RESEARCH ARTICLE

Characterizing multi-hop localization for
Internet of things

Walid M. Ibrahim\(^1\), Najah Abu Ali\(^2\), Hossam S. Hassanein\(^1\) and Abd-Elhamid M. Taha\(^3\)*

\(^1\) School of Computing, Queen’s University, Canada
\(^2\) College of IT, United Arab Emirates University, United Arab Emirates
\(^3\) College of Engineering, Alfaisal University, Saudi Arabia

ABSTRACT

Deployments over large geographical areas in the Internet of Things (IoT) pose a major challenge for single-hop localization techniques, giving rise to applications of multi-hop localizations. While many proposals have been made on implementations for multi-hop localization, a close understanding of its characteristics is yet to be established. Such an understanding is necessary, and is inevitable in extending the reliability of location-based services in IoT. In this paper, we study the characteristics of multi-hop localization and propose a new solution to enhance the performance of multi-hop localization techniques. We first examine popular assumptions made in simulating multi-hop localization techniques, and offer rectifications facilitating more realistic simulation models. We identify the introduced errors to follow the Gaussian distribution, and the estimated distance follows the Rayleigh distribution. We next use our simulation model to characterize the effect of the number of hops on localization in both dense and sparse deployments. We find that, contrary to common belief, it is better to use long hops in sparse deployments, while short hops are better in dense deployments – despite the traffic overhead. Finally, we propose a new solution that decreases and manages the overhead generated during the localization process. Copyright © 2016 John Wiley & Sons, Ltd.

KEYWORDS

Internet of Things; Localization; Multi-hop Wireless Localization; DV-Distance; DV-Hop

*Correspondence
Abd-Elhamid M. Taha, College of Engineering, Alfaisal University, PO Box 50927, Riyadh 11533, Saudi Arabia.
E-mail: hamid.taha@gmail.com

1. INTRODUCTION

Identifying the physical location of things or Sensor Nodes (SNs) is a major requirement in the Internet of Things (IoT) [1,2]. In many applications, the sensed data is rendered more useful if coupled with the location of the event [3,4]. A straightforward solution to realize this requirement is to add a GPS module to individual SNs. GPS circuitry, however, is costly in its monetary, volume, and energy requirements. An alternative solution is to estimate the location of unlocalized SNs with the aid of what can be called anchor nodes. Anchor nodes know their absolute locations by either using GPS, or by being attached to predefined locations with known coordinates. To localize SNs, anchor nodes broadcast their location with the operating instructions to SNs, which use the received locations of anchor nodes to estimate their own locations [5].

Depending on the application and size of the sensed area, localization techniques can be based either on single-hop or multi-hop localization. The work in [6] discusses the application of single-hop and multi-hop localization and their characteristics. In a small scale and/or sparse deployment, using single-hop techniques, unlocalized SNs would require a minimum of three anchor nodes (in a 2-D setting) or four anchor nodes (in a 3-D setting) within their transmission range in order to estimate their location. In large scale and/or dense deployments, however, especially when the sensed area is vast, most of the SNs are not located within the transmission range of three anchor nodes at the same time unless the number of anchor nodes is increased to cover the entire sensed area by at least three anchor nodes. Thus, multi-hop localization [7] is used to estimate the locations of SNs in large scale deployments. Multi-hop localization uses two or more wireless hops to convey location information from anchor node to SN.

Localization error is defined as the Euclidean distance between the estimated location of the SN and its actual location. In large scale deployments, as the sensed area increases, the localization error increases. An understanding of the relationships or trade-off between number of
hops, transmission range, and localization error is necessary to reduce the localization error for SNs in large deployment scenarios. Intuitively, decreasing the transmission range would increase the number of packets used in the localization process, which would shorten the SN’s lifespan. Meanwhile, increasing the lifespan of the SNs is important as it should be operational for several months or even years without requiring a change of batteries. Moreover, decreasing the number of packets exchanged between SNs is essential as the generated traffic could increase the collision rate, which could also affect the overall localization process. It is therefore elemental to understand the behavior of different localization techniques, as well as the relationship between hop count, transmission, and localization error. We furthermore need to identify design bases that reduce the traffic generated from localization algorithms.

The contributions of this work are thus as follows:

- Creating a more realistic simulation model to simulate the localization error to represent actual measurement used in estimating the distance between SNs. Previous simulation models added Gaussian noise to the actual distance between SNs [8,9]. However, in this work, we show that the simulation can be more accurately represented by using Rayleigh distribution instead of using Gaussian distribution. By analyzing real measurements, we show that using Rayleigh distribution gives a more realistic representation of the localization error. Then we show, by using multi-hop simulation, the difference between using Gaussian and Rayleigh distribution to validate our model.

- Characterizing the error behavior of multi-hop localization, studying the effect of changing the transmission range of SNs and observing how changing the number of hops affects the localization error. Existing works lack quantitative analysis of the relation between the hop numbers, transmission range, and localization error [8,9]. Researchers in multi-hop localization commonly believe that increasing the number of hops between the anchor nodes and SNs will increase the localization error. In this work, we show that this is not always the case, and that this belief indeed holds but only under special conditions, under which, using a larger number of hops gives a lower localization error than using a smaller number of hops.

- Characterizing the overhead and the amount of traffic generated during the localization process. We study how the density of SNs would affect the number of generated packets during the localization process. We also verify the different parameters that have an effect on the amount of traffic generated. Then, we propose a new solution to reduce the number of packets exchanged between SNs without worsening (i.e., increasing) localization error. By reducing the traffic generated, the lifespan of SNs will be longer.

The remainder of this paper is organized as follows. Section 2 offers the background and motivation for this work. Section 3 introduces a realistic error model to estimate the distance between SN using Receive Signal Strength Indicator (RSSI), while Section 4 explores the effect of the transmission range of SNs in a multi-hop localization environment and its impact on localization error. Section 5 proposes a new aggregation scheme that reduces the number of packets exchanged during the localization process, and evaluates the proposed scheme using various operational aspects, including number of packets sent, collisions, localization error, in addition to the percentage of unlocalized SNs. Finally, Section 6 concludes and highlights intended future work.

2. BACKGROUND

In this section, we provide the background about the main multi-hop localization techniques used in WSNs that cover range-based and range-free categories. We also elaborate on the multilateration process used to estimate nodes location, we discuss the elements of error computation utilized in this work.

2.1. Multi-hop localization techniques

We adopt two generic techniques that represent the major categories of multi-hop localization techniques, namely DV-Hop and DV-Distance. DV-Hop represents the connectivity based category, while DV-Distance represents the distance based category [10]. In the following subsection, we give an overview of DV-Hop and DV-Distance.

2.1.1. DV-hop localization technique.

The DV-Hop localization technique has two stages [11,12]. In the first stage, the anchor nodes broadcast their actual locations to the SNs. The SNs save the shortest number of hops to each anchor node along with the anchor node’s location. At the end of the first stage, each SN maintains a table of \([x_i, y_i, h_i]\), where \(x_i\) and \(y_i\) are the coordinates of anchor \(i\) and \(h_i\) is the shortest number of hops to reach anchor \(i\). SNs exchange the shortest hop location packets only with their neighbors. When an anchor node receives a location packet from another anchor nodes, it estimates the average distance for a single hop for the entire network.

The average distance of a single hop of anchor \(i\) is calculated as follows:

\[
C_i = \sum_{j=1}^{M} \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{h_j}, \quad \text{where } i \neq j. \quad (1)
\]

In the second stage, the anchor nodes broadcast their estimates of the average distance for a single hop. The SN saves the average single hop from the closest anchor node
node and forwards it to its neighboring nodes. The SNs use the received average distance for a single hop and multiply it with the total number of hops for each anchor node using the information saved in the table \( \{x_i, y_i, h_i\} \). Finally, these values are plugged in the multilateration equation described in the next subsection.

The error in the DV-Hop localization technique appears because it assumes all hops to have the same value.

Figure 1 shows an example of DV-Hop. SNs A1, A2, and A3 are anchor nodes. Anchor node A1 has both the Euclidean distance and hop numbers to anchor nodes A2 and A3. Anchor node A1 calculates the average distance for a single hop in meters as follows \( \frac{150 + 50}{5 + 3} = 25 \) m. Anchor node A1 then broadcasts the average distance for a single hop to the network. In a similar manner, anchor nodes A2 and A3 compute the average distance for a single hop \( \frac{50 + 65}{3 + 4} = 16.43 \) and \( \frac{65 + 150}{4 + 5} = 23.89 \), respectively. The SN receives the average distance for a single hop from anchor nodes. Note that the SN only saves the average hop distance received from the closest anchor node, that is, for SNi only the average hop distance information received from A2 are saved. SNi then estimates the distances to the three anchor nodes to be: \( A_1 = 2 \times 25 \), \( A_2 = 2 \times 16.43 \), and \( A_3 = 3 \times 23.89 \). These values are next plugged into the triangulation procedure to calculate the estimated location of the SNs.

### 2.1.2. DV-distance localization technique.

The DV-Distance localization technique [13] has only one stage, in which the anchor nodes broadcast their locations to the entire network. The location packet contains the actual location \( x_i, y_i \) of anchor node \( i \) and the total distance traveled \( d_i \). The anchor node initializes the distance to zero. When an SN receives the packet, it estimates the distance the packet traveled for a single hop using either RSSI or Time of Arrival (ToA). After which, the SN node adds the estimated distance to the total distance traveled by the packet and forwards the packet.

Thus, each SN maintains a table of \( \{x_i, y_i, d_i\} \), where \( x_i \) and \( y_i \) are the coordinates of anchor \( i \) and \( d_i \) is the cumulative traveling distance estimated in meters from anchor node \( i \). The DV-Distance technique is prone to distance estimation errors due to obstacles between the two SNs, multi-path fading, noise interference, and irregular signal propagation. However, the hops between SNs are not assumed to be physically equidistant.

### 2.2. Multilateration using linear least squares

Multilateration is the process that uses the estimated distance between anchor nodes and un-localized SNs to estimate the location of the SNs. Ideally, the multilateration computation utilizes distance measurements assumed to be accurately estimated and noise-free. (Figure 2(a) illustrates the effect of noise on distance measurements.) Such assumptions, however, are not practically applicable as the estimation of distance measurements is easily affected by surrounding noise. These inaccuracies prevent the circles to intersect in a single point making the localization process more challenging. Figure 2(b) shows that the three circles do not intersect in a single point. In order to estimate the SN’s location, a possible solution is to use the least squares optimization [14].

![Figure 1. An example for DV-hop localization.](image)

![Figure 2. The difference between trilateration based on noise-free and noisy distance measurements.](image)
3. MODELING LOCALIZATION ERROR

Estimating the distance between two SNs, that is, ranging, is the main component of localization. The most common techniques used in WSN localization are based on RSSI and ToA. Both techniques are prone to adding noise to the estimated distance [15]. Using RSSI, the SN estimates the distance between itself and the sender using the strength of the received radio frequency signal by using RSSI profiling measurements or estimating the distance using the analytical model by mapping the RSSI to distance using the path-loss propagation model. Meanwhile, using ToA, the SN estimates the distance by measuring the amount of time it takes the signal to travel between the sender and itself, and assuming a constant signal speed, multiplies the time taken by the speed of the signal.

In RSSI-based ranging, signal attenuation over distance is assumed to be previously known and the distance is estimated using the following equation [16].

$$d_{ij} = d_0 \left( \frac{P_j}{P_{0j}(d_0)} \right)^{-1/n_p} \frac{\sigma_i^2}{e^{2\eta d}}.$$  

(2)

Here, $P_{0j}(d_0)$ is a known reference power value at a reference distance $d_0$ from the transmitter, $P_j$ is the RSSI measurement between a transmitter $i$ and receiver $j$, $n_p$ is the path loss exponent that indicates the rate at which the RSSI decreases with distance, and $\eta = 10^{10/(10)}$.

RSSI-based ranging is sensitive to channel noise, interference and reflections, all of which have significant impact on signal amplitude. And while ToA-based ranging relies on the signal speed rather than the signal strength, it is fairly immune to various sources of noise including signal attenuation, refraction, and reflection. However, the estimated distance of ToA is affected by processing time, queuing time and rarity of perfect line-of-sight between SNs.

Previous works in localization that use RSSI and ToA in their theoretical analysis or simulation often adopt the noisy disk model [15,17–19], which estimates the distance between SNs in order to: (i) evaluate and compare different localization techniques; (ii) mathematically derive the maximum likelihood for localization; and/or (iii) study the lower bounds on localization error. The model comprises two parts: node connectivity and noise component. The node connectivity component represents the actual distance between the two SNs, while the noise component represents the noise distribution of the estimated distance and the actual distance.

To be sure, while the works in [15,17,18], and [19], studied different localization problems, they all used the noisy disk model and the Gaussian noise to account for the estimation error in distance between the $i^{th}$ and $j^{th}$ SN. In these works, the estimated distance was defined as follows.

$$d_{ij} = d_{ij} + \varepsilon_{ij} \quad \forall i, j = 1, 2, \ldots, M$$  

(3)

where $r_{ij} = \|x_i - x_j\|$ is the noise free distance between node $i$ and $j$, and $\varepsilon_{ij}$ is a non-negative zero mean Gaussian noise where $\sigma^2$ is a large number estimated and is known a priori [15,17–19].

Liu et al. proposed an iterative least square to localize SNs using a small number of anchors [20]. The authors proposed an error control mechanism. Their algorithm was evaluated using MATLAB and they simulated three different noise models. The first experiment, did not have noise to the distance. In the second experiment, they added Gaussian noise to the distance similar to Equation 3 and fixed the $\sigma$ to 1.7 inches. In the third experiment, they used the following equation:

$$\varepsilon = \begin{cases} 
\delta_1 + \varepsilon_1 & \text{if } d < d_0, \text{ where } \varepsilon_1 \sim \mathcal{N}(0, \sigma_1) \\
\delta_0 + \varepsilon_2 & \text{otherwise, where } \varepsilon_2 \sim \mathcal{N}(0, \sigma_2)
\end{cases}$$

(4)

where $d_0 = 120$ inches and $\sigma_1 = K \delta_1$ where $K$ is a large number ($10^6$), with the assumption that noise increases rapidly when the distance exceeds a defined threshold.

To take the range into consideration, Chan et al. [21] added a zero-mean white Gaussian process with the variance $\sigma^2 = \frac{d_0^2}{K}$ to propose a new weighted multidimensional scaling as a localization scheme, where $K$ is a constant used to make longer distances have a larger measurement error. So and Chan [22], Wei et al. [23] and Qin et al. [24] take the quality of the channel into consideration and replaced constant $\sigma$ with the signal-to-noise ratio (SNR) in the equation of the variance of the zero-mean white Gaussian process with variance. The equation they used is as follows: $\sigma^2 = \frac{d_{in}^2}{SNR}$, where $SNR$ is the signal-to-noise ratio and $d_{in}$ is the actual distance.

3.1. Enhancing error modeling used in localization

In each of the previous works [15,17–20,22–24] Gaussian noise was added to the actual distance in a manner similar to the computation in Equation 3.1 but with different variances. The Gaussian error introduced to the estimated distance, however, is added to the displacement of SN location, that is, in the $x$ and $y$ co-ordinate, not to the absolute distance, that is, $d$, between the SNs. Figure 3 shows that the error added to the estimated distance $d_{ij}$ results from the displacement in both $x$ and $y$ of the SN location.

If we assume that the displacement in $x$ and $y$ follows the Gaussian distribution:

$$x_{est} = x_j + x_{err} \quad \text{where } x_{err} \sim \mathcal{N}(0, \sigma^2),$$

(5)

and,

$$y_{est} = y_j + y_{err} \quad \text{where } y_{err} \sim \mathcal{N}(0, \sigma^2),$$

(6)

then the estimated distance can be represented as follows:

$$d_{ij} = \sqrt{(x_{est} - x_i)^2 + (y_{est} - y_i)^2},$$

(7)
By substituting Equations 5 and 6 in Equation 7, we yield:

$$d_{ij} = \sqrt{(x_j - x_i + x_{err})^2 + (y_j - y_i + y_{err})^2}$$

where $x_{err}$ and $y_{err} \sim \mathcal{N}(0, \sigma^2)$.

From the definition of the Rayleigh, $\gamma \sim \text{Rayleigh}(\sigma)$ if $\gamma = \sqrt{X^2 + Y^2}$, where $X$ and $Y \sim \mathcal{N}(0, \sigma^2)$ are independent normal random variables. This is the case in Equation 8. Therefore $d_{ij} \sim \text{Rayleigh}(\sigma_{ij})$.

To validate that $d_{ij} \sim \text{Rayleigh}(\sigma_{ij})$, we use real data provided by Patwari et al. [25]. The choice of using this data set is motivated by the enhancements they did for the RSSI model to estimate the distance between SNs and they reached 2-m location error using the RSSI. In their experiment, they used a wideband DSSS transceiver (Sigtek ST-515), and maintained an SNR $>25$ dB during the experiment to reduce the effect of the noise and ISM-band. They modeled the wideband radio channel impulse response as a sum of attenuated signal, phase-shifted and multi-path [26,27].

Patwari et al. deployed 44 SNs within a $14 \times 13$ m area as shown in Figure 4. The distance between each SN pair is estimated using RSSI measurements to have in total $44 \times 43 = 1892$ measurements. The histogram of the absolute noise, that is, $\epsilon_{ij}$, resulting from estimating the distance between the SNs is plotted as shown in Figure 5(a). The output of the histogram follows a Gaussian distribution with $\mu = 0.4$ and $\sigma^2 = 8.41$. The data can be replicated easily using the same values as shown in Figure 5(b). Patwari et al. have shown a similar finding. They use this finding and suggest that the added noise to actual distance follows the Gaussian distribution. The authors therefore added the generated noise to the absolute distance to represent the estimated distance. However, when the histogram of the estimated distance is plotted using the real data, that is, a histogram of the actual
distance with the noise $d_{i,j} = r_{i,j} + \epsilon_{i,j}$, the histogram is observed to follow a Rayleigh distribution with $\sigma = 6.6$, as shown in Figure 6(a).

When we replicate the estimated distance by adding Gaussian noise resulting from Figure 5(b) to the actual distance using the following equation 3, the estimated distance follows the normal distribution with $\mu = 7.7$ and $\sigma^2 = 4.7$ as shown in Figure 6(b). However, by using Equation 8, we obtain a Rayleigh distribution with $\sigma = 6.72$ as shown in Figure 6(c). The histogram resulting using Equation 8 therefore gives a more realistic representation of the error, as it gives an almost similar distribution resulting from using the estimated distances using real measurements. This would indicate the added noise is not a pure Gaussian distribution and it is affected by the change in both $x$ and $y$ co-ordinates.

To test the validity of fitting, the empirical histogram to the standard Rayleigh distribution, we performed the chi-square test on the estimated distance provided by Patwari et al. [25] shown in Figure 7. As illustrated, the Rayleigh distribution represents the data more accurately than the Gaussian distribution. Thus, the estimated distance between SNs follows the Rayleigh distribution not the Gaussian distribution.

### 3.2. The effect of using Rayleigh distribution on localization error

We preformed an experiment using simulation to study the effect of adding Rayleigh distribution using Equation 8 to the distance error between SNs instead of adding Gaussian using Equation 3. We use ns-3 to study the effect of using Normal verses Rayleigh distribution on a multi-hop localization technique that uses DV-Distance [28]. 500 SNs are randomly placed an area $200 \times 200$ m$^2$. In order to minimize the effect of placing the anchor nodes on the transmission range, we placed four anchor nodes at the corner of the simulated area. The transmission range of the anchors and SNs are increased gradually from 20 to 100 m in increments of 20 m. The same $\sigma^2$ is used for both Gaussian and Rayleigh distribution.

---

1An abridged exposition limited to preliminary results was in made in [29].
Characterizing Multi-hop Localization for Internet of Things

W. Ibrahim et al.

Figure 8. The relation between transmission range and localization error (both in meters). Number of anchors = 4 at the edge of the studied area.

Rayleigh distribution. All the results are from an average of 10 runs.

Figure 8 shows the relation between increasing the transmission range and localization error. When the error is small ($\sigma^2 = 2$), the localization error is the same for both Gaussian and Rayleigh distribution as shown in Figure 8(a). As we increase the transmission range, however, the localization error increases until the transmission range is 20 m. At 20 m, the density of SNs is less and, in turn, the SNs require a larger number of hops to reach the anchor node. In the next section, we study the effect of transmission range on localization error in detail.

When the error is large ($\sigma^2 = 8$) and the transmission range is small (20 m), the difference between Gaussian and Rayleigh distribution is at maximum (12 m). As we increase the transmission range, the difference between Gaussian and Rayleigh distribution decreases, until both Rayleigh and Gaussian distribution have similar localization errors when the transmission range = 60 m, as shown in Figure 8(b).

The above indicates that by decreasing the transmission range, the difference between using Gaussian and Rayleigh increases as the variance of the error increases. Also, when the transmission range is large both Rayleigh and Gaussian give similar localization error as the variance of the error increases.

To validate our findings and check the effect of the variance on the localization error, we increased the value of the variance gradually while keeping the transmission range constant. When the transmission range = 20 m the difference between using Gaussian and Rayleigh increases rapidly until the difference reaches 12 m, as shown in Figure 9(a). However, when the transmission range = 60 m, the difference between using Gaussian and Rayleigh increases slowly until the difference is 2 m, as shown in Figure 9(b). Results in Figure 9 validate the findings in Figure 8.

4. THE EFFECT OF TRANSMISSION RANGE ON LOCALIZATION ERROR

We use our realistic model for distance estimation error to study the relation between the transmission range and localization error. Localization estimation errors can be a result of either extrinsic or intrinsic errors [30]. An intrinsic error is usually caused by the imperfections of the sensor hardware and/or software, while an extrinsic error is attributed to the physical effects on the measurement channel and multi-hop communication. Savvides et al. [30] studied a range of intrinsic error characteristics for different measurement technologies. They did not, however, study the effect of extrinsic error.

Our focus in this paper is studying the effects of extrinsic errors on multi-hop localization. Such errors result from estimating the distance of hops between SNs. A common belief held by researchers is that by increasing the number of hops between the anchor nodes and SNs, the localization error will increase. To the best of our knowledge, no data has been published regarding the relationship between the number of hops and localization errors.
4.1. Performance evaluation setup for transmission range

We consider three different deployment scenarios to study the effect of SNs transmission on localization error. The scenarios are shown in Figure 10. The first scenario represents a practical one where SNs are deployed randomly in the sensed area as shown in Figure 10(a) with a total of 1000 SNs. In order to obtain a better understanding of the effects of the error, we used a controlled scenario in the second and third deployments. In the second scenario, the SNs are deployed in a fixed grid with a 20 m step between SNs with a total of 500 SNs as shown in Figure 10(b). In the third, the SNs are deployed in a dynamic grid, where the number of SNs and their placement are changed based on the SN transmission range as shown in Figure 10(c).

The WSN is deployed with N unlocalized SNs and four anchor nodes. Different numbers and locations of anchor nodes have been experimented with and resulted in similar observations. The simulation area is set to 200×200 m². To minimize the effect of the anchor nodes location on localization error and to overcome the collinearity problem, the four anchor nodes are placed at the four corners of the simulation area. The collinearity problem appears when the anchor nodes are on the same line [31]. In this case, it is hard to identify whether the SN is on the left or right of the anchor nodes, which causes the location of the SN to be flipped [32]. The measurement noise used in estimating the distance for DV-Distance is a zero-mean white Gaussian process with variance $\sigma_i^2 = d^2/\text{SNR}$ added to the x and y coordinates, as discussed in Section 3. We use two values for SNR: SNR = 10 dB is used to represent a communication channel with a high noise and SNR = 30 dB is used for a communication channel with low noise. All results are averages of 10 different independent runs with distinct random seeds.

The number of SNs used in the dynamic grid is calculated as follows. (The width and length of the simulation area, and the SN’s transmission range, are respectively shorthanded by SimWidth, SimLength, and TxRange.)

$$\frac{\left(\sqrt{(\text{SimWidth})^2 + (\text{SimLength})^2}\right)^2}{\text{TxRange}} + 2$$ (9)

while the vertical step is calculated as follows.

$$= \frac{\text{SimWidth} \times \text{TxRange}}{\sqrt{(\text{SimWidth})^2 + (\text{SimLength})^2} + \text{TxRange}}$$ (10)

and horizontal step is calculated in the following manner.

$$= \frac{\text{SimLength} \times \text{TxRange}}{\sqrt{(\text{SimWidth})^2 + (\text{SimLength})^2} + \text{TxRange}}.$$ (11)

To make the fixed grid deployment more realistic, we assume that there is randomness in the deployment of individual SNs modeled by a random error disk with a radius of 1 m [33]. The second scenario represents a dense deployment of SNs, in which more SNs are covered by increasing the transmission range. The third scenario represents a sparse deployment where SNs are only covered by the border of the maximum transmission range. In Figure 10(b), for example, when the transmission range is 50 m, there are more than 50 SNs covered with 3 or 4 SNs within 10 m.
from the anchor nodes; around 10 SNs that are near 50 m away; and the rest of the SNs are in between. This variation of distances increases the overall localization error. Meanwhile, in Figure 10(c), we only have three nodes which are approximately 50 m away from the anchor node when the transmission range is 50 m.

### 4.2. Performance evaluation for transmission range

All our experiments strongly indicate that DV-Hop localization outperforms the DV-Distance localization when the SNR is low. Previous works also conclude similar findings, for example, [34]. However, the performance of DV-Distance improves when the SNR increases.

In the following subsections, we elaborate on our findings made using random, fixed grid and dynamic grid deployment scenarios.

#### 4.2.1. Random deployment.

In the first scenario, we deployed 500 SNs randomly. Figure 11 shows that using shorter transmission range yields a lower error for both DV-Hop and DV-Distance localization techniques. The average error increases when we increase the transmission range of the SNs for both DV-Hop and DV-Distance except when the transmission range is 20 m.

The increase in error, of DV-Hop, is greater than in DV-Distance, when we increase the transmission range (the error in DV-Distance with 30 dB is almost constant). The error of DV-Hop increased by 11% while the error of DV-Distance increased by 8% for SNR = 10 dB and 2% for SNR = 30 dB when the transmission range of the sensor nodes increases from 40 to 100 m. Figure 12 illustrates why the localization error increases when the transmission range of the SN is increased to 20 m.

In Figure 12(a) the transmission range of the SNs is set to 20 m and the distance between anchor node A1 and SN4 is 48 m. Thus, the packets that are used to estimate the distance between A1 and SN4 are from SN1, SN2, and SN3, respectively. In DV-Distance, by adding the actual distance of the four hops, we obtain a distance of 65 m, where the actual distance is 48 m. The same effect applies when the DV-Hop is used, since we have four hops. Assuming the average hop distance to be 15 m, then the total estimated distance is 60 m. However, when the transmission range is 40 m, the estimated distance is estimated through only two hops, which decreases the estimation error. Increasing the density of SNs by adding an extra SN5 between A1 and SN3 as shown in Figure 12(b), the estimated distance becomes 51 m if we used DV-Distance, and 45 m if we used DV-Hop. These estimates are closer to the actual distance.

To validate our assumption, we repeated the experiment using 400, 500, and 600 SNs in the same area. Figure 13 shows that as we increase the number of SNs, the localization error increases when the transmission range of the SNs is equal to 20 m for both DV-Hop and DV-Distance. However, the change in the localization error is insignificant when the transmission range is larger than 40 m. This result shows that in highly dense deployments, it is better to have a larger number of hops with shorter transmission

![Figure 11. The localization error (in meters) using random deployment of SNs.](image1)

![Figure 12. The effect of dense and sparse deployment on localization error.](image2)
ranges (approximately 40 m) than to have a small number of hops with larger transmission ranges.

The earlier findings contrast to a commonly held belief that using shorter hops would give less error. In order to more closely understand this behavior, we use the two controlled experiments presented in section 4.1, namely fixed grid, and dynamic grid. The details of these evaluations follow.

4.2.2. Fixed grid.

In this scenario, SNs are deployed in a fixed grid in the sensed area. Figure 14 shows the error for localizing SNs when the SNs are placed in a fixed grid. Neither the number of SNs nor the distance between the SNs is changed when we increase the transmission range for SNs. The error of DV-Hop localization is increased by 27% when the transmission range is increased from 20 to 100 m. Whereas, the error of DV-Distance only increases by 12% and 10% when SNR = 10 dB and 30 dB, respectively. This result supports the previous finding that using shorter transmission for SNs decreases the overall localization error. The only difference with the previous finding is at transmission range = 20 m using fixed grid deployment gives a lower localization error than using random deployment. The reason is in fixed grid the shortest path is always granted, as shown in Figure 12(b).

To explain why the localization error for DV-Hop increases as we increase the transmission range (decreasing the number of hops), we can consider the scenario in Figure 15. Assuming the distance between $A_1$ and $A_2$ is 100 m and one SN between the two anchor nodes shown in Figure 15(a), both anchor nodes will ideally calculate
Characterizing Multi-hop Localization for Internet of Things

W. Ibrahim et al.

that the average hop distance to be 50 m as the packet reached $A_2$ in two hops. Thus, when $SN$ multiples the average hop distance which is 50 by the number of hops which is 1, the result is 50 m that leads to lower estimation error. This, however, is not the case in a dense deployment. Figure 15(b) shows that there are 5 $SN$s between the two anchor nodes. The minimum number of hops between $A_1$ and $A_2$ is two hops through $SN_3$. Thus, the average hop distance is 50 m. When $SN_1$, $SN_2$, and $SN_3$ estimate the distance between themselves and $A_1$ by multiplying the average hop distance which is 50 m by the number of hops which is one, we obtain 50 m that is an accurate estimation for only $SN_3$.

4.2.3. Dynamic grid.

In this scenario, we use a dynamic grid, where the spacing between $SN$s increases as we increase the transmission range of $SN$s. Figure 16 shows that as we increase the transmission range, the overall localization error lessens for both DV-Hop and DV-Distance using $SNR = 30$ dB and descends for DV-Distance using $SNR = 10$ dB. This result was as expected, however, the error increased dramatically when we used DV-Distance using $SNR = 10$ dB. The previous results show that in sparse deployments, it is better to have a lower number of hops with longer transmission ranges.

To explain why DV-Distance with high noise (low $SNR$ value) has high error in a sparse deployment when using longer transmission ranges, refer to Figure 17. When the distance between the anchor nodes and the $SN$ is accurately estimated, the area we perform multilateration is relatively small, which decreases the localization error as shown in Figure 17(a). However, when the error for the estimated distance is high, the area on which we perform multilateration is larger. This increases the error for the estimated location. Thus, when we use DV-Distance in an environment that has a high noise level in the channel, it is better to use shorter hops with a shorter transmission range than to use a long transmission range.

Previous results show that to achieve low localization error in highly dense deployments it is better to use short transmission ranges for the $SN$s. However, by decreasing the transmission range, the number of packets transmitted in the network will increase as shown in Figure 18. The
number of packets sent decreased by 62% and 48% for DV-Distance and DV-Hop, respectively, as we increased the transmission range from 20 to 100 m.

As the number of packets transmitted increases in the network, the number of collisions will increase which will have a direct impact on the overall localization error. Figure 18 shows that DV-Hop generates almost twice the packets generated by DV-Distance for a transmission range of 100 m. Also, as we increase the transmission range, the number of packets transmitted decreases gradually. Thus, it is important to study the effect of packets transmitted during the localization process. In the next section, we perform this study to characterize the overhead of the multi-hop localization.

5. CHARACTERIZING THE LOCALIZATION OVERHEAD

In the previous Section, we show that to achieve low localization error in highly dense environments, it is better to use short transmission ranges for the SNs. However, using short transmission range for SN in high dense deployments increase the number of packets transmitted in the network, as in previous localization techniques, when an SN receives a packet containing the anchor node’s location, the SN forwards this packet to the surrounding SNs at once. This behavior (receive then forward) generates huge traffic from SNs that affects the performance of the transmission. This may be acceptable in sparse WSN deployments, but it adversely affects the performance of the communication in high dense deployments. The reason for the huge traffic generated is that the packet is broadcasted to all neighboring SNs. Thus, it is important to reduce the number of packets generated while implementing the localization techniques.

5.1. Packet aggregation

To reduce the amount of traffic generated during the localization process, we propose that SNs only transmit localization packets in a predefined time. This means when a SN receives a packet, it first stores the packet and aggregates it with the previous received packets. When the timer for the transmitter expires the SN forwards the aggregated packet. In connectivity based localization techniques, SNs aggregate two types of packets: the first for the anchor location packet, while the second is for the average hop packet. After applying this behavior (receive – save – forward), the number of packets sent is reduced dramatically.

Figure 19 shows the number of packets sent to localize both DV-Hop and DV-Distance. The number of packets decreased for DV-Distance by 70% when the transmission range of SNs is 20 m and 50% when the transmission range of SNs is 100 m. For DV-Hop, the number of packets decreased by 55% and 40% when the transmission range of SNs is set to 20 and 100 m, respectively. From the earlier result, it is clear that the aggregation technique we propose dramatically decreases the number of packets sent without affecting the localization error.

In the rest of the section, we investigate the impact of the aggregation concept on the performance of localization technique in multi-hop environment. We used connectivity based localization technique for two reasons. The first reason is the results from the previous section shows that using connectivity based localization techniques performs better than using range based localization techniques in noisy deployments. The second reason, connectivity based localization techniques generate twice the number of packets generated by distance based localization techniques.

5.2. Performance evaluation for packet aggregation

In this setup, SNs are uniformly distributed in a simulated area of 100 × 100 unit blocks. The choice of area is made to ensure the full connectivity of the WSNs when even the least number of SNs is distributed, which is 25 SNs. The transmission range for sensor and anchor nodes are set to 40 m. In the beginning, 25 anchor nodes in addition to one sink SN are used. Simulations are made to run for a period of 100 s. Impact of scale is evaluated through varying size of the sensor network from 25 SNs to 200 in increments of 25 SNs. The performance metrics are averaged over 10 different topology runs generated using distinct random seeds.

In our evaluations, we investigate several aspects of operation. We describe below our findings giving performance metrics that best illustrate the impact of scale and...
mobility on wireless localization algorithms. The metrics are defined below.

- **Total number of packets sent**: comprises the total number of packets transmitted for two types of packets. The first type includes weight update packets, which are generated by anchor nodes and forwarded by the SNs. The second includes the weight table packets, which are generated and forwarded by SNs.
- **Total number of packets received**: the total number of packets received at anchor nodes and SNs.
- **Total number of packets dropped**: comprises the packets dropped due to redundancy, that is, a weight update or weight table packets that have already been received.
- **Total number of packet collisions**: reports the number of packet collisions at the MAC layer.
- **Mean error in Euclidean distance**: reports the mean error in computing the Euclidean distance between estimated location and actual location for each SN.
- **Percentage of unlocalized SNs**: localizing any SN requires at least three reports. This metric reports the percentage of the SNs for which at least three distance reports were not received.

In order to investigate the effect of scalability on localization, the number of SNs is increased from 25 to 200 by increments of 25 SNs. The weight parameter, \( k \), is set to four to allow SNs to connect to anchor nodes using four hops. The impact of network size is evaluated in Figure 20.

In each, the sub-figures illustrate the impact on the total numbers of (i) packets sent (ii) packets received (iii) packets dropped and (iv) packet collisions. Recall that the aggregation simply involves the use of a hold time prior to forwarding received packets and tables while aggregating them in one transmission. In the static setting, this modification reduces the number of sent and forwarded packets from almost 300,000 to less than 20,000 at 200 SNs as shown in Figure 20(a), which leads to relatively decrease the number of packet collisions experienced in the network as shown in Figure 20(d).

The value of the different metrics consistently increases as the number of SNs is increased. This is observed in both static and mobile settings. When aggregation is not employed, despite the redundancy check, a great number of packets are generated. For example, at 125 SNs, 150,000 packets are either sent or forwarded in the static setting as shown in Figure 20(a), while almost 2,000,000 packets are received as shown in Figure 20(b). A similar trend can also be observed in the mobile setting (100,000 to 800,000 at 125 SNs). Note that this great discrepancy is due to the broadcast nature of the wireless medium, that is, all receivers in the vicinity of a sender or forwarder SN receive the broadcasted packet. This very nature also justifies the numbers of packets dropped and collisions as shown in Figures 20(c) and 20(d). Lesser numbers are also experienced in the mobile setting as SNs move beyond each other’s vicinity more frequently.

**Figure 20.** The effect of increasing the number of SNs on the evaluation metrics.

### 5.3. Discussion and further observations

The earlier results indicate a strong need for carefully designed localization algorithms. Even for the proposed aggregation technique, further optimization remains possible. For example, the exact relationship between hold time and mobility need to be further explored. Another issue of concern is that of MAC operation, especially when it comes to functionality such as localization. This impact of MAC is particularly apparent in Figure 20(d) that shows the number of collisions in MAC layer.

It is also important in optimizing or re-designing the localization procedures to keep in mind the ultimate objective, which is accurate localization. In investigating the suggested aggregation, we explored its impact on localiza-
Characterizing Multi-hop Localization for Internet of Things

6. CONCLUSION

Evaluating multi-hop localization techniques in large areas is expensive and time consuming. Creating a realistic simulation model is required. Using a network simulation provides a rich opportunity for efficient experimentation, as simulation gives practical feedback before designing real-world systems. This allows us to determine the correctness and efficiency of the localization techniques before the actual deployment of the SNs. Thus, we build a realistic simulation model to use to study and investigate the behavior of multi-hop localization techniques in large scale deployments and propose new schemes for multi-hop localization. After that, we characterize the error behavior in multi-hop localization. There has been a belief in the literature that the smaller the number of hops, the lower the error is as discussed in [35]. We assess this belief for representative generic schemes of range-based and range-free localization. We show that in dense deployments, it is better to decrease the transmission to achieve low localization error, which generates a large amount of traffic. Thus, we propose a new solution that substantially decreases the number of packets generated in highly dense deployments during the localization process. The proposed solution decreased the number of messages exchanged by almost 70% for DV-Distance and 55% for DV-Hop.

While some possibilities of future works are discussed in Section 5.3, an extensive empirical validation for our localization error model remains. This validation will complement our foundational validation offered in Section 3, and offer a stronger basis for the design of wireless multi-hop localization algorithm. Meanwhile, considerations for sensor/thing mobility (along with reasonable model for localization error) are to be investigated, especially for possibilities of occasional sensor/thing collinearity with localization anchors.

ACKNOWLEDGMENT

This work is supported by the Roadway Transportation & Traffic Safety Research Center, United Arab Emirates University through project number 31R014-Research Center-RTT SRC-4-2013. The work is also supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant number CRDPJ445517.

REFERENCES

Walid M. Ibrahim received the BSc degree in Computer Science from Cairo University in 2004, the MSc degree in Information technology from Cairo University in 2008, and the PhD degree in Computer Science from Queen’s University, Canada, in 2014. He is currently an Assistant Professor at the Faculty of Computers and Information at Cairo University. His current research interests are in wireless communications, localization, smart environments, wireless sensors networks in harsh environments, security and Internet of Things.

Dr. Najah Abu Ali is currently an Associate Professor at the Faculty of Information Technology in the United Arab Emirates University. She earned her PhD at the Department of Electrical and Computer Engineering at Queen’s University in Kingston, Canada; specializing in resource management in computer networks. Her MSc and BSc were both attained in Electrical Engineering at the University of Jordan. Her general research interest includes modelling wireless communications, resource management in wired and wireless networks, and reducing the energy requirements in wireless sensor networks. More recently, she has strengthened her focus on the Internet of Things, particularly at the nano-scale communications, in addition vehicle-to-vehicle networking. Her work has been consistently published in key publications venues for journals and conference. She has further co-authored a Wiley book on 4G and beyond cellular communication networks. She has also delivered various seminar and tutorials at both esteemed institutions and flagship gatherings. Dr. Najah has also been awarded several research fund grants, particularly from the Emirati NRF/United Arab Emirates University funds and the Qatar National Research Foundation.

Hossam Hassanein is a leading authority in the areas of broadband, wireless, and mobile networks architecture, protocols, control, and performance evaluation. His record spans more than 500 publications in journals, conferences and book chapters, in addition to numerous keynotes and plenary talks in flagship venues. Dr. Hassanein has received several recognitions and best papers awards at top international conferences. He is also the founder and director of the Telecommunications Research Lab at Queen’s University School of Computing, with extensive international academic and industrial collaborations. He is a senior member of the IEEE and is a former chair of the IEEE Communication Society Technical Committee on Ad hoc and Sensor Networks. Dr. Hassanein is an IEEE Communications Society Distinguished Speaker (Distinguished Lecturer 2008–2010).

Abd-Elhamid M. Taha is currently an Assistant Professor in Electrical Engineering at Alfaisal University, his BSc (Honours) and MSc in Electrical Engineering were earned at Kuwait University in 1999 and 2002, and his PhD in Electrical and Computer Engineering was earned Queen’s University, Canada, in 2007. He also holds an adjunct post at the School of Computing at Queen’s. His general research interest is in the area of computer networks and communications. His current efforts span simplified device localization, mobile security in the Internet of Things, and modelling in networked cyber-physical systems. He has written and lectured extensively on broadband wireless networks, focusing on radio resource management techniques. He is also the co-author of the book LTE, LTE-Advanced and WiMAX: Toward IMT-Advanced Networks (Wiley) and a presenter for several tutorials at key IEEE Communications Society events. Dr. Taha is a Senior Member of the IEEE Communications Society, as well as a member of the ACM.