is proved in [11]. In fact, for other hops W_i , $i = 2, 3, \dots, N$, the distribution of $r - W_i$ is truncated at $r - W_{i-1}$. However, this dependence is negligible for large n (i.e., large λ). Therefore, in a Poisson point process with large λ , we consider the i th hop, where $i \leq N$, the distance between the next hop relay to the radio radius limit, denoted by $T_i = r - W_i$, has an exponential distribution with parameter λ , in particular, T_i is less than or equal to radio transmission range r , thus it is a truncated exponential distribution, with probability distribution:

$$F_T(t) \approx \frac{1 - \exp(-\lambda t)}{1 - \exp(-\lambda r)}, \quad 0 < t \leq r \quad (13)$$

which leads to equations (10), (11), and (12) follows.

For $r \ll N/\lambda$, the expected length of the last hop will be half the expected length of the other hops, since the destination will be approximately uniformly distributed within the maximum possible range of the last hop. Thus

$$E(NW_0 + W_0/2) \approx E(L). \quad (14)$$

Lemma 2 implies that it is reasonable and accurate to replace $E(L|C)$ with $E(L)$ when computing $E(N)$ via (1) for dense networks. The basic idea of above lemmas is follow: in a dense network that all the nodes forming Poisson process, it is highly possible to be connected. So we can use the network span $X_{(n)} - X_{(1)}$, to approximate the coverage of a connected network, $E(L)$ can serve as an accurate approximation of $E(L|C)$. Therefore, combining (14) with (2) and (12) gives

$$E(N) \approx \frac{d - 2/\lambda + 2 \exp(-\lambda)/\lambda}{r/(1 - \exp(-\lambda)) - 1/\lambda} - \frac{1}{2} \approx \frac{d - 2/\lambda}{r - 1/\lambda} - \frac{1}{2} \quad (15)$$

neglecting terms of $O(e^{-\lambda})$.

IV. DISTRIBUTION OF N

In this section, we propose an approach to compute the distribution of N numerically. In order to derive the distribution of N , note that the following proposition is a straightforward result of the law of total probability.

Proposition 1: Consider sensor nodes randomly deployed according to Poisson process with parameter λ in 1-D space. Let n be the Poisson distributed number of nodes in the network terrain $[0, d]$. The probability mass function of cardinality of MCDS of this network, $P(N = k - 1)$, is given by:

$$P(N = k - 1) = \int_0^d P(N = k - 1 | L = l) f_L(l) dl. \quad (16)$$

The following proposition derives the probability that $N = k - 1$ given that the network coverage is l , $P(N = k - 1 | L = l)$.

Proposition 2: In 1-D space, if all nodes are randomly deployed forming a Poisson process with parameter λ , then the conditional probability that the size of the minimum connected dominating set is $k - 1$, given the network coverage l , is:

$$P(N = k - 1 | L = l) = P(D_{k-1} < l) - P(D_k < l), 0 < l < d \quad (17)$$

where $D_k = \sum_{i=1}^k W_i$, is the sum of k approximately IID RV's W , with probability density given by (11).

Proposition 2 follows because the event $D_{k-1} < l$ is equivalent to the union of two mutually exclusive events: $D_k < l$, or $D_{k-1} < l$ while $D_k \geq l$.

We compute $P(D_{k-1} < l)$ by numerical inversion of Laplace transform.

The Laplace transform of (11) for $r = 1$ is

$$\mathcal{L}(f_W(t)) = \frac{\lambda \exp(-\lambda)}{1 - \exp(-\lambda)} \left[\frac{\exp(\lambda - s) - 1}{\lambda - s} \right]. \quad (18)$$

As the pdf of D_k is the k fold convolution of $f_W(t)$, by transform method, we can obtain the Laplace transform of pdf of D_{k-1} , which denote as $\mathcal{L}(D_{k-1})$ is:

$$\mathcal{L}(D_{k-1}) = \left[\frac{\lambda \exp(-\lambda)}{1 - \exp(-\lambda)} \right]^{k-1} \left[\frac{\exp(\lambda - s) - 1}{\lambda - s} \right]^{k-1}. \quad (19)$$

By numerical inversion (19), we can obtain pdf of D_{k-1} and D_k , denote as $f_{D_{k-1}}(x)$ and $f_{D_k}(x)$, which lead to $P(D_{k-1} < l)$ and $P(D_k < l)$ where the probability that $D_{k-1} = l$ and $D_k = l$ are ignored as they are zero for continuous distributed nodes.

Combining (4), (17) and (19), by (16), we can numerically compute the probability mass function of N .

For networks with uniformly distributed nodes, if n is large, we adopt Poisson approximation yielding $n \approx E(n) = \lambda d$, which enables the above approach to compute the distribution of N .

V. SIMULATION, NUMERICAL RESULTS AND MODEL VALIDATION

To verify our preliminary assumptions that W 's are approximately independent of each other in dense sensor networks, as well as are independent of N , we conducted 10^5 Monte Carlo simulation sessions for $n = 25, 30, \dots, 500$, in a network with all nodes uniformly distributed within $[0, 12]$ and $r = 1$, and obtain the values of W 's and N . The Correlation Coefficients (CC) between different W 's, and between W 's and N are computed. The results are shown in Figure 2 and Figure 3. It is clear in Figure 2 that the correlation coefficients between different W 's are close to zero as the number of nodes n increasing, as we expected.

In Figure 3, the W 's are negatively correlated with N , and the correlation coefficients become smaller in magnitude as n increases but remain negative. That is to be expected, as N is inversely proportional to the sample mean of W . The magnitude of correlation coefficient between N and W_3 increases slightly for $n < 150$ before quickly increasing. This can be understood as follows. When n is small, the correlation

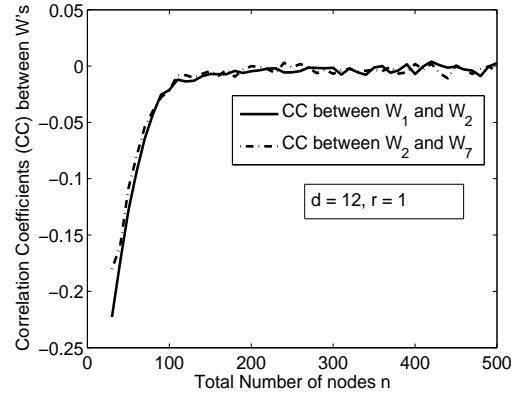


Fig. 2. Simulation results for Correlation Coefficients (CC) between W 's

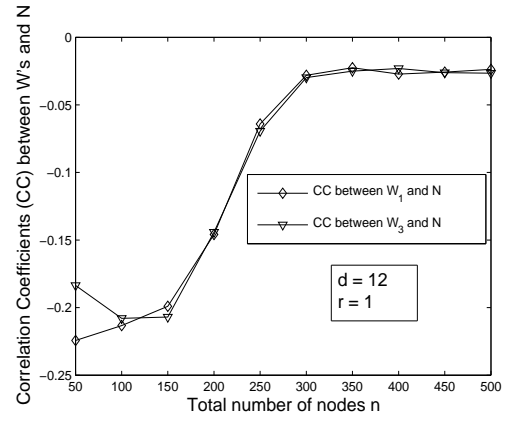


Fig. 3. Simulation results for Correlation Coefficients (CC) between W 's

between the W 's and N is high, however, the variance of the W 's and N are also large, which produces a relatively small correlation coefficient. As n increases, strong correlation between the W 's and N still exists, while the variances of the W 's and N become small. This results in large correlation coefficients. As n increases further, the dependence between the W 's and N decreases, resulting in correlation close to zero. These figures verify that the approximation of the W 's being independent of each other and of N is valid for large n .

To illustrate Lemma 2, Figure 4 compares the CDF of the network coverage given by equation (5) and the network coverage conditioning on the network connectivity obtained by simulation, for a network that all nodes are uniformly distributed within interval $[0, 12]$, other parameters are indicated in the graph. The solid lines in the graph represent the CDF obtain from (5), while the markers indicate the simulation results. We can see from the figure that for a network with normalized total distance $d = 12$, $r = 1$, results given by (5) shows good agreement with simulation results when the number of nodes n is high, which is as expected.

The algorithm to do the numerical inversion of (19) is based on [12]. To demonstrate the accuracy of our approaches to obtain the mean and the pmf of N , we conducted 10^6 Monte Carlo simulation sessions for $d = 12$, $n = 60, 100, 140, 180, 200$, obtain the expected value of minimum

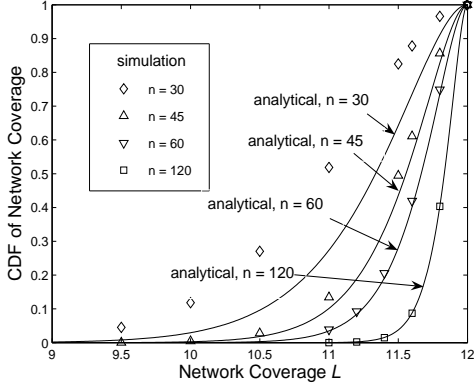


Fig. 4. Comparison of analytical and Simulation results for Cumulative Distribution Function of the network coverage

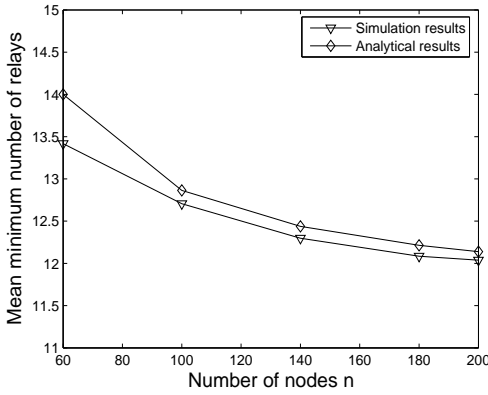


Fig. 5. Comparison of analytical and simulation results for expected minimum number of relays

number of relays nodes that maintain the whole network coverage, comparing with the analytical results calculated by (1). It can be seen that good agreement is achieved between simulation and analytical results, shown in Figure 5. More importantly, as shown in Figure 5, our analytical results provide conservative estimates for the expected number of N .

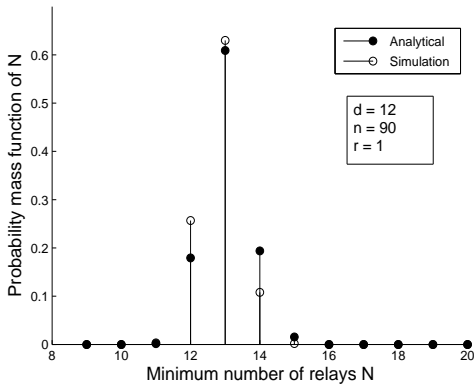


Fig. 6. Example for probability mass function of N

As the number of nodes n becomes large, it is with increasing probability that the minimum number of nodes is

equal to the ceiling of d/r , i.e., $E[N] \rightarrow \lceil d/r \rceil$ as $n \rightarrow \infty$, as the example obtained from simulation shown in Figure 5. Figure 6 compares the simulation results of probability mass function of N for an sensor network with uniformly distributed nodes and analytical result of probability mass function of N , given by (16). For $n = 90$, $d = 12$ and $r = 1$, 10^5 Monte Carlo simulation yields the empirical pmf of N . Filled circles represent analytical results while empty circles for simulation results. This example illustrates that our approach provides an acceptable accuracy level.

VI. EXTENSION TO 2-D ANALYSIS

Obtaining the MCDS and evaluating its cardinality in a 2-D network are more difficult as there are then multiple routes available to the destination. However, our approach serves as a stepping stone towards understanding 2-D networks, and suggests a potential way to evaluate MCDS in 2-D networks, especially for dense networks. Moreover, even though the optimal 2-D case asymptotes to parallel 1-D backbones in a first dimension and a small number of 1-D ribs in the other dimension [7], the results in this paper will require extension before they can be applied to even dense 2-D networks, because the maximum spacing allowed between the ribs depends on how well the relays on neighbouring ribs align. This is the subject of ongoing research.

VII. CONCLUSION

In this paper, we show that for a 1-D sensor network with n uniformly distributed nodes, the cardinality of the minimum connected dominating set can be evaluate probabilistic. The classic result that a point process with all n points uniformly distributed can be well approximated by Poisson process as n becomes large is adopted here, which enables our results to be valid for networks with nodes have either uniform or Poisson distribution.

The expected value and distribution of N obtained from our approaches are important measures to evaluate existing and potential routing protocols, especially for strictly energy constraint sensor networks.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council. The authors thank Iradj Oveysi for helpful comments and Taka Sakurai for discussions about algorithms for numerical inversion of Laplace transform.

REFERENCES

- [1] S. Guha and S. Khuller, "Connected facility location problems," in *Network Design: Connectivity and Facilities Location*, ser. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, P. M. Pardalos and D. Du, Eds. American Mathematical Society, 1998, vol. 40, pp. 179 – 190.
- [2] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: W. H. Freeman, 1979.
- [3] B. Das and V. Bharghavan, "Routing in ad-hoc networks using minimum connected dominating sets," in *Proceedings of 1997 IEEE International Conference on Communication*, 1997, pp. 376 –380.

- [4] J. Wu, M. Gao, and I. Stojmenovic, "On calculating power-aware connected dominating sets for efficient routing in ad hoc wireless networks," in *Proceedings of the 2001 International Conference of Parallel Processing*, 2001, pp. 346–356.
- [5] T. Hsieh, "Using sensor networks for highway and traffic applications," *IEEE Potentials*, vol. 23, no. 2, pp. 13 – 16, Apr - May 2004.
- [6] M. Nekovee, "Sensor networks on the road: the promises and challenges of vehicular adhoc networks and vehicular grids," in *Proceedings of the Workshop on Ubiquitous Computing and e-Research*, Edinburgh, UK, May 2005.
- [7] X. Bai, K. Santosh, X. Dong, Y. Ziqiu, and H. L. Ten, "Deploying wireless sensors to achieve both coverage and connectivity," in *Proc. of the 17th ACM International Symposium on Mobile ad hoc networking and computing MobiHoc'06*, May 2006.
- [8] I. Stojmenovic and X. Lin, "Gedir: Loop-free location based routing in wireless networks." in *Proceedings of the IASTED Int. Conf. on Parallel and Distributed Computing and Systems.*, Boston, USA, November 1999.
- [9] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. John Wiley & Sons, Inc, 1971, vol. 2.
- [10] Y. Cheng and T. Robertazzi, "Critical connectivity phenomena in multihop radio models," *IEEE Trans. Commun.*, vol. 37, July 1989.
- [11] T. Hou and V. Li, "Performance analysis of routing strategies in multihop packet radio network," in *Proc. GLOBECOM'84*, vol. 1, Atlanta, GA, Nov 1984, pp. 487–492.
- [12] J. Abate and W. Whitt, "Numerical inversion of Laplace transforms of probability distributions," *ORSA Journal on Computing*, vol. 7, no. 1, 1995.