Range-Free Node Localization in Multi-Hop Wireless Sensor Networks

Sim Zaidi*, Ahmad El Assaf*†, Sofiène Affes*†, and Nahi Kandil*†
*INRS-EMT, Université du Québec, Montreal, QC, H5A 1K6, Canada, Email: {zaidi,elassaf,affes}@emt.inrs.ca
†LRTCS, University of Quebec in Abitibi-Témiscaming, Rouyn-Noranda, QC, J9X 5E4, Canada, Email: nahi.kandil@uqat.ca

***Abstract*—To localize multi-hop wireless sensor networks (WSN) nodes, only the hop-based information (i.e., hops' number, average hop size, etc.) have been so far exploited by range-free techniques, with poor-accuracy, however. In this paper, we show that localization accuracy may greatly benefit from joint exploitation, at no cost, of the information already provided by the forwarding nodes between each anchor (i.e., position-aware) and sensor nodes pair. As such, we develop a novel range-free localization algorithm, derive its average location estimation error (LEE) in closed-form, and compare it in LEE performance with the best representative algorithms in the literature. We show that the proposed algorithm outperforms in accuracy all its counterparts. In contrast to the latter, we further prove that it is able to achieve a LEE average and variance of about 0 when the number of sensors is large enough, thereby achieving an unprecedented accuracy performance among range-free techniques.**

***Index Terms*—Wireless sensor networks (WSNs), multi-hop, localization, low-cost, location estimation error (LEE).**

I. Introduction

Recent advances in wireless communications and low-power circuits technologies have led to proliferation of wireless sensor networks (WSNs). A WSN is a set of small and low-cost sensor nodes often equipped with small batteries. The latter are often deployed in a random fashion to sense or collect from the surrounding environments some physical phenomena such as temperature, light, pressure, etc. [1]. Since power is a scarce resource in such networks, sensors usually recur to multi-hop transmission in order to send their gathered data to an access point (AP). However, the received data at the latter are often fully or partially meaningless if the location from where they have been measured is unknown, making the sensors' localization an essential task in multi-hop WSNs. Designed to comply with such networks, many localization algorithms exist in the literature [2]- [14]. To properly localize each sensor, most of these algorithms require the distance between the latter and at least three position-aware nodes called hereafter anchors. Since it is very likely in multi-hop WSNs that some sensors be unable to directly communicate with all anchors, the distance between each anchor-sensor pair is usually estimated using their shortest path. The latter is obtained by summing the distances between any consecutive intermediate nodes located on the shortest path between the two nodes. Depending on the process used to estimate these distances, localization algorithms may fall into three categories: measurement-based, heuristic, and analytical [2]-[14].

Measurement-based algorithms exploit the measurements of the received signals' characteristics such as the received signal strength (RSS) [2] or the time of arrival (TOA) [3], etc. Using such algorithms, additional hardware is usually required at both anchors and regular nodes [?], thereby increasing the overall cost of the network. Furthermore, in the presence of noise and/or multipath, the TOA measurement is severely affected thereby hindering sensors' localization accuracy.As far as heuristic algorithms [4]-[6] are concerned, most of them are based on variations of the DV-HOP technique [4], whose implementation in multi-hop WSNs requires the computation of the average hop size (i.e., average distance between any two consecutive intermediate nodes). The latter is usually computed in a non-localized manner and broadcasted in the network by each anchor. This incurs undesired prohibitive overhead and power consumption, thereby increasing the overall cost of the localization process.

More popular alternatives suitable for multi-hop WSNs are the analytical algorithms [7]-[14] which evaluate theoretically $h_{\rm av}$ using the statistical characteristics of the network deployment. The obtained $h_{\rm av}$ is actually locally computable at each regular node, thereby avoiding the unnecessary overhead and power consumption incurred by heuristic techniques if, likewise, it had to be broadcasted in the network. In spite of their valuable contributions, the localization algorithms developed so far in [7]-[14] do not provide unfortunately sufficient accuracy, due to large errors occurred when mapping n_h into distance units. This is primarily caused by the lack of information provided by both $h_{\rm av}$ and n_h. Actually, the distance between an anchor-sensor pair depends not only on the latter hop-based information, but also on the number m of forwarding nodes (i.e., which forward any data between the two nodes). Indeed, when n_h and the total nodes' number are fixed, the distance increases (decreases) if m increases (decreases). Consequently, if this easily-obtained information is taken into account when designing a localization algorithm, its accuracy would definitely be improved.

Hence we propose in this paper, a novel analytical localization algorithm that properly exploits m alongside the hop-based information, derive its average location estimation error (LEE) in closed-form, and compare it in LEE performance with the best representative algorithms in the literature. We

978-1-4673-9814-5/16/$31.00 ©2016 IEEE

show that the proposed algorithm outperforms in accuracy all its counterparts. In contrast to the latter, we further prove that it is able to achieve a LEE average and variance of about 0 when the number of sensors is large enough, thereby achieving an unprecedented accuracy performance among range-free techniques.

II. Network model and motivation

Consider the system model of M anchor and N sensor nodes deployed in a 2-D square area S. The anchors are aware of their positions while the sensors are oblivious to this information. These sensors are assumed to be uniformly distributed in S. All anchor and sensor nodes are assumed to have the same range (i.e., transmission capability) denoted by R. Each node is then able to directly communicate with any other node located in the disc having that node as a center and R as a radius, while it communicates in a multi-hop fashion with the nodes located outside it. Let $(x_i, y_i)\,,\ i = 1, \ldots, N$ be the coordinates of the sensors and $(a_k, b_k)\,,\ k = 1, \ldots, M$ be those of the anchors.

In what follows, we propose an efficient anchor-based localization algorithm aiming to accurately estimate the sensors' positions. Such an algorithm requires that the latter estimate their distances to at least 3 anchors and be aware of their coordinates. The k-th anchor should then broadcast its coordinates (a_k, b_k) through the network. If the i-th sensor is located at a distance less than or equal to R from that anchor, it receives the coordinates in $n_h = 1$ hop. Otherwise, it receives them after $n_h > 1$ hops. So far, in most previous algorithms, the i-th sensor estimates its distance to the k-th anchor d_{i-k} using only the information n_h as

$$\hat{d}_{i-k} = n_h h_{\text{av}} \tag{1}$$

where h_{av} is a predefined average hop size. Note that this distance estimation (DE) approach relies on the fact that in highly dense WSNs,

$$d_{i-k} \approx \sum_{l=1}^{n_h} h_l, \tag{2}$$

holds. In (2), h_l is the l-th hop's distance. Unfortunately, this approach exhibits a major drawback. Indeed, h_{av} is usually derived either analytically (i.e., $h_{\text{av}} = \mathrm{E}\{h_l\}$) [7]- [14] or heuristically by computing the mean hop size of all the shortest paths between anchors as in [4]

$$h_{\text{av}} = \frac{1}{M(M-1)} \sum_{k=1}^{M} \sum_{l=1}^{M} \frac{\sqrt{(a_k - a_l)^2 + (b_k - b_l)^2}}{n_{k,l}} \tag{3}$$

where $n_{k,l}$ is the number of hops between the k-th and l-th anchors. It is then very likely that h_{av} be different from the mean hop size of the shortest path between the k-th anchor and the i-th sensor (i.e., $h_{\text{av}} \neq \left(\sum_{l=1}^{n_h} h_l\right)/n_h$ and, hence, large DE errors may occur, thereby hindering the i-th sensor's localization accuracy. This motivates us to seek for more efficient DE approach for exploitation by our localization algorithm.

III. Proposed DE approach

In this work, we propose to exploit, in addition to n_h, another easily obtained information, in order to reduce the distance estimation error, thereby improving the localization accuracy. According to the parity of n_h, we distinguish below between two cases and develop two different approaches suitable for each case.

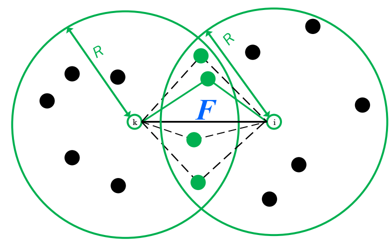


Fig. 1: Two-hop communication.

A. n_h is even

For simplicity, let us first assume that $n_h = 2$. Let $D_k(R)$ and $D_i(R)$ be the discs with radius R and having, respectively, the k-th anchor and the i-th sensor as centers. $F = D_k(R) \cap D_i(R)$ is then the forwarding area wherein the forwarding nodes, which forward the messages sent from the k-th anchor to the i-th sensor, are located. An in depth look at this area reveals that it is strongly dependant on d_{i-k}; a fact that could be exploited to estimate the latter. Indeed, as can be observed from Fig. 1, if d_{i-k} increases (decreases). then F decreases (increases). Using some geometrical properties and trigonometric transformations, one can even show that

$$F = \Phi\,(d) = 2R^2 \cos^{-1}\left(\frac{d_{ki}}{2R}\right) - \frac{1}{2} d_{ki} \sqrt{4R^2 - d_{ki}^2}. \tag{4}$$

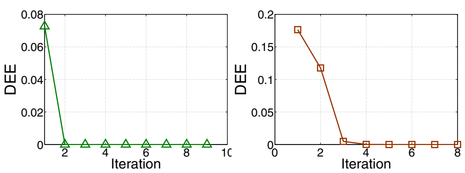
It follows from (4) that $\Phi\,(d)$ is a decreasing function of d, which confirms the above observation. As such, computing F is crucial in order to estimate the distance between the k-th anchor and the i-th sensor. From Fig. 1, the latter receives m times the k-th anchor' coordinates, each from a distinct forwarding node. Since nodes are uniformly distributed in S, knowing m, the i-th sensor is able to locally approximate F as $\hat{F} = m/\lambda$ where $\lambda = N/S$ is the WSN density. $\hat{d}_{i-k}$ could then be obtained as

$$\hat{d}_{i-k} = \Psi(\hat{F}), \tag{5}$$

where $\Psi(x) = \Phi^{-1}(x)$ is the inverse function of Φ. Unfortunately, to the best of our knowledge, there is no closed-form expression for $\Psi(x)$. It is then impossible to obtain $\hat{d}_{i-k}$ using (5). In order to circumvent this impediment, a look-up table may be envisaged at each sensor. However, such a table usually requires a large memory space; a scarce resource for these often-primitive devices. Even if it is possible to implement an additional memory space at each node, this would substantially increase the overall cost of the network, especially for large-scale WSNs. Alternatively, one may numerically compute $\hat{d}_{i-k}$. To this end, we propose to equivalently reformulate this problem as a root-finding problem of the function $\widetilde{\Phi}(x) = \Phi(x) - \hat{F}$. Many root-finding iterative algorithms already exist in the literature such as Newton-Raphson method, Brent's method, Secant method, etc.. Due to its simplicity, only the latter is of concern in this work. Using the Secant method, $\hat{d}_{i-k}$ is derived by iteratively executing the following instruction:

$$\hat{d}_{i-k}^{p+1} = \hat{d}_{i-k}^{p} - \tilde{\Phi}\left(d_{i-k}^{p}\right) \frac{d_{i-k}^{p} - d_{i-k}^{p-1}}{\Phi\left(d_{i-k}^{p}\right) - \Phi\left(d_{i-k}^{p-1}\right)}, \tag{6}$$

where p refers to the p-th iterations, until convergence (i.e., $p = p^{\max} = \inf_p\{\hat{d}_{i-k}^{p} = \hat{d}_{i-k}^{p+s}, \quad \forall s \in \mathbb{N}^{\star}\}$). From (6), two initial values $\hat{d}_{i-k}^{0}$ and $\hat{d}_{i-k}^{1}$ are required to properly compute $\hat{d}_{i-k} = \hat{d}_{i-k}^{p^{\max}}$. To guarantee fast convergence of the Secant method, $\hat{d}_{i-k}^{0}$ and $\hat{d}_{i-k}^{1}$ must be chosen among the range of possible values of $\hat{d}_{i-k}$ (i.e., [R,2R]). In this work, we opt for $\hat{d}_{i-k}^{0} = R$ and $\hat{d}_{i-k}^{1} = 2R$. As can be observed from



(a) (b)

Fig. 2: Distance estimation error (DER) vs. the number of iterations.

Fig. 2, using these values, $p^{\max}$ does not exceed 5 iterations. Knowing that the required power to execute one instruction is in the range of 10^{-4} of the power consumed per transmitted bit [15], the power needed to execute the Secant method is then very negligible with respect to the overall power consumed by each sensor. Consequently, the proposed DE approach complies with WSNs where the power is considered as a scarce resource.

Now, let us generalize the proposed DE approach by considering $n_h > 2$. In such a case, d_{i-k} would simply be the summation of $n_h/2$ two-hop distances between the k-th anchor and the i-th sensor. $\hat{d}_{i-k}$ is then given by

$$\hat{d}_{i-k} = \sum_{l=1}^{n_h/2} \Psi\left(\frac{m_l}{\lambda}\right), \tag{7}$$

where m_l is the number of forwarding nodes at the l-th 2-hop distance.

B. n_h is odd

If n_h is odd, d_{i-k} would be the summation of $(n_h-1)/2$ 2-hop distances plus the last-hop distance d^{Last}. Using the fact that the minimum square error (MMSE) of the last-hop distance estimation is obtained as $d_{\text{av}}^{\text{Last}} = \mathrm{E}\{d^{\text{Last}}\}$, $\hat{d}_{i-k}$ is given by

$$\hat{d}_{i-k} = \sum_{l=1}^{(n_h-1)/2} \Psi\left(\frac{m_l}{\lambda}\right) + d_{\text{av}}^{\text{Last}}. \tag{8}$$

Now, let us focus on $d_{\text{av}}^{\text{Last}}$. In order to derive it, one should compute the conditional cumulative distribution function (CDF) $F_Z(z) = P\left(Z \leq z / Z \leq R\right)$ where, for the sake of clarity, Z refers to the random variable d^{Last}. Actually, as

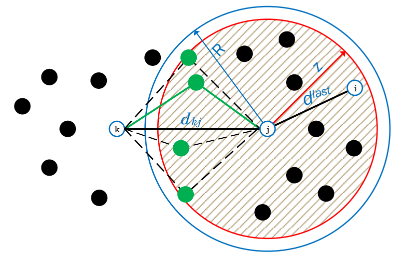


Fig. 3: Last-hop distance estimation.

shown in Fig. 3, the probability that the event $\{Z \leq z\}$ occurs is the probability that the i-th sensor be in the disc $D_j(z)$ having the j-th sensor as center and z as radius. Therefore, $F_Z(z)$ can be defined as

$$F_Z\left(z\right) = P\left(\mathcal{A}|\mathcal{B}\right) = \frac{P\left(\mathcal{A}\right)}{P\left(\mathcal{B}\right)}, \tag{9}$$

where $P\left(\mathcal{A}|\mathcal{B}\right)$ is the probability that the event $\mathcal{A} = \{$the i-th sensor is in the dashed disc $D_j(z)\}$ given $\mathcal{B} = \{$the i-th sensor is in $D_j(R)\}$ occurs. Since the nodes are uniformly distributed in S, we have $P\left(\mathcal{A}\right) = \pi z^2/S$ and $P\left(\mathcal{B}\right) = \pi R^2/S$. It follows from these results that $F_Z\left(z\right) = \left(z/R\right)^2$ and, hence, the probability density function (PDF) $f_Z(z)$ of Z is given by

$$f_Z\left(z\right) = \frac{2z}{R^2}. \tag{10}$$

Exploiting (10), we easily show that

$$\hat{d}^{\text{Last}} = \frac{2R}{3}. \tag{11}$$

In what follows, we introduce a new localization algorithm for multi-hop WSN that exploits the proposed DE approach and analytically prove its accuracy.

IV. Proposed localization algorithm

A. Initialization

As a first step of any anchor-based localization algorithm, the k-th anchor broadcasts through the network a packet which consists of a header followed by a data payload. The packet header contains the anchor position (a_k, b_k), while the data payload contains $(n, \hat{d})$, where n is the hop-count value initialized to one and $\hat{d}$ is the estimated distance initialized to zero. If the packet is successfully received by a node, it stores the k-th anchor position as well as the received hop-count $n_k = n$ in its database, adds one to the hop-count value and broadcasts the resulting message. Once this message is received by the another node, its database information is checked. If the k-th anchor's position does not exist, the node adds the received information to its database and checks the parity of n. If it is odd, the message is broadcasted after incrementing it by 1. Otherwise, the node creates a variable

m_k, which represents the number of received packets from the k-th anchor with the same data payload, and initializes it to one. However, if the node is aware of the k-th anchor's coordinates, it compares n and $\hat{d}$ with the stored ones n_k and $\hat{d}_k$, respectively. If $n > n_k$ or $n = n_k$ but $\hat{d} > \hat{d}_k$, the packet is immediately discarded. If $n < n_k$ or $n = n_k$ and $\hat{d} < \hat{d}_k$, the node updates n_k to n and $\hat{d}_k$ to $\hat{d}$. Otherwise, the parity of n is checked. If it is odd, the packet is broadcasted after incrementing it by 1. If not, m_k is incremented by 1. At this stage, a waiting-time τ, before transmitting the k-th anchor information, is envisaged to ensure that all similar packets are received. Afterwards, using m_k and the approach in Section III-A, the node estimates the last two-hop distance, adds the estimate to $\hat{d}_k$ and broadcasts the resulting packet in the network. This process will continue until each sensor in the network becomes aware of all anchors' position. It is noteworthy that, at this stage, if n_k is even, the sensor is already aware of its distance to the k-th anchor. Otherwise, it is obtained by adding, as discussed in Section III-B, $2/3$ to the stored $\hat{d}_k$.

B. Positions' computation

Once the i-th sensor obtains all the anchors' coordinates and their corresponding distances, it computes its position by solving the system of the following nonlinear equations: $(a_k - \hat{x}_i)^2 + (b_k - \hat{y}_i)^2 = \hat{d}_{i-k}^2, k = 1, \ldots, M$ where $(\hat{x}_i, \hat{y}_i)$ are the estimated i-th sensor's coordinates. After some rearrangements aiming to linearize the system above, we obtain

$$\boldsymbol{\Upsilon}\hat{\boldsymbol{\alpha}}_i = -\frac{1}{2}\boldsymbol{\kappa_i}, \tag{12}$$

where $\hat{\boldsymbol{\alpha}}_i = [\hat{x}_i, \hat{y}_i]^T$, $[\boldsymbol{\Upsilon}]_{k1} = a_k - a_M$ and $[\boldsymbol{\Upsilon}]_{k2} = b_k - b_M$, and $[\boldsymbol{\kappa_i}]_k = \hat{d}_{i-k}^2 - \hat{d}_{i-M}^2 + a_M^2 - a_k^2 + b_M^2 - b_k^2$. Since $\boldsymbol{\Upsilon}$ is a non-invertible matrix, $\hat{\boldsymbol{\alpha}}_i$ could be estimated with the pseudo-inverse of $\boldsymbol{\Upsilon}$ as follows:

$$\hat{\boldsymbol{\alpha}}_i = -\frac{1}{2}\left(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T\right)^{-1}\boldsymbol{\Upsilon}^T\boldsymbol{\kappa_i}. \tag{13}$$

Therefore, the i-th sensor is able to obtain an estimate of its coordinates as $\hat{x}_i = [\hat{\boldsymbol{\alpha}}_i]_1$, and $\hat{y}_i = [\hat{\boldsymbol{\alpha}}_i]_2$. It is noteworthy that $\hat{x}_i$ and $\hat{y}_i$ are solely dependent on the anchors' coordinates $(a_k, b_k), k = 1, \ldots, M$ and the estimated distances $\hat{d}_{k-i}, k = 1, \ldots, M$ which are all locally available at the i-th sensor. Therefore, their computation does not require any additional overhead (i.e., additional power cost), making our algorithm compliant with WSNs' power restrictions.

In what follows, the performance of the proposed localization algorithm is analyzed and compared to the most representative benchmarks in the literature.

V. Performance analysis of the proposed algorithm

A. Performance metrics

One way to prove the efficiency of the proposed localization algorithm is undoubtedly analyzing its accuracy. To this end, we introduce the following performance metric:

$$\mathcal{E}_{\mathrm{P},i} = \|\boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i\|^2 \tag{14}$$

where $\mathcal{E}_{\mathrm{P},i}$ denotes the i-th sensor's location estimation error (LEE) and $\boldsymbol{\alpha}_i = [x_i, y_i]^T$ is a vector whose entries are the true i-th sensor coordinates. From (14), $\mathcal{E}_{\mathrm{P},i}$ is an excessively complex function of the random variables $(x_i, y_i), i = 1, \ldots, N$, d_{i-k} and $\hat{d}_{i-k}, k = 1, \ldots, M$ and, hence, a random quantity of its own. Therefore, it is practically more appealing to investigate the behavior and the properties of the average LEE $\bar{\mathcal{E}}_{\mathrm{P}}(N) = \mathrm{E}\{\mathcal{E}_{\mathrm{P},i}\}$ achieved using the proposed algorithm. Actually, $\bar{\mathcal{E}}_{\mathrm{P}}(N)$ could be differently defined as

$$\bar{\mathcal{E}}_{\mathrm{P}}(N) = \mathrm{E}\left\{\mathcal{G}_{\mathrm{P}}^{\mathrm{Net}}(N)\right\}, \tag{15}$$

where $\mathcal{G}_{\mathrm{P}}^{\mathrm{Net}}(N) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}_{\mathrm{P},i}$ refers to the global LEE through the network, which is commonly used as a performance metric in the context of localization in WSNs [4]-[11]. Furthermore, using the strong law of large numbers, we show for large N that we have

$$\mathcal{G}_{\mathrm{P}}^{\mathrm{Net}}(N) \xrightarrow{p1} \bar{\mathcal{E}}_{\mathrm{P}}(N), \tag{16}$$

where $\xrightarrow{p1}$ stands for convergence with probability one. From (16), $\bar{\mathcal{E}}_{\mathrm{P}}(N)$ is not only the statistical average of $\mathcal{G}_{\mathrm{P}}^{\mathrm{Net}}(N)$, but also it approaches the latter for any given realization (i.e., any given $(x_i, y_i), i = 1, \ldots, N$). All this proves that $\bar{\mathcal{E}}_{\mathrm{P}}(N)$ is a meaningful and useful performance metric.

B. Proposed algorithm's average LEE

It follows from (13) that

$$\mathcal{E}_{\mathrm{P},i} = \frac{1}{4}\left\|\left(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T\right)^{-1}\boldsymbol{\Upsilon}^T\boldsymbol{\delta}_i\right\|^2. \tag{17}$$

where $[\boldsymbol{\delta}_i] = [\epsilon_1 - \epsilon_M, \ldots, \epsilon_{M-1} - \epsilon_M]^T$ with $\epsilon_k = \hat{d}_{i-k}^2 - d_{i-k}^2$ being the squared-distance estimation error. $\mathcal{E}_{\mathrm{P},i}$ is then given by

$$\begin{aligned}\mathcal{E}_{\mathrm{P},i} &= \mathrm{Tr}\left(\left(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T\right)^{-1}\boldsymbol{\Upsilon}^T\boldsymbol{\delta}_i\boldsymbol{\delta}_i^T\boldsymbol{\Upsilon}\left(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T\right)^{-1}\right)\\ &= \sum_{k=1}^{M-1}\boldsymbol{\Omega}_{kk}\left([\boldsymbol{\delta}_i]_k\right)^2 + \sum_{k=1}^{M-1}\sum_{l=1,l\neq k}^{M-1}\boldsymbol{\Omega}_{kl}[\boldsymbol{\delta}_i]_l[\boldsymbol{\delta}_i]_k,\end{aligned} \tag{18}$$

where $\mathrm{Tr}(\mathbf{X})$ is the trace of the matrix $\mathbf{X}$ and $\boldsymbol{\Omega} = \boldsymbol{\Upsilon}\left(\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^T\right)^{-2}\boldsymbol{\Upsilon}^T$. Note in the first line of (18) that we exploit the cyclic property of the trace. Since $\epsilon_k, k = 1, \ldots, M$ are i.i.d random variables, we have from (18) the following

$$\bar{\mathcal{E}}_{\mathrm{P}}(N) = \sigma_\epsilon^2\left(2\mathrm{Tr}(\boldsymbol{\Omega}) + \sum_{k=1}^{M-1}\sum_{l=1,l\neq k}^{M-1}\boldsymbol{\Omega}_{kl}\right). \tag{19}$$

Now let us turn our attention to σ_ϵ^2. For the sake of clarity, we first assume that there are exactly two hops between the i-th sensor and each anchor. The obtained results will be thereafter generalized. In such a case, from (4) and (5), the Taylor series expansion of $\Psi(x)$ at F_k yields

$$\hat{d}_{i-k} = d_{i-k} + \sum_{n=1}^{\infty}\frac{\Psi^{(n)}(F_k)}{n!}\Delta F^n, \tag{20}$$

where $\Psi^{(n)}(x)$ is the n-th derivative of $\Psi(x)$ and $\Delta F = m_k/\lambda - F_k$. Assuming that ΔF is small enough to allow approximation of $\hat{d}_{k-i}$ by the first three non-zero terms of the right-hand-side (RHS) of (20), we obtain

$$\epsilon_k \simeq 2d_{i-k}\Psi^{(1)}(F_k)\Delta F + \left(\left(\Psi^{(1)}(F_k)\right)^2 + d_{i-k}\Psi^{(2)}(F_k)\right)\Delta F^2, \tag{21}$$

where $\Psi^{(1)}(x) = \left(4R^2 - \Psi(x)^2\right)^{-1/2}$ and $\Psi^{(2)}(x) = \Psi(x) / \left(4R^2 - \Psi(x)^2\right)^2$. Since the nodes are uniformly deployed in S, the probability of having m_k nodes in F_k follows a Binomial distribution $\text{Bin}(N,p)$ where $p = \frac{F_k}{S}$ and, therefore, the first and second order statistics of m_k are $\text{E}\{m_k\} = \lambda F_k$ and $\text{E}\{m_k^2\} = \lambda F_k\left(1 - \frac{F_k}{S} + \lambda F_k\right)$, respectively. Using the latter along with (21) yields

$$\text{E}_{m_k}\{\epsilon_k\} = 4R^2\lambda^{-1}F_k\left(1 - \frac{F_k}{S}\right) / \left(4R^2 - \Psi(F_k)^2\right)^2, \tag{22}$$

where the expectation is taken with respect to m_k. As could be observed from (22), the probability density function $f_{F_k}(F)$ of F_k is crucial to derive σ_ϵ^2 in closed-form. For the sake of mathematical tractability, F_k is assumed to be Uniform in $[0, F_{\max}]$ where $F_{\max} = \Phi(R) = \left(-\frac{\sqrt{3}}{2} + \frac{2\pi}{3}\right)R^2$. Despite this simplifying assumption, we will shortly see in Section VI that the obtained analytical results closely match those obtained empirically by Monte Carlo simulations. $\text{E}\{\epsilon_k\}$ is then given by

$$\begin{aligned}\text{E}\{\epsilon_k\} &= \frac{1}{9\left(3\sqrt{3} - 4\pi\right)N}\left(12\left(-9 + \sqrt{3}\pi\right)S\right.\\ &+ \left.\left(27\sqrt{3} + 4\pi(9 - 2\sqrt{3}\pi)\right)R^2\right).\end{aligned} \tag{23}$$

Following similar steps as above, we show that

$$\begin{aligned}\text{E}\{\epsilon_k^2\} = &\frac{1}{27\left(3\sqrt{3} - 4\pi\right)N}\left(\left(4\pi\left(-189 + 4\pi\left(9\sqrt{3} + 4\pi\right)\right)\right.\right.\\ &\left.\left.-1053\sqrt{3}\right)R^4 - 9\left(8\pi\left(3\sqrt{3} + 2\pi\right) - 243\right)R^2S\right),\end{aligned} \tag{24}$$

and, hence, σ_ϵ^2 is obtained. It can be then inferred from (19)-(24) that the achieved average LEE $\bar{\mathcal{E}}_{\text{P}}(N)$ using the proposed algorithm linearly decreases with N when S and R are fixed. Furthermore, for sufficiently large N, we have $\bar{\mathcal{E}}_{\text{P}}(N) \simeq 0$. This property is actually a desired feature for any sensor localization algorithm since WSNs are typically dense. It is noteworthy here that the best representative benchmarks in the literature lack such a feature [4], [9]. Recall, however, that the results in (23) and (24) were derived assuming that the number of hops between any anchor-sensor pair is $n_h = 2$ hops. For the sake of generalization, we consider now that n_h is a random variable with mean $\bar{n}_h$. In such a case, it can be shown that $\text{E}\{\epsilon_k\}$ is given by [16]

$$\begin{aligned}\text{E}\{\epsilon_k\} = R^2&\Big(\left(\xi_{0,1}^2(2\bar{n}_h(\bar{n}_h - 3) + 3)\right)/8N^2 + \big(3(2\bar{n}_h - 1)(\xi_{0,2} +\\ &2\xi_{1,1}) + \xi_{0,1}(3\xi_{1,0}(2\bar{n}_h(\bar{n}_h - 3) + 3) + 4(\bar{n}_h - 1))\big)/12N - 1/36\Big).\end{aligned} \tag{25}$$

where $\xi_{n,m}$ $n,m = 0,1,2$ are parameter functions of R and S whose expressions are listed in Tab. I. In turn, $\text{E}\{\epsilon_k^2\}$ could be expressed in (26), as shown on the top of the next page. It is worth mentioning that the results in (25) and (26) are very interesting in terms of implementation strategy, since they allow, through (19), to easily find the smallest N that keeps $\bar{\mathcal{E}}_{\text{P}}(N)$ below a certain level. They also allow to find the best anchor placement strategy that minimizes $\bar{\mathcal{E}}_{\text{P}}(N)$ for a given N. Moreover, in contrast with the two-hop case, it follows from (25) and (26) that we have

$$\bar{\mathcal{E}}_{\text{P}}(N) \simeq 0.16R^4 + \frac{2R^3}{15}\xi_{1,0} + \frac{R^2}{9}\xi_{2,0} \neq 0, \tag{27}$$

when N is large enough. Note that x is nothing but the error incurred when estimating the last hop of an odd distance between any anchor-sensor pair in the network. A proper anchor selection scheme should then be envisaged to make our proposed algorithm reach its optimal accuracy (i.e., $\bar{\mathcal{E}}_{\text{P}}(N) \simeq 0$) at large N. Indeed, if each sensor selects among the list of anchors only those with an even number of hops, its achieved average LEE would approach 0 when N is large enough. This, of course, requires that at least 3 anchors comply with the above criterion. Please note that such a selection scheme could be easily implemented in each sensor without burdening neither the implementation complexity of the proposed localization algorithm nor the overall cost of the WSNs.

C. Proposed algorithm's asymptotic LEE

So far, we derived the average LEE achieved by our localization algorithm and studied its behavior and properties. Motivated by the fact that the LEE is a more practical metric than its average, we investigate in this section its statistical properties more thoroughly for the sake of further highlighting the proposed algorithm's accuracy.

Let us consider again the 2-hop case (i.e., two hops between the i-th sensor and the k-th anchor nodes). Exploiting the fact that m_k is a Binomial random variable, we have from the Chebyshev's inequality we have

$$1 - P(|\Delta F| < \kappa) \leq \frac{F_k(S - F_k)}{N\kappa^2}, \tag{28}$$

where κ is any given strictly positive real. If the latter is chosen small enough to guarantee the equivalence $|\Delta F| < \kappa \Leftrightarrow |\Delta F| \simeq 0$, it holds for sufficiently large N that

$$P(|\Delta F| \simeq 0) \simeq 1. \tag{29}$$

Exploiting this result along with (21) we obtain

$$P(\epsilon_k \simeq 0) \simeq 1, \tag{30}$$

$$
\begin{aligned}
\mathrm{E}\{\epsilon_k^2\} = {} & \frac{R^2}{4N}\Bigg(\frac{16\xi_{0,2}}{9} - \frac{4\xi_{0,1}}{27} + \frac{(\bar{n}_h-1)}{72}\Big(2(\xi_{0,2}+2\xi_{1,1}) + \xi_{0,1}(\bar{n}_h-3)\big(\frac{\xi_{0,1}}{N}+2\xi_{1,0}\big)\Big)\Big(\frac{12\xi_{0,1}(\bar{n}_h-1)}{N}-1\Big) + (\bar{n}_h-1) \\
& \Bigg(4\xi_{2,2} + \frac{\bar{n}_h-3}{2}\Big(\frac{3\xi_{0,2}^2}{N} + 8\big(\frac{\xi_{1,1}^2}{N}+\xi_{1,2}\xi_{1,0}\big) + \frac{2\xi_{0,1}}{N}\big(6\xi_{1,2}+4\xi_{2,1}+(3\xi_{0,2}+4\xi_{1,1})(\bar{n}_h-5)\xi_{1,0}\big) + \frac{6\xi_{0,1}^3\xi_{1,0}}{N^2}(\bar{n}_h-7)(\bar{n}_h-5) + \\
& + \frac{2\xi_{0,2}\xi_{0,1}^2}{N^2}\big((\bar{n}_h-5)\xi_{1,0}^2+2\xi_{2,0}\big)\Big(\frac{6\xi_{0,2}}{N} + (\bar{n}_h-5)\big(\frac{12\xi_{1,1}}{N} + (\bar{n}_h-7)\xi_{1,0}^2+\xi_{2,0}\big)\Big)\Big)\Bigg) + \bar{n}_h\Bigg(4\xi_{2,2} + (\bar{n}_h-2)\Big(\frac{3\xi_{0,2}^2}{2N} + 4\big(\frac{\xi_{1,1}^2}{N} \\
& + \xi_{1,2}\xi_{1,0}\big) + \frac{\xi_{0,1}}{N}\big(6\xi_{1,2}+4\xi_{2,1}+(\bar{n}_h-4)(3\xi_{0,2}+4\xi_{1,1})\xi_{1,0}\big) + 3\frac{\xi_{0,1}^3}{N^2}(\bar{n}_h-6)(\bar{n}_h-4)\xi_{1,0} + \xi_{0,2}\big((\bar{n}_h-2)\xi_{1,0}^2+\frac{\xi_{2,0}}{2}\big) + \frac{\xi_{0,1}^2}{N} \\
& \Big(\frac{3\xi_{0,2}}{N} + (\bar{n}_h-4)\big(\frac{6\xi_{1,1}}{N} + (\frac{\bar{n}_h-6}{2})\xi_{1,0}^2+\xi_{2,0}\big)\Big)\Big)\Bigg)\Bigg) + R^2\Big(\frac{25R^2}{162} + \frac{2R\xi_{1,0}}{15} + \frac{\xi_{2,0}}{9}\Big). \qquad (26)
\end{aligned}
$$

Parameter	Closed-form expression
$\xi_{1,0}$, $\xi_{2,0}$	$6\sqrt{3}R \big/ \big(4(\pi - 3\sqrt{3})\big)$, $(8\sqrt{3}\pi + 9)R^2 \big/ \big(8\sqrt{3}\pi - 18\big)$
$\xi_{0,1}$	$\Big(\big(297\sqrt{3} - 8\pi(9+2\sqrt{3}\pi)\big)R^2 + 6\big(4\sqrt{3}\pi - 27\big)S\Big) \Big/ \big(36(3\sqrt{3}-4\pi)R\big)$
$\xi_{0,2}$	$\Big((64\pi^3 - 567\sqrt{3} - 216\pi)R^2 + 36(27 - 4\pi^2)S\Big) \Big/ \big(216(3\sqrt{3}-4\pi)\big)$
$\xi 1,1$	$\Big(\Big(1215\sqrt{3} - 8\pi\big(8\pi(3\sqrt{3}+\pi) - 135\big)\Big)R^2 + 36\big(4\pi(2\sqrt{3}+\pi) - 99\big)S\Big) \Big/ \big(432(3\sqrt{3}-4\pi)\big)$
$\xi_{1,2}$	$\Big(\big(-1701\sqrt{3} + 40\pi(13+2\sqrt{3}\pi)\big)R^3 + 30(19 - 4\sqrt{3}\pi)RS\Big) \Big/ \big(30(3\sqrt{3}-4\pi)\big)$
$\xi_{2,1}$	$\Big(9R\big(3\sqrt{3}R(-1+45R) - 104S\big) - 54S - 16\sqrt{3}(1+3R)\pi^2R^2 + 24\pi\big(3(1-5R)R^2 + \sqrt{3}(1+6R)S\big)\Big) \Big/ \big(36(3\sqrt{3}-4\pi)\big)$
$\xi_{2,2}$	$\Big(9\big(243 - 8\pi(3\sqrt{3}+2\pi)\big)R^2S + \Big(4\pi\big(4\pi(9\sqrt{3}+4\pi) - 189\big) - 1053\sqrt{3}\Big)R^4\Big) \Big/ \big(108(3\sqrt{3}-4\pi)\big)$

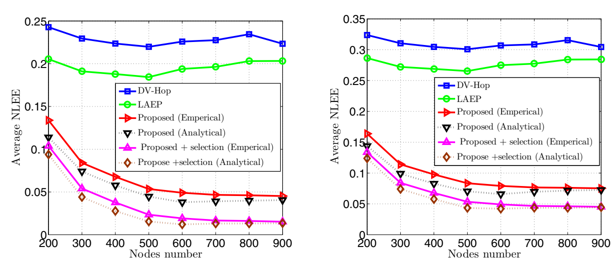
TABLE I: Closed-form expressions of $\xi_{n,m}$ $n, m = 0, 1, 2$.

and, hence, for large N we have $\mathcal{E}_{\mathrm{P},i} \simeq 0$. This further proves the accuracy of the proposed algorithm. Furthermore, it is straightforward to show that $\mathcal{E}_{\mathrm{P},i} \simeq 0$ also holds when the number of hops between the i-th sensor and all anchors is even (but not necessarily 2). This emphasizes even more the importance of the anchor selection scheme discussed above.

VI. Simulations Results

Monte Carlo simulations are provided, in this section, to support the theoretical results. These simulations are conducted to compare, under the same network settings, the R^2-normalized LEEs (NLEE)s achieved by the proposed algorithm and two of the best representative localization algorithms currently available in the literature, i.e., DV-Hop [4] and LAEP [9]. All simulation results are obtained by averaging over 500 trials. In all simulations, sensors are uniformly deployed in a 2-D square area. S, R, and M are set to 10^4 m^2, 22 m, and 20, respectively. Two commonly used anchor placement strategies in the context of WSNs are considered: the perimeter and grid placements.

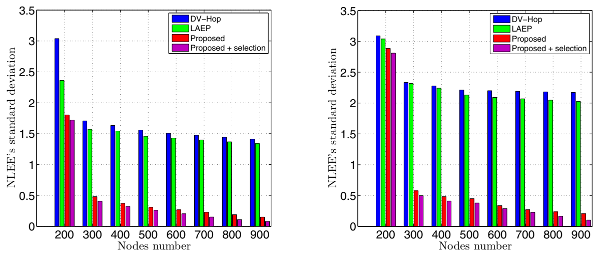
Fig. 4 plots the average NLEE achieved by the proposed algorithm, DV-Hop, and LAEP versus N with two anchor placement strategies: perimeter in Fig. 4(a) and grid placement in Fig. 4(b). From these figures, the proposed localization algorithm always outperforms in accuracy its counterparts. It is, for instance at $N = 700$, until 12 times more accurate than DV-Hop and until 10 times more accurate than LAEP. This further proves the proposed algorithm's efficiency and highlights the advantage of using it in WSNs instead of its counterparts.



(a) Perimeter anchor placement. (b) Grid anchor placement.

Fig. 4: Average NLEE achieved by the proposed algorithm, DV-Hop, and LAEP versus N.

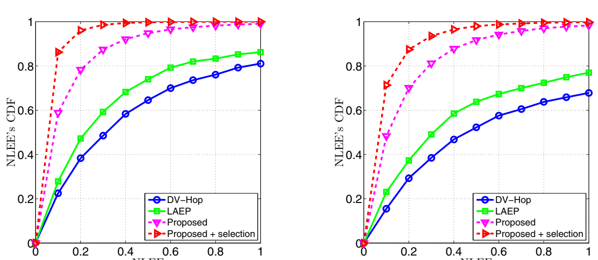
Fig. 5 plots the NLEE's standard deviation achieved by all localization algorithms versus N, for two anchor placement strategies: perimeter in Fig. 5(a) and grid placement in Fig. 5(b). As can be observed from these figures, regardless of the anchor placement strategy, the one achieved by the proposed algorithm substantially decreases when N increases while those achieved by the other algorithms slightly decrease. Furthermore, the NLEE's standard deviation achieved by the proposed algorithm with or without anchor selection approaches 0 for any placement strategy. This is due to the fact that the LEE itself being around 0 occurs almost certainly (i.e., with almost probability 1) as stated in Section V-C. On the other hand, Figs. 5(a) and 5(b) suggest that the proposed algorithm's performance is further improved if the low-cost anchor selection scheme introduced in Section V-B is implemented at each sensor. All these observations corroborates the results and discussions disclosed in Section V.



(a) Perimeter anchor placement. (b) Grid anchor placement.

Fig. 5: NLEE's standard deviation achieved by the proposed algorithm, DV-Hop, and LAEP versus N.

Fig. 6 illustrates the NLEE's CDF achieved by our proposed localization algorithm with and without the anchor selection scheme as well as that achieved by the other algorithms for two anchor placement strategies: perimeter in Fig. 5(a) and grid placement in Fig. 5(b). From these figures, using the proposed algorithm, 80% (98% with anchor selection) of the sensors could estimate their position with NLEE less than 0.2. In contrast, 45% of the nodes achieve the same accuracy with LAEP and only about 38% with DV-Hop using the perimeter anchor placement strategy. This proves even more the accuracy of the proposed localization algorithm.



(a) Perimeter anchor placement. (b) Grid anchor placement.

Fig. 6: NLEE's CDF achieved by the proposed algorithm, DV-Hop, and LAEP when $N = 300$.

VII. Conclusion

In this paper, we proposed a novel localization algorithm which properly exploits, in addition to the hop-based information, the forwarding nodes' number between any anchor-sensor pair. Its average location estimation error (LEE) was derived in closed-form and compared to those of the best representative algorithms in the literature. We showed that the proposed algorithm outperforms in accuracy all its counterparts. Furthermore, we proved that, in contrast to the latter, our algorithm is able to achieve an average LEE of about 0, when the total sensors' number N is large enough. We also proved in such a condition that any realization of its achieved LEE approaches 0, which confirms unambiguously its high accuracy.

References

[1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.,* vol. 40, no. 8, pp. 102-114, August 2002.

[2] J. Rezazadeh, M. Moradi, A.S. Ismail and E. Dutkiewicz, "Superior Path Planning Mechanism for Mobile Beacon-Assisted Localization in Wireless Sensor Networks," *IEEE Sensors J.,* vol. 14, no. 9, pp. 3052-3064, May 2014.

[3] S. Hong, D. Zhi, S. Dasgupta, and Z. Chunming, "Multiple Source Localization in Wireless Sensor Networks Based on Time of Arrival Measurement," *IEEE Trans. Signal Process.,* vol. 62, no. 8, pp. 1938-1949, February 2014.

[4] D. Niculescu and B. Nath, "Ad hoc positioning system (APS),“ *Proc. IEEE GLOBECOM'2001,* San Antonio, TX, USA, November 25-29, 2001.

[5] A. Boukerche, H.A.B.F. Oliveira, E.F. Nakamura, A.A.F. Loureiro, "DV-Loc: a scalable localization protocol using Voronoi diagrams for wireless sensor networks," *IEEE Wireless. Commun. Mag.,* vol. 16, no. 2, pp. 50-55, April 2009.

[6] L. Gui, T. Val, A. Wei, "Improving Localization Accuracy Using Selective 3-Anchor DV-Hop Algorithm," *Proc. IEEE VTC'2011,* San Francisco, CA, USA, Septembre 5-8, 2011.

[7] B. Huang, C. Yu, B. D. O. Anderson and G. Mao, "Estimating distances via connectivity in wireless sensor networks," *Wirel. Commun. Mob. Com.,* vol. 14, no. 5, pp. 541-556, April 2014.

[8] X. Ta, G. Mao, and B. D. Anderson, "On the probability of k-hop connection in wireless sensor networks," *IEEE Commun. Lett.,* vol. 11, no. 9, pp. 662-664, August 2007.

[9] Y. Wang, X. Wang, D. Wang, and D. P. Agrawal, "Range-Free Localization Using Expected Hop Progress in Wireless Sensor Networks," *IEEE Trans. Parallel Distrib. Syst.,* vol. 25, no. 10, pp. 1540-1552, October 2009.

[10] A. El Assaf, S. Zaidi, S. Affes, and N. Kandil, "Efficient Range-Free Localization Algorithm for Randomly Distributed Wireless Sensor Networks," *IEEE GLOBECOM'2013,* Atlanta, GA, December 9-13, 2013.

[11] A. EL Assaf, S. Zaidi, S. Affes, and N. Kandil, "Range-free localization algorithm for heterogeneous wireless sensors networks," *Proc. IEEE WCNC'2014,* Istanbul, Turkey, Apr. 6-9, 2014.

[12] S. Vural and E. Ekici, "On Multihop Distances in Wireless Sensor Networks with Random Node Locations," *IEEE Trans. Mobile Comput.,* vol. 9, no. 4, pp. 540-552, April 2010.

[13] JC. Kuo, W. Liao, "Hop Count Distribution of Multihop Paths in Wireless Networks With Arbitrary Node Density: Modeling and Its Applications," *IEEE Trans. Veh. Technol.,* vol. 56, no. 4, pp. 2321-2331, July 2007.

[14] Q. Xiao, B. Xiao, J. Cao, J. Wang, "Multihop Range-Free Localization in Anisotropic Wireless Sensor Networks: A Pattern-Driven Scheme," *IEEE Trans. Mobile Comput.,* vol. 9, no. 11, pp. 1592-1607, November 2010.

[15] N. Patwari, J.N. Ash, S. Kyperountas, A.O. Hero, R.L. Moses and N.S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks,“ *IEEE Signal Processing Mag.,* vol. 22, no. 4, pp. 54-69, July 2005.

[16] S. Zaidi, A. El Assaf, S. Affes, and N. Kandil, "Accurate range-free localization in multi-hop wireless sensor networks," Submitted to *IEEE Trans. Commun.,* August 2015.

[17] S. Biaz, Ji. Yiming, Q. Bing, Wu. Shaoen, "Realistic radio range irregularity model and its impact on localization for wireless sensor networks," *IEEE WiCOM'2005,* Wuhan, China, September 23-26 2005.