

Relation Between Gradients and Geographic Distances in Dense Sensor Networks with Greedy Message Forwarding

Antonio Caruso
 Department of Mathematics
 University of Salento
 (73100) Lecce, Italy
 Email: antonio.caruso@unisalento.it

Stefano Chessa
 Department of Computer Science
 University of Pisa and ISTI-CNR
 Pisa, Italy
 Email: ste@di.unipi.it

Swades De
 Department of Electrical Engineering
 Indian Institute of Technology
 New Delhi, India
 Email: swadesd@ee.iitd.ac.in

Abstract

The distributed gradient protocol is a common building block to perform several tasks in a wireless sensor network. The gradient calculates the minimum hop-distances between each sensor and a specified set of anchor sensors. This calculation is performed using a distributed greedy forwarding of messages in the network. Several virtual localization protocols use gradients to compute the virtual coordinates of the sensors. The quality of these coordinate systems depends on the relation between the value of the gradient and the real geographic distances between sensors.

In this paper a formal proof of such relation is provided in the case of dense sensor networks with homogeneous sensor communication range r . The minimum-hop distances between a sensor and an anchor is bounded to be in a range defined by two geographic distances. The size of this range decreases with increasing density of the network and it is equal to the maximum resolution (the communication range r) when the density is high enough.

1. Introduction

The effective development of reliable and efficient communication mechanisms in Wireless Sensor Networks (WSN) [1] is particularly challenging. In fact, due to their limited transmission range, sensors must organize into a multihop wireless network and each sensor must be able to relay messages i.e., act as router in order to allow arbitrary point-to-point communications. Given the meagre resources of the sensors (in particular in terms of energy) and their disposable nature, it is not feasible to use other routing protocols (such as those created for ad hoc networks [2], [3]) where addressing hierarchies are created and maintained, since they require network-wide exchange of routes information and a considerable memory overhead to store routing tables in the sensors.

For this reason most recent approaches exploit greedy routing based on the physical position of the destination and

the neighbors of the forwarding sensor. This approach provides reasonable performance in densely, regularly shaped networks but may fail in relatively low densities (4-8 neighbors per sensor) or when obstacles hinder sensor connectivity. In these cases these protocols switch to the face traversal mode [4] (or similar rescue modes) in order to turn around the void area and then to continue with greedy forwarding. The main disadvantage of these approaches is that the sensors must be aware of their physical coordinates, an assumption that is not realistic in some settings (for example indoor), and that requires additional costs since the sensors must be equipped with GPS or other positioning systems.

On the other hand, recent solutions base geographic routing on virtual coordinates rather than on physical coordinates [5],[6],[7],[8],[9],[10],[11]. Virtual coordinates completely disregard sensor positions and take into account only the WSN connectivity to assign coordinates to the sensors.

A common building block used by this class of protocols is the calculation of a *gradient* - an operation in which each sensor estimates its shortest path distance from a set of gradient sources or *anchors*. Gradients are used not only in coordinate assignment but for several applications like data harvesting (directed diffusion [12]) and routing [5], [7].

Gradients are generally calculated through the iterative (distributed) application of a triangle inequality constraint. In its most basic form the calculation of the gradient g_j^x at a sensor x with respect to the anchor j is given by:

$$g_j^x = \begin{cases} 1 & \text{if } x = j \\ \min\{g_j^y + d(x, y) \mid y \in N_x\} & \text{if } x \neq j \end{cases} \quad (1)$$

where N_x is the neighborhood of x (excluding itself) and $d(x, y)$ is the distance between neighboring sensors x and y . Iterated applications of the above calculus converge in each sensor to a stable value.

When the distance function among neighboring sensors is defined as $d(x, y) = 1$ the protocols are called *Range-Free*, i.e., they do not estimate geographic distances among sensors, but, rather, they use only the information implied by the connectivity.

With virtual coordinates, each sensor is assigned a tuple of coordinates, where each coordinate is the gradient with respect to a different anchor. The coordinate assignment is achieved by means of distributed protocols, and the different virtual coordinate systems differ in the number and displacement of the anchors and on small variations in the definition of the gradients. It is important to note that the virtual coordinate vector of a sensor is not necessarily unique. The set of sensors sharing the same virtual coordinates is called a *zone* of the network. Once the virtual coordinate system is established, a geographic routing protocol delivers the packets to their destination zone. Then each packet can be delivered to the desired destination sensor using, for example, a proactive ID-based approach within the zone. For this reason, in order to efficiently support geographic routing, the zones should be small and span a limited number of hops.

The size and shape of zones depend on the number of anchors and their position in the network, and on the algorithm used to compute the gradients. Although it is desirable to have small zones (possibly one zone per sensor), this may be costly to achieve, since it may require a large number of anchors, and thus a higher message complexity of the protocol. Note that, for unambiguous geographic forwarding we require at least three anchors (and thus the resulting coordinate space is three dimensional). In fact, if only two coordinates were used (say x and y), there would exist physically distant sensors, symmetric to the directrix connecting two anchors that would share the same virtual coordinate (see Figure 1 for an example). A third anchor is thus needed to break the symmetry. As a matter of fact any protocol that define a consistent coordinate system uses at least three anchors.

The Virtual Coordinate Assignment Protocol (VCAP) presented in [5] defines the coordinates of each sensor using a variation of gradients where the anchor sensors are elected jointly by means of a distributed protocol. Using this approach it was formally proved that the radius of the zones (in terms of number of hops) is probabilistically bounded by a small constant. The proof is valid in dense sensor networks, but it is based on a conjecture that relate the gradient, i.e., the minimum hop-count distance between a sensor and an anchor, with the geographic distance between the sensor and the anchor, and this conjecture has not been formally proven yet. In this paper we remove such conjecture by providing a formal proof of this relation, i.e., we show in a dense sensor network the minimum hop-distance, defined as in gradients, is geometric probabilistically related to geographic distance. This result applies to any virtual coordinate system constructed by means of gradients (for example to [6] or [11]) since it provide bounds that can be used to determine the maximum size of the zones for dense sensor networks. In particular, it can be directly applied to the virtual coordinate system defined by VCAP [5] since the analysis on the size

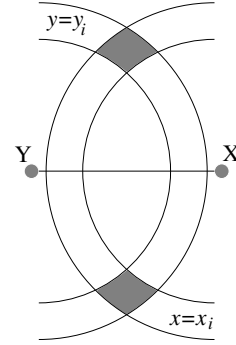


Figure 1: If only two anchors (X, Y) are used, sensors within the gray areas receive the same coordinate (x_i, y_i) .

of the zones depends on this result.

The paper is organized as follow: In Section 2 we review the VCAP protocol and the virtual coordinate system that it defines. In Section 3 we review part of the analysis done in [5] that left as an open problem the relation between the two distances. In Section 4 we present the proof of the above statement. Finally in Section 5 we draw some conclusions.

2. The VCAP Protocol

We consider a sensor network composed of a large number of sensors uniformly scattered in a sensing field. The sensors are assumed static, or they have very low mobility with respect to signal propagation speed. Every sensor has the same transmission range r , and each sensor is able to communicate with all sensors within its range.

The VCAP protocol [5] has four phases, each phase jointly elects an anchor and computes the corresponding gradients (as shown in Figure 2). The overall goal of the protocol is to select three anchors that are near the border of the network (the most external sensors).

In each phase the function that computes the gradients uses also the values of the gradients computed in previous phases. The anchors are denoted with capital letters W, X, Y, Z , the value of the gradient in each sensor with respect to a given anchor is denoted with a small letter corresponding to the anchor identity, i.e., w, x, y, z .

The first anchor W is any sensor in the network, i.e., it could be the sink or any other sensor specifically programmed. After the gradient calculation with respect to W , the *border* of the network is defined as the set of sensors that have the maximum value of w within their two-hop neighborhood. Since this phase is only used to determine the sensors on the border it could be possibly replaced by using any other protocol for the identification of the sensor network borders (see for example [13]).

Each sensor in the border attempts to be elected as X anchor. The election is done jointly with the calculation of the gradient with respect to X . Only messages with a higher

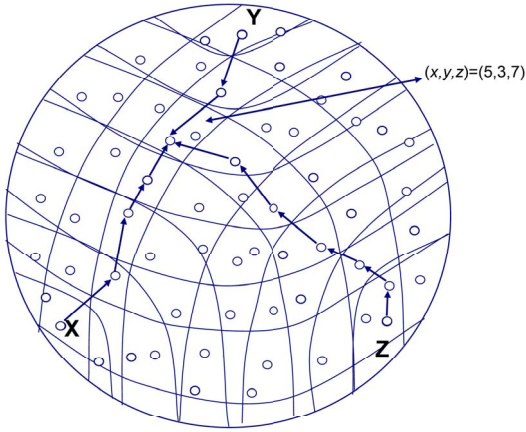


Figure 2: The three anchors elected by VCAP in a circular sensor network.

w (or higher sensor ID) are used to update the value of the gradient for X . At the end of this phase a single sensor on the border of the network is elected as anchor X , and each sensor know the gradient with respect to it.

The third phase is similar to the previous, it elects a sensor Y among the sensors that are on the border and with maximum distance with respect to X , i.e., maximum x . The position of the anchor Y is therefore opposite to X in the network. To avoid a situation in which Y is too close to W (this situation may impair the correct selection of anchor Z), Y is selected with the constraint that its w coordinate is such that $w > \gamma$, where γ is a parameter of the protocol.

The fourth phase involves sensors that satisfy a specific condition on the value of w, x, y . The best choice for Z (formally proved in [5]) would be a sensor orthogonal to the center of the XY directrix. The rule used in the last protocol phase to choose Z is a heuristic aimed at finding a sensor that satisfies this last constraint. In [5] a rule was suggested that selects a sensor with maximum w among the sensors with $x = y \pm 1$.

When the last phase is completed, each sensor is assigned with a coordinate triplet (x, y, z) . The gradient with respect to W is used to let anchors X, Y , and Z be close to the boundary of the network with high probability, it is not used in the definition of the virtual coordinates.

3. Properties of the VCAP protocols

The VCAP protocol [5] causes a partition of the network into zones, where the sensors within a zone are labeled with the same virtual coordinates. The size of a zone depends on the position of the three anchors X, Y , and Z and on the sensor node density, measured as the average number of neighbors per sensor. As already mentioned in Section 2, in order to efficiently support geographic routing, the size of the zones should be bounded.

In [5] we proved that the size of a zone in a sensor network with density Δ is limited. In particular, the Euclidean distance d between two sensors in the same zone is bounded:

Theorem 1. Consider a sensor network deployed in a circular space of diameter \mathcal{D} , and assume that the anchors X, Y , and Z are placed on the vertexes of an equilateral triangle inscribed in the circle of diameter \mathcal{D} . Let r be the communication range of sensors. The maximum distance d between two sensors in the same zone satisfy:

$$\lim_{\Delta \rightarrow \infty} d = \frac{8}{3}r.$$

Although limited to circular domains, the result of Theorem 1 indicates why VCAP selects the anchors as far as possible from each other and on the boundary of the network. In fact this choice contributes to reduce the size of the zones and, in turn, reduce the number of sensors which share the same set of coordinates.

Note that during the distributed calculation of the gradients the sensors forward messages using a greedy approach. This is because, dissemination of the coordinates is such that the message originated by an anchor proceeds via broadcast, and only the message along the shortest path (with the least hop count) is considered at a sensor. This implies that the messages setting the gradients are forwarded at each intermediate hop by the sensors that minimize the distance from the respective anchors. In other words, the message propagation is based on a greedy forwarding approach.

The proof of Theorem 1 in [5] is based on an assumption that relate the hop distance between two sensors when using a greedy forwarding approach with the Euclidean distance between the sensors d . This assumption was verified only by simulations. In this paper we remove the assumptions and present a formal analysis of the above relationship for dense networks. In particular, given two sensors A and B at hop distance h we show that their physical distance can be approximated with high probability as a function of h , and that the error of this approximation tends to be at most r for dense sensor networks. This result is formally stated in Theorem 2 in the next section.

4. Existence of bounds on Euclidean distance

In this section we discuss the relationship between a given hop count and the bounds on Euclidean distance when using a greedy forwarding approach. As shown in Figure 3, the current distance between a sensor A and a destination B is d . $P(t, \theta)$ is a potential neighbor of A and its distance to B is z , which is a random variable.

The greedy approach that we consider chooses as next hop the sensor P that minimizes z . We consider the sensor density to be sufficiently high to avoid any potential backward movement at any step due to network holes.

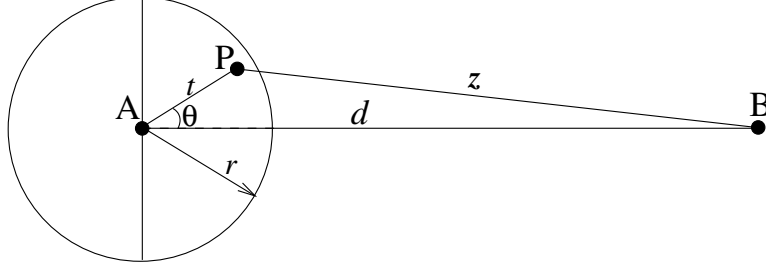


Figure 3: Remaining Euclidean distance to the destination in greedy forwarding approach.

Referring to Figure 3 and considering that a potential forwarding sensor can be anywhere in the right half circle, the joint probability density function (pdf) of the random variables t and θ is

$$f_{t\theta}(t, \theta) = \begin{cases} \frac{2t}{\pi r^2}, & 0 \leq t \leq r \text{ and } -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \\ 0, & \text{elsewhere.} \end{cases}$$

By transformation of variables, the pdf of remaining distance z is obtained as

$$f_z(z) = \begin{cases} \frac{4z}{\pi r^2} \arccos\left(\frac{d^2+z^2-r^2}{2dz}\right), & \text{if } d-r \leq z \leq d \\ \frac{4z}{\pi r^2} \left[\arcsin\left(\frac{d}{z}\right) - \arcsin\left(\frac{d^2+z^2-r^2}{2dz}\right) \right], & \text{if } d \leq z \leq \sqrt{d^2+r^2} \\ 0, & \text{elsewhere} \end{cases}$$

We denote the remaining distances from $\Delta/2$ potential forwarding sensors as $z_1, z_2, \dots, z_{\Delta/2}$, where Δ is the average number of neighbors of a sensor, also called network (sensor) density. Since the sensors are uniformly random distributed, $z_i \forall i = 1$ to $\Delta/2$ are independent and identically distributed (i.i.d.) random variables. Therefore, the pdf of least remaining distance $\xi = \min\{z_1, z_2, \dots, z_{\Delta/2}\}$ is given by

$$f_\xi(\cdot) = \frac{\Delta}{2} f_z(\cdot) R_z^{\frac{\Delta}{2}-1}(\cdot),$$

where R_z is the complementary cumulative distribution function (cdf) of z . The corresponding pdf of forward progress in one hop $\varepsilon = d - \xi$ is

$$f_\varepsilon(\varepsilon) = f_\xi(d - \varepsilon). \quad (2)$$

Let $G(d, \varepsilon, r) = \arccos\left(1 + \frac{\varepsilon^2 - r^2}{2d(d - \varepsilon)}\right)$. Assuming high sensor density such that ε is always positive, the pdf of ε can be approximately expressed as

$$f_\varepsilon(\varepsilon) = \begin{cases} \Delta \left(\frac{2}{\pi r^2}\right)^{\frac{\Delta}{2}} (d - \varepsilon) G(d, \varepsilon, r) \left[\frac{\sqrt{4r^2 d^2 - (r^2 - \varepsilon^2 + 2d\varepsilon)^2}}{2} \right. \\ \left. - (d - \varepsilon)^2 G(d, \varepsilon, r) + r^2 \arcsin\left(\frac{r^2 - \varepsilon^2 + 2d\varepsilon}{2dr}\right) \right]^{\frac{\Delta}{2}-1} & \text{if } 0 \leq \varepsilon \leq r \\ 0, & \text{elsewhere.} \end{cases} \quad (3)$$

The n -th moment of ε can be calculated from (3) via numerical integration, $\bar{\varepsilon}^n = \int_0^r \varepsilon^n f_\varepsilon(\varepsilon) d\varepsilon$, from where the standard deviation of ε , $\sigma = \sqrt{\bar{\varepsilon}^2 - (\bar{\varepsilon})^2}$, is obtained. As shown in Figure 4, especially at large sensor density, σ is largely independent of d . Therefore, we take σ approximately as a constant, and compute it using d as the expected distance between any source-destination pair d_{S-D} . For example, it was shown in [14] that the expected distance between any two randomly located points within a circular disc of diameter \mathcal{D} is $d_{S-D} = \frac{64}{45\pi} \mathcal{D}$.

Theorem 2. Consider two sensors at minimum hop distance h . For any fixed probability p there exist two values $l(h)$ and $u(h)$ such that the Euclidean distance d between the two sensors is bounded in probability, i.e., $l(h) \leq d \leq u(h)$ with probability p . The quality of the bound depends on the network density Δ and the probability p . At sufficiently large network density,

$$\lim_{\Delta \rightarrow \infty} u(h) - l(h) = r,$$

where r is the transmission range of the sensors.

Proof: Assume that the forward progress in each hop is independent of the progress in previous hops. The total progress in first $h - 1$ hops is a random variable, $\varepsilon' = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{h-1}$, where $\varepsilon_i, \forall i = 1$ to $h - 1$, are i.i.d. random variables with mean $\bar{\varepsilon}$ and variance σ^2 . For moderately large network (with $h > 3$), by central limit theorem, ε' is normal distributed, with mean $(h - 1)\bar{\varepsilon}$ and variance $(h - 1)\sigma^2$. Thus, the standard deviation of Euclidean distance progress in $h - 1$ hops is $\sqrt{h - 1}\sigma$.

Since ε' is normal distributed, a proper multiplication factor k can be chosen such that the distance ambiguity $u(h - 1) - l(h - 1)$ in first $h - 1$ hops lies within $2k\sqrt{h - 1}\sigma$ around the mean with probability p , where p is a function of k .

To study the limiting case, we obtain the cdf of one hop

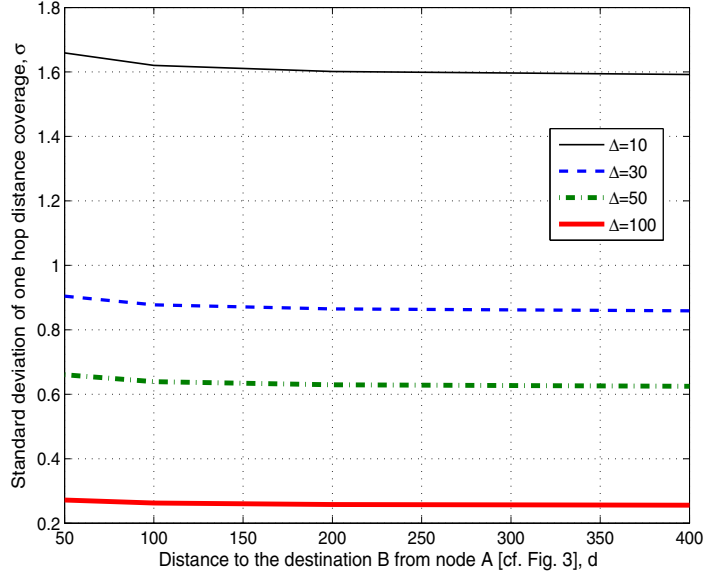


Figure 4: Variation of standard deviation of one hop progress with d . $r = 10$.

progress ε from (2) and (3):

$$F_{\varepsilon}(\varepsilon) = \begin{cases} 0, & \varepsilon < 0 \\ \left[\frac{2}{\pi r^2} \left\{ \frac{\sqrt{4r^2 d^2 - (r^2 - \varepsilon^2 + 2d\varepsilon)^2}}{2} - (d - \varepsilon)^2 \arccos \left(1 + \frac{\varepsilon^2 - r^2}{2d(d - \varepsilon)} \right) + r^2 \arcsin \left(\frac{r^2 - \varepsilon^2 + 2d\varepsilon}{2dr} \right) \right\} \right]^{\frac{1}{2}}, & \text{if } 0 \leq \varepsilon \leq r \\ 1, & \varepsilon > r. \end{cases}$$

Denote

$$x = \frac{2}{\pi r^2} \left\{ \frac{\sqrt{4r^2 d^2 - (r^2 - \varepsilon^2 + 2d\varepsilon)^2}}{2} - (d - \varepsilon)^2 \arccos \left(1 + \frac{\varepsilon^2 - r^2}{2d(d - \varepsilon)} \right) + r^2 \arcsin \left(\frac{r^2 - \varepsilon^2 + 2d\varepsilon}{2dr} \right) \right\}$$

By the property of cdf, for all $\varepsilon < r$ (i.e., $x < 1$), $F_{\varepsilon}(\varepsilon) < 1$, and for all $\varepsilon \geq r$, $F_{\varepsilon}(\varepsilon) = 1$. Since $\lim_{\Delta \rightarrow \infty} x^{\Delta/2} = 0$ for all $x < 1$, the cdf $F_{\varepsilon}(\varepsilon)$ is a step function at $\Delta \rightarrow \infty$; $F_{\varepsilon}(\varepsilon) = 0$ for all $\varepsilon < r$ and $F_{\varepsilon}(\varepsilon) = 1$ for all $\varepsilon \geq r$. Thus, in the limit, the pdf $f_{\varepsilon}(\varepsilon)$ is a shifted delta function: $f_{\varepsilon}(\varepsilon) = \delta(\varepsilon - r)$, and correspondingly, $\varepsilon = r$, a constant. Therefore, the standard deviation of ε , and hence the distance ambiguity $u(h-1) - l(h-1)$ tends to 0 in the limit.

One way to capture the effect of last hop is explained through Figure 5, which shows one dimensional hop counts and distance coverages. Say, a sensor with x -coordinate x_i needs to reach an h -hop away sensor B (with x -coordinate

x_{i+h}). There are two extreme cases involving the two nearly r distance apart sensors, A_1 and A_2 . At very high sensor density, in first $h-1$ hops, the sensor C is reached from sensor A_1 , whereas the sensor B' is reached from sensor A_2 . In both cases, the distance covered is $(h-1)r$ (since $\lim_{\Delta \rightarrow \infty} u(h-1) - l(h-1) = 0$). To reach sensor B, in the first case the distance covered is r (C to B), whereas in the second case the distance is nearly zero (B' to B). Therefore, in total, $\lim_{\Delta \rightarrow \infty} u(h) - l(h) = r$.

Since the distance coverage in the last hop is uniformly distributed in $(0, r]$, the maximum distance ambiguity in the last hop is r , irrespective of the sensor density. Thus, at finite sensor density, $u(h) - l(h) = 2k\sqrt{h-1}\sigma + r$ in probability p , where p is a function of k . \square

5. Conclusions

In this paper we have considered the problem of determining the relationship between the distances calculated with gradient, i.e., the minimum hop distances between a sensor and an anchor calculated by a distributed greedy forwarding algorithm, and the respective geographic distances.

This relationship is of great practical importance since an important class of virtual coordinate systems assign the coordinates to the sensors based on hop distances (i.e., based on gradients range free), thus using this result it is possible to determine the physical distance between sensors of given coordinates. This result can also be exploited in the formal analysis of localization protocols and in other distributed protocols based on gradients. An immediate application of

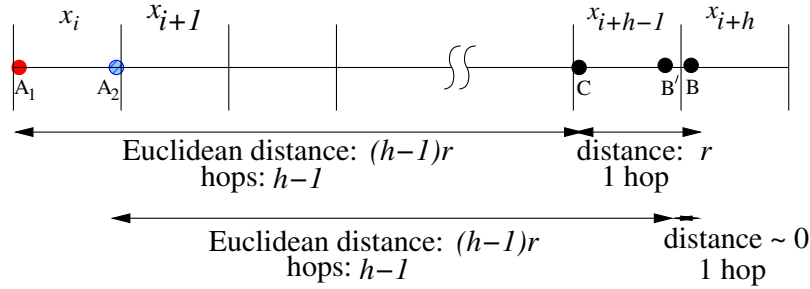


Figure 5: One dimensional schematic showing the maximum difference in distance coverage in h hops at very large sensor density. r is the nodal range. All sensors within two consecutive vertical lines have the same virtual coordinate (hop count).

this result is related to the VCAP protocol. In fact this protocol defines zones (i.e., set of sensors assigned with the same virtual coordinates), and the size of the zones is critical to the performance of routing. The proof given in [5] that defines a bound on the size of the zones is however dependent on an unproven assumption on the relationship between the gradient distances and the geographic distances. The proof given in the Theorem 2 in this paper fills this gap and shows that this relationship is also dependent on the network density. Future work include the extension of such results to network of heterogeneous sensors having different communication and energy capabilities.

References

- [1] P. Baronti, P. Pillai, V. W. C. Chook, S. Chessa, A. Gotta, and Y. F. Hu. Wireless sensor networks: A survey on the state of the art and the 802.15.4 and zigbee standards. *Computer Communications*, 30(7):1655–1695, 2007.
- [2] C. E. Perkins and E. M. Royer. Ad hoc on-demand distance vector routing. In *Proceedings of the 2nd IEEE workshop on mobile computing systems and applications*, pages 90–100, New Orleans, LA, USA, February 1999.
- [3] D.B. Johnson and D.A. Maltz. Dynamic source routing in ad hoc wireless networks. *KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE*, pages 153–179, 1996.
- [4] B. Karp and H. T. Kung. Greedy perimeter stateless routing (gpsr) for wireless networks. In *Proc. sixth annual ACM/IEEE international conference on mobile computing and networking (mobicom)*, pages 243–254, Boston, MA, USA, 2000.
- [5] A. Caruso, S. Chessa, S. De, and A. Urpi. Gps free coordinate assignment and routing in wireless sensor networks. In *Proc. IEEE INFOCOM*, pages 150–160, 13–17 March 2005.
- [6] R. Fonseca, S. Ratnasamy, J. Zhao, C. Tien Ee, D. Culler, S. Shenker, and I. Stoica. Beacon vector routing: scalable point-to-point routing in wireless sensornets. In *Proc. 2nd USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 329–342, May 2005.
- [7] Q. Fang, J. Gao, L. J. Guibas, V. de Silv, and L. Zhang. Glider: Gradient landmark-based distributed routing for sensor networks. In *Proc. IEEE INFOCOM*, March 2005.
- [8] A. Rao, C. Papadimitriou, S. Shenker, and I. Stoica. Geographic routing without location information. In *Proceedings of the 9th annual international conference on mobile computing and networking*, pages 96–108, San Diego, CA, USA, 2003.
- [9] J. Beal, J. Bachrach, D. Vickery, and M. Tobenkin. Fast self-healing gradients. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pages 1969–1975, New York, NY, USA, 2008. ACM.
- [10] J. Bachrach, R. Nagpal, M. Salib, and H. Shrobe. Experimental results for and theoretical analysis of a self-organizing global coordinate system for ad hoc sensor networks. *Telecommunication Systems*, 26(2):213–233, 2004.
- [11] N. Filardi, A. Caruso, and S. Chessa. Virtual naming and geographic routing on wireless sensor networks. In *IEEE Symposium on Computers and Communications (ISCC '07)*, pages 609–614, Aveiro, Portugal, 2007.
- [12] C. Intanagonwivat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva. Directed diffusion for wireless sensor networking. *IEEE/ACM Transactions On Networking*, 11(1), 2003.
- [13] J. S. Deogun, S. Das, H. S. Hamza, and S. Goddard. An algorithm for boundary discovery in wireless sensor networks. In *12th Int. Conf. on High Performance Computing HiPC 2005*, pages 343–352, Goa, India, 2005.
- [14] M. G. Kendall and P. A. P. Moran. *Geometrical Probability*. Griffin, London, 1963.