Source Node Location Estimation in Large-scale Wireless Sensor Networks

Dmitriy Penkin†, Gerard Janssen* and Alexander Yarovoy†
†*Microwave Technology and Systems for Radar Group, Delft University of Technology, The Netherlands*
Email: D.Penkin@tudelft.nl; A.Yarovoy@irctr.tudelft.nl
**Wireless and Mobile Communications Group, Delft University of Technology, The Netherlands*
Email: G.J.M.Janssen@tudelft.nl

***Abstract*—An algorithm suitable to estimate the source node location in a large-scale wireless sensor network without relying on a particular network infrastructure is developed based on a discrete percolation approach. The algorithm's applicability is evaluated in terms of confidence intervals for the Gaussian distribution of hop count in a peripheral channel and cooperative routing effect is exploited to improve the localization approach.**

***Index Terms*—Large-scale wireless sensor network, localization algorithm, lattice structure, percolation, cooperative effect.**

I. Introduction

Due to the achievements in digital circuit integration and micro-electromechanical systems, small and autonomous sensor nodes capable of environmental sensing, data processing and communicating with each other over short distances via radiofrequency (RF) signals, will be soon brought in the real world [1]. Therefore, the large-scale wireless sensor network (LWSN), composed of a multitude of such nodes deployed in a random manner, is a subject of intensive research. Such networks might find a wide spectrum of breakthrough applications in the field of environmental sensing and monitoring [2]. Since communication and collaboration between miniature nodes should be minimized due to the very limited capabilities of these nodes, the LWSN requires innovative solutions for a number of problems at networking level. In particular, the location estimation of a source node in LWSN is further addressed here because in many environmental sensing and alarm applications, such as intrusion detection, water quality monitoring and biochemical surveillance, the sensed data are pointless without knowing the local area of its origin.

Two generic localization approaches are widely used in a WSN: individual node localization based on signal time difference of arrivals to external reference points or cooperative localization with a central processing unit and knowledge of the absolute positions of either few external (beacons) or internal (anchors) nodes. The fact that the approach taken by GPS is unacceptable in terms of energy consumption, node dimensions and low precision for close range led scientists to focus on GPS-free localization in a LWSN. However, the present GPS-free localization techniques (e.g. [3]–[5]) typically need energy for performing the prerequisite self-organization stage (i.e., nodes reconnoitre their surroundings to form a network topology) and, thus, they can hardly be employed on a LWSN due to very modest node processing capabilities and limited energy resources [6]. In this context, the major novelty of this contribution is to develop the framework suitable to estimate the location of a source node without relying on a particular positioning infrastructure as well as minimizing computation and communication costs by avoiding the self-organization step. This framework is based on simplification of a network topology to a regular lattice structure and uses time differences of last arriving signals to the network borders in order to localize a signal origin.

The paper is organized as follows. The concept of discrete percolation and modelling of LWSN topology using this concept are discussed in Section II, meanwhile the algorithm to localize a source node in such a network is proposed in Section III. The algorithm's applicability is investigated from a statistical perspective in Section IV, whereas cooperative routing is exploited in the Section V to refine the algorithm proposed. Finally, the paper concludes in the Section VI.

II. System Model

Before discussing the localization algorithm in detail, first the system model is outlined. Although the localization approach can be exploited in the 3-dimensional space, the LWSN area is assumed to be limited to the 2-dimensional. Due to very limited energy available, the small node is only going to transmit a signal either when it is triggered by its sensor or when a signal is received from its adjacent nodes. As the directional broadcast brings down the efficiency of a LWSN [7], a sensor node is assumed to be equipped with a short-range RF transceiver radiating omnidirectionally (with coverage radius R) to provide such a large-scale system with the best performance. Furthermore, the signal originating within a LWSN can be detected and processed by the control unit only if it reaches any of the outward nodes (i.e., the nodes situated at the borders of the deployment area). Last but not least, we focus our analysis on static LWSNs, since nodes do not move over significant distances during propagation time in most dynamic scenarios. Also, we suppose that there is no inter-node interference (e.g. it can be omitted by applying particular routing protocols).

Our localization algorithm is based on the concept of the discrete percolation [8]: the priori assumption is that a rectangular deployment area is represented by a regular lattice structure with each site either occupied by a node with a given occupation probability P_{occ} independent of its neighbours. The value of P_{occ} is typically determined in terms of a lattice type, the density of sensor nodes and the lattice spacing D. Since a dipole-like antenna will most likely be used for communication between tiny devices (e.g., [9]), the spacing D should be less or equal to $R/\sqrt{2}$ to compensate for the random attenuation caused by a polarization mismatch between node antennas [10]. Therefore, to meet the above requirement, the LWSN topology is further modelled here by the planar structure with $D = R/\sqrt{2}$ and the number of adjacent, i.e., potentially interconnected, sites $m = 8$ (Fig.1).

Evidently, when the value P_{occ} is small, there is a sparse population of occupied sites, and clusters of small numbers of these sites predominate in the lattice. However, by increasing P_{occ} more occupied sites become interconnected and thus form a single cluster. Finally, when P_{occ} matches the percolation threshold P_c, the structure experiences a second-order phase transition: i.e., the cluster of interconnected sites, spanning the lattice from border to border, occurs for the first time (this cluster is also called percolating). In this way, in order to safely reach the lattice borders to be detected by the control unit, the signal has only to originate and is conveyed via a number of node-to-node transmissions within the spanning cluster. Therefore, the threshold value P_c and the size of the spanning cluster are of interest: the former delimits connected from disconnected structures, whereas the latter gives a probability that a source node belongs to the percolating cluster. For the selected lattice, we have previously studied these parameters in [11]: in particular, the spanning cluster only exists in such a structure when $P_{occ} > P_c = 0.40725 \pm 0.0003$.

The signal propagating in a wave-front manner within the percolating cluster can reach the same network border via different multihop paths. Each of these paths will be of a different number of hops (or hop count) and thus take a different amount of time to traverse. Apparently, the last detected signal at the lattice border is transferred via the channel composed of sites located at the hull of the spanning cluster, i.e., the peripheral path (Fig.2a). Since the percolating cluster is typically characterized by a fractal geometry [12], the major parameter that affects the hop count of peripheral channel h with increasing lattice dimension N is the fractal dimension D, also known as the Hausdorff dimension [11] ($h \sim D^N$). Since this property is only applicable for a peripheral path, time differences of last arriving signals to the network borders are used to derive below the algorithm to estimate a source node location.

III. Location Estimation of a Source Node

Once a signal originates in a percolating cluster, its spreading will only be terminated by the lattice borders. Let us assume that the absolute time of last arriving signals at the left, right, top and bottom borders are denoted as T_l, T_r, T_t and T_b, respectively. Without loss of generality, we may also suppose that $T_l \geqslant T_r$ and $T_t \geqslant T_b$. As the delay time of a sensor node is commonly much larger than the actual signal propagation, these time values can be given as the product of the pre-specified signal delay time in each node T and the hop count of the corresponding peripheral path (with respective suffix letters - see in Fig.2b). In [11], we have shown that such a hop count can be considered as a normally distributed value defined by its mean and variance obtained as functions of the occupation probability P_{occ} and the lattice dimension N. In this way, the delay times $\Delta t_i = T_l - T_r$ and $\Delta t_j = T_t - T_b$ are distributed by the normal law (i.e., a resultant of two normally distributed variables is also normally distributed) and their mean values $\langle \Delta t_i \rangle$, $\langle \Delta t_j \rangle$ as well as the standard deviations $\{\Delta t_i\}$, $\{\Delta t_j\}$ can be expressed as follows:

$$\begin{aligned} \langle \Delta t_i \rangle &= \langle T_l \rangle - \langle T_r \rangle = T[\langle h_l \rangle - \langle h_r \rangle] = \\ &= TC_h[i^{D_h} - (N-i)^{D_h}], \\ \langle \Delta t_j \rangle &= \langle T_t \rangle - \langle T_b \rangle = T[\langle h_t \rangle - \langle h_b \rangle] = \\ &= TC_h[j^{D_h} - (N-j)^{D_h}]. \end{aligned} \tag{1}$$

$$\begin{aligned} \{\Delta t_i\} &= \sqrt{\{T_l\}^2 + \{T_r\}^2} = T\sqrt{\{h_l\}^2 + \{h_r\}^2} = \\ &= TC_\sigma \sqrt{i^{2D_\sigma} + (N-i)^{2D_\sigma}}, \\ \{\Delta t_j\} &= \sqrt{\{T_t\}^2 + \{T_b\}^2} = T\sqrt{\{h_t\}^2 + \{h_b\}^2} = \\ &= TC_\sigma \sqrt{j^{2D_\sigma} + (N-j)^{2D_\sigma}}, \end{aligned} \tag{2}$$

where effective amplitudes C_h,C_σ and the fractal dimensions D_h,D_σ being only functions of the occupation probability have been determined for the sequence of P_{occ} in [11]. In this context, the source site (i, j) can be predicted

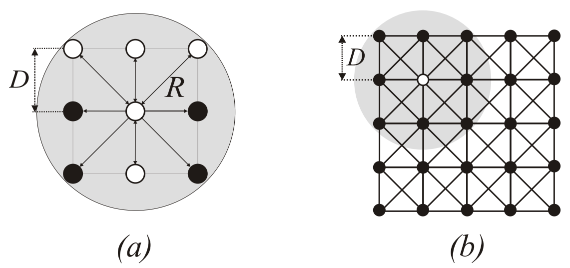

Fig. 1. *(a)* The sketch of the node coverage when $R = \sqrt{2}D$ and *b)* the 2-dimensional lattice structure with a number of adjacent sites $m = 8$. Note that sites occupied with nodes are drawn in white, blank sites are shown in black, whereas the node coverages are indicated in grey.

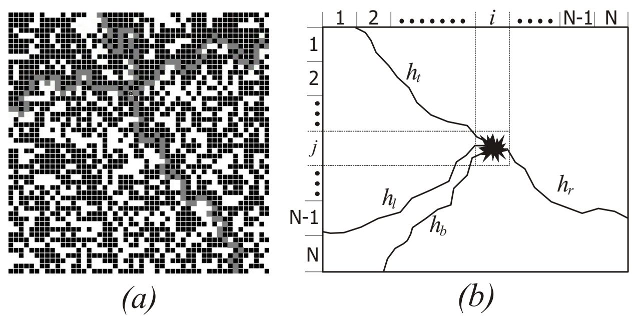

Fig. 2. *(a)* The peripheral paths from the source node in the lattice of 50×50 (occupied sites are shown in white, empty sites are illustrated in black); *(b)* the outline of peripheral paths distributing in a LWSN, which is simplified by a lattice of dimension $N \times N$ with a source node situated in (i, j) site.

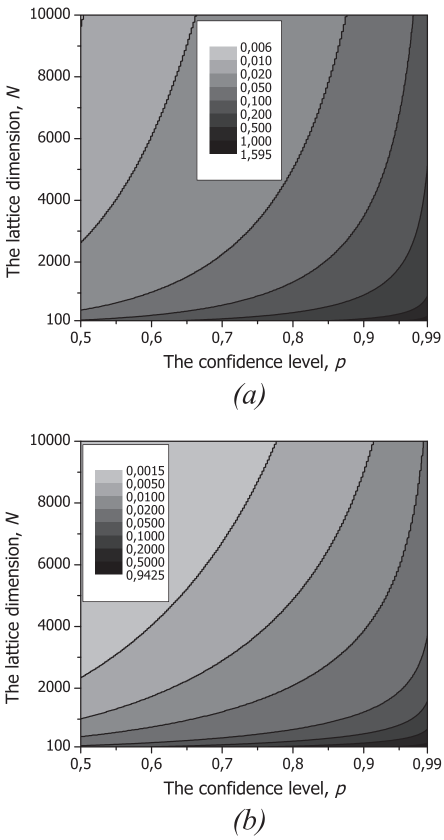

Fig. 3. The accuracy of source node localization Δ (i.e., the ratio of the CA to the whole lattice area) as a function of the lattice dimension N and the confidence level p for the cases when *(a)* $P_{occ} = 0.45$ and *(b)* $P_{occ} = 0.5$.

by measuring the delay times, as mentioned above, and solving Eq.1. Since Δt_i and Δt_j are normally distributed, we further apply quantiles of the Gaussian distribution of hop count in a peripheral path [13] to test the applicability of the localization approach by comparing a confidence area (CA) of signal origin to the entire lattice area.

IV. Performance of Localization Approach

From Eq.2 by taking $d\{\Delta t_i\}/di = 0$ and $d\{\Delta t_j\}/dj = 0$ (variables i and j are assumed to be continuous), it follows that the maximum value of both the standard deviations is equal to as $\{\Delta t_i\}^{max} = \{\Delta t_j\}^{max} = TC_\sigma N^{D_\sigma}$. Hence, being unaware of where the source node would appear the calculations are further performed by assuming that the standard deviations always match this upper bound and, thus, can be given *in hops* as follows: $\{\Delta t\}^{hops} = \{\Delta t_{i(j)}\}^{max}/T = C_\sigma N^{D_\sigma}$.

As was demonstrated in [11], the hop count $h = C_h N^{D_h}$ is required to pass through a lattice of dimension N. Hereby, to get the next lattice column (or row), a signal, on average, has to make $\Delta h = C_h N^{D_h}/N$ hops. In this context, the standard deviation *in sites* can be expressed as:

$$\{\Delta t\}^{sites} = \frac{\{\Delta t\}^{hops}}{\Delta h} = \frac{C_\sigma N^{D_\sigma}}{C_h N^{D_h}} N. \qquad (3)$$

The accuracy of source node localization (i.e., the ratio between the CA and the whole lattice area) can be addressed in terms of confidence intervals for normal distribution of hop count h. Evidently, that such a ratio is equivalent to a number of sites Q, that belongs to the CA, divided by the total number of sites in the lattice N^2. The value of Q can be determined by squaring the length of confidence interval $Z_p\{\Delta t\}^{sites}$, where Z_p is the normal distribution quantile function of the confidence level p. In this way, using Eq.3 the ratio between the CA and the entire lattice area is derived as follows:

$$\Delta = \frac{[Z_p\{\Delta t\}^{sites}]^2}{N^2} = \left[Z_p \frac{C_\sigma N^{D_\sigma}}{C_h N^{D_h}}\right]^2. \qquad (4)$$

From the Eq.4 it is clear that our localization technique is limited to the cases when the fractal dimension D_σ is smaller than the value of D_h, otherwise, the algorithm becomes inaccurate as the value of Δ would diverge with increasing N. This requirement puts a limit on the occupation probability, such as $P_{occ} \gtrsim 0.42$ (based on the results in [11]). Meanwhile the performance of the localization approach is reasonably affordable when P_{occ} is close to the above value (Fig.3a), the algorithm accuracy noticeably improves for the cases when P_{occ} is moderately larger than 0.42 due to the fast decrease of D_σ with increasing the value of P_{occ} (Fig.3b).

Furthermore, if the the probability to find the minimum hop count path between two nodes in the lattice P' is almost the unity, the other technique which relies on the difference in arrival times of the first signals can be used to estimate the source node location. For example, when $P' \cong 1$, the signal in Fig.2 firstly reaches the left lattice border in $i-1$ hops, whereas the right one in $N-i$ hops. Based on this information and knowing the delay difference between the first signal receptions at these borders, the number i is able to be simply calculated. In this context, the probability P' has been under consideration: in particular, this value is shown through a set of numerical experiments to be almost the unity only when the lattice size $N > 50$ and the occupation probability is larger than 0.58. In other words, the second algorithm is merely applicable to localize the source node in the large-scale network with $P_{occ} > 0.58$.

V. On Cooperative Routing in LWSN

As a consequence of signal wave-front spreading, the cooperative routing effect [14] coming from arbitrary sets of synchronized RF nodes enables larger hops in a LWSN. This leads to smaller hop counts in the network and, thus, affects the

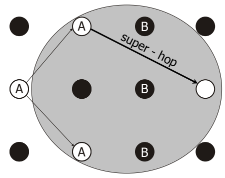

Fig. 4. The scenario when a LWSN is connected due to cooperative routing solely. Sites occupied with nodes are drawn in white, empty sites are shown in black and the coverage of two synchronized nodes is indicated in grey.

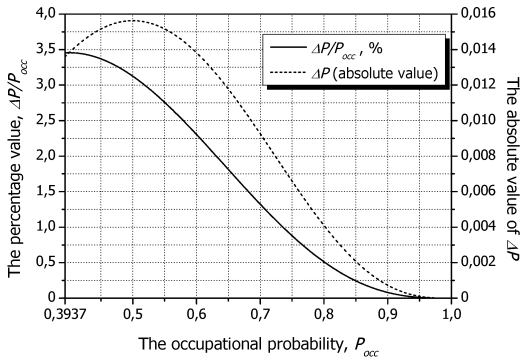

Fig. 5. The percentage and absolute value of the probability ΔP_{ref} as a function of the occupation probability P_{occ}.

localization approach, as discussed above. Since the probability to get three or more nodes synchronized is negligible compared to the probability of having a pair of such nodes, we only study the impact of cooperative routing of two RF nodes.

From the perspective of electrodynamics, the pair of synchronized nodes is regarded as the elementary antenna array consisting of two elements. Since the coverage of such an array is larger than that of a single node, the scenario when the lattice is connected due to cooperative routing only, i.e., a super-hop over a row/column is available, is of interest and under evaluation (Fig.4). Also, based on the underlying electrodynamics principles [15] it was proved that the super-hop over two (or more) rows/columns is infeasible in the selected lattice (i.e., the scenario, as mentioned above, is unique).

By taking into account cooperative routing effect, a hop count will be obviously reduced. Such a decrease is further estimated in terms of the effective occupation probability $P_{occ}^{eff} = P_{occ} + \Delta P$, where ΔP is the probability to maintain the connectivity in the lattice due to only super-hops. As shown in Fig.4, the latter condition comes true only when the sites of type A are occupied whereas the sites of type B are blank. In this way, the value of ΔP can be derived as $\Delta P = P_{occ}^3(1-P_{occ})^3$ and, in particular, its percentage and absolute value as a function of P_{occ} is plotted in the Fig.5.

Eventually, being able to calculate the value of P_{occ}^{eff}, we can estimate the new hop counts in the lattice by using the corresponding effective amplitude $C(P_{occ}^{eff})$ and the fractal dimension $D(P_{occ}^{eff})$ instead of the initial $C(P_{occ})$ and $D(P_{occ})$, respectively. In this regard, cooperative routing effect is of particular interest for the probability P_{occ} being in the vicinity of the percolation threshold P_c as the fractal dimension D vastly decreases with slightly increasing P_{occ}.

Last but not least, by taking into account cooperative routing effect the percolation threshold of the selected lattice is also changed and obtained to be equal to $P_c^{rev} = 0,3937$ (Fig.5). And this result matches well with the data gained by us from numerical simulations using the Monte Carlo method.

VI. Conclusion

The approach suitable to estimate the location of a source node without relying on a particular positioning infrastructure as well as minimizing computation and communication costs by avoiding the self-organization step has been developed. This approach is particularly based on simplification of a network topology to a regular lattice structure and uses time differences of last arriving signals to the network borders in order to localize a signal origin. In this respect, the opportune structure for modelling the 2-dimensional RF network has been selected. For this lattice, the threshold value that delimits connected from disconnected structures has been shown as well as the way to find a signal origin through knowing time differences of last arriving signals to the network borders has been demonstrated. The accuracy of such a source node localization has been addressed in terms of confidence intervals for the Gaussian distribution of hop count in a peripheral path: in particular, it has been estimated that the algorithm, as mentioned above, can only be applied if the occupation probability is larger than 0.42. Also, the localization technique which relies on the delay between times of the first signal reception at opposite lattice borders has been explored: e.g. it has been found that such a technique is only able to be used when $P_{occ} > 0.58$. Furthermore, the impact of cooperative routing effect on the algorithm performance has been addressed: in particular, the way to calculate the refined hop counts and the percolation threshold has been demonstrated.

References

[1] I. F. Akyildiz and J. M. Jornet, "Electromagnetic wireless nanosensor networks," *Nano Communication Networks*, vol. 1, no. 1, pp. 3 – 19, 2010.

[2] M. Ilyas and I. Mahgoub, *Smart Dust: Sensor Network Applications, Architecture and Design*. CRC Press, 2006.

[3] G. Mao, B. Fidan, and B. D. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, no. 10, pp. 2529 – 2553, 2007.

[4] S. Capkun, M. Hamdi, and J.-P. Hubaux, "Gps-free positioning in mobile ad-hoc networks," in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, jan. 2001, p. 10 pp.

[5] L. Hu and D. Evans, "Localization for mobile sensor networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*, ser. MobiCom '04, 2004, pp. 45–57.

[6] S. Roundy, D. Steingart, L. Frechette, P. K. Wright, and J. M. Rabaey, "Power Sources for Wireless Sensor Networks," in *EWSN'04*, pp. 1–17.

[7] C.-C. Shen, Z. Huang, and C. Jaikaeo, "Directional broadcast for mobile ad hoc networks with percolation theory," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 4, pp. 317 – 332, april 2006.

[8] D. Stauffer and A. Aharony, *Introduction to percolation theory*, 2nd ed. Taylor&Francis, London, 1992.

[9] G. Hanson, "Radiation efficiency of nano-radius dipole antennas in the microwave and far-infrared regimes," *Antennas and Propagation Magazine, IEEE*, vol. 50, no. 3, pp. 66 –77, june 2008.

[10] D. Penkin, G. Janssen, and A. Yarovoy, "Feasibility analysis of peer-to-peer microwave communications between self-powered miniature electronic devices," in *Antennas and Propagation (EUCAP), Proceedings of the 5th European Conference on*, april 2011, pp. 122 –125.

[11] D. Penkin, A. Yarovoy, and G. Janssen, "A study on communication aspects of two-dimensional large-scale wireless sensor networks using percolation principles," in *Communications and Vehicular Technology in the Benelux, 2010 17th IEEE Symposium on*, nov. 2010, pp. 1 –6.

[12] J. Feder, *Fractals*, ser. Physics of solids and liquids. Plenum Press, 1988.

[13] R. Cooper and T. Weekes, *Data, models, and statistical analysis*. Barnes & Noble Books, 1983.

[14] A. Khandani, E. Modiano, J. Abounadi, and L. Zheng, "Cooperative routing in wireless networks," in *Advances in Pervasive Computing and Networking*. Springer US, 2005.

[15] J. A. Stratton, *Electromagnetic Theory*. Lightning Source Inc, 2008.