Hop-Distance Estimation in Wireless Sensor Networks with Applications to Resources Allocation

Liang Zhao and Qilian Liang

Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76010, USA

Received 22 May 2006; Revised 7 December 2006; Accepted 26 April 2007

Recommended by Huaiyu Dai

We address a fundamental problem in wireless sensor networks, how many hops does it take a packet to be relayed for a given distance? For a deterministic topology, this hop-distance estimation reduces to a simple geometry problem. However, a statistical study is needed for randomly deployed WSNs. We propose a maximum-likelihood decision based on the conditional pdf of \( f(r \mid H_i) \). Due to the computational complexity of \( f(r \mid H_i) \), we also propose an attenuated Gaussian approximation for the conditional pdf. We show that the approximation visibly simplifies the decision process and the error analysis. The latency and energy consumption estimation are also included as application examples. Simulations show that our approximation model can predict the latency and energy consumption with less than half RMSE, compared to the linear models.

Copyright © 2007 L. Zhao and Q. Liang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The recent advances in MEMS, embedded systems, and wireless communications enable the realization and deployment of wireless sensor networks (WSN), which consist of a large number of densely deployed and self-organized sensor nodes [1]. The potential applications of WSN, such as environment monitor, often emphasize the importance of location information. Fortunately, with the advance of localization technologies, such location information can be accurately estimated [2–5]. Accordingly, geographic routing [6–8] was proposed to route packets not to a specific node, but to a given location. An interesting question arises as “how many hops does it take to reach a given location?” The prediction of the number of hops, that is, hop-distance estimation, is important not only in itself, but also in helping, estimate the latency and energy consumption, which are both important to the viability of WSN.

The question could become very simple if the sensor nodes are manually placed. However, if sensor nodes are deployed in a random fashion, the answer is beyond the reach of simple geometry. The stochastic nature of the random deployment calls for a statistical study.

The relation between the Euclidean distance and network distance (in terms of the number of hops), also referred to as hop-distance relation, catches a lot of research interest recently. In [9], Huang et al. defined the \( \Gamma \)-compactness of a geometric graph \( G(V,E) \) to be the minimum ratio of the Euclidean distance to the network distance,

\[
y = \min_{ij \in V} \frac{d(i,j)}{h(i,j)},
\]

where \( d(i,j) \) and \( h(i,j) \) are the Euclidean distance and network distance between nodes \( i \) and \( j \), respectively. The constant value \( y \) is a good lower bound, but might not be enough to describe the nonlinear relation between Euclidean distance and network distance. In fact, their relation is often treated as linear for convenience, for example, \( r/R + 1 \) is widely used to estimate the needed number of hops to reach distance \( r \) given transmission range \( R \). Against this simple intuition, the relation between Euclidean distance and network distance is far more complex. Fortunately, a lot of probabilistic studies have been applied to this question. In [10], Hou and Li studied the 2D Poisson distribution to find an optimal transmission range. They found that the hop-distance distribution is determined not only by node density and transmission range, but also by the routing strategy. They showed results for three routing strategies, most forward with fixed radius, nearest with forward progress, and most forward with variable radius. Cheng and Robertazzi in [11] studied the one-dimensional Poisson point and found the pdf of \( r \) given the number of hops. They also pointed out that the 2D Poisson
point distribution is analogous to the 1D case, replacing the length of the segment by the area of the range. Vural and Ekici reexamined the study under the sensor networks circumstances in [12], and gave the mean and variance of multihop distance for 1D Poisson point distribution. They also proposed to approximate the multihop distance using Gaussian distribution. Zorzi and Rao derive the mean number of hops of the minimal hop-count route through simulations and analytic bounds in [8]. Chandler [13] derives an expression for \( t \)-hop outage probability for 2D Poisson node distribution. However, Mukherjee and Avidor [14] argue that one of Chandler’s assumptions is relaxed, and thus his expression is in fact a lower bound on the desired probability. Using the same assumption, they also derive the pdf of the minimal number of hops for a given distance in a fading environment. Although these analytic results are available in the literature, their monstrous computational complexity limits their applications. Therefore, we try to approximate the hop-distance relation and simplify the decision process and error analysis in this paper. Considering the application of resource allocation, only large-scale path loss is considered, and thus the fading is ignored.

The rest of this paper is organized as follows. The number of hops prediction problem is addressed and solved in Section 2. Since this problem has no closed-form solution, we propose an attenuated Gaussian approximation and show how to simplify the error analysis in Section 2.1. Application examples are shown in Section 3. Section 4 concludes this paper.

## 2. ESTIMATION OF NETWORK DISTANCE BASED ON EUCLIDEAN DISTANCE

Suppose the sensor nodes are placed on a plane at random, and \( N(A) \), the number of nodes in a given area \( A \), follows two-dimensional Poisson distribution with average density \( \lambda \). The problem of interest is to find the number of hops needed to reach a distance \( r \) away. We can make a maximum-likelihood (ML) decision,

\[
\hat{H} = \arg \max f(r \mid H_n), \quad n = 1, 2, 3, \ldots
\]

where the event \( H_n \) can be described as “the minimum number of hops is \( n \) from the source to the specific node at Euclidean distance \( r \).” In the following discussion, we are trying to approximate \( f(r \mid H_n) \) for 2D Poisson distribution. Note that \( r < R \) implies \( H_1 \), that is, the specific node is within one hop from the source. We are more interested in multiple-hop distance relation, especially when \( n \) is moderately large.

### 2.1. Attenuated Gaussian approximation

Since \( f(r \mid H_n) \) is awkward to evaluate even using numerical methods, we use histograms collected from Monte Carlo simulations as substitute to the joint pdf. All the simulation data are collected from a scenario where \( N \) sensor nodes were uniformly distributed in a circular region of radius of \( R_{\text{bound}} \) meters. For convenience, polar coordinates were used. The source node was placed at \((0, 0)\). The transmission range was set as \( R \) meters. For each setting of \((N, R_{\text{bound}}, R)\), we ran 300 simulations, in each of which all nodes are redeployed at random. We ran simulations for extensive settings of node density \( \lambda \) and transmission range \( R \). Due to space constraints, only the histograms for \((N = 1000, R_{\text{bound}} = 200, R = 30)\) are plotted in Figure 1, which approximately shows that \( f(r \mid H_n) \) approaches the normal when \( H_n \) increases. Table 1 lists the first-, second-, third-, and fourth-order statistics of \( f(H, r) \).

![Figure 1: Histograms of hop-distance joint distribution (\(N = 1000, R_{\text{bound}} = 200, R = 30\)).](image)

Skewness is a third-order statistic used to measure of symmetry, or more precisely, the lack of symmetry. Skewness is zero for a symmetric distribution and positive skewness indicates right skewness while negatives indicates left skewness.

**Definition 1** (see [15]). For a given sample set \( X \),

\[
\begin{align*}
   m_3 &= \frac{\Sigma(X - \overline{X})^3}{n}, \\
   m_2 &= \frac{\Sigma(X - \overline{X})^2}{n},
\end{align*}
\]

where \( \overline{X} \) is the sample mean of \( X \), and \( n \) is the size of \( X \). Then a sample estimate of skewness coefficient is given by

\[
g_1 = \frac{m_3}{m_2^{3/2}}.
\]
Kurtosis is a fourth-order statistic indicating whether the data are peaked or flat relative to a normal distribution.

**Definition 2** (see [15]). A sample estimate of kurtosis for a sample set $X$ is given by

$$g_2 = \frac{m_4}{m_2^2} - 3,$$

where $m_4 = \Sigma(X - \bar{X})^4/n$ is the fourth-order moment of $\bar{X}$ about its mean.

Skewness and kurtosis are useful in determining whether a sample set is normal. Note that the skewness and kurtosis of a normal distribution are both zero; significant skewness and kurtosis clearly indicate that data are not normal. Table 1 clearly shows that the skewness and kurtosis satisfy the Gaussianity condition within tolerance of error. Furthermore, The postulated distribution and histogram are drawn together in Figures 2(a), 2(b), 2(c), and 2(d), which clearly shows a close match for each case. Also, note that $f(r \mid H_n)$ attenuates exponentially with $n$ increase, we need to introduce an attenuation factor to model this behavior.

Thus, the objective function can be approximated by

$$f(r \mid H_n) = \alpha^n \mathcal{N}(m_n, \sigma_n) = \frac{\alpha^n}{2\pi\sigma} e^{-(r-m_n)^2/2\sigma^2},$$

where $\alpha$ is the equivalent attenuation base, $m_n$ and $\sigma_n$ are the mean and standard deviation (STD), respectively. Since $f(r \mid H_n)$ attenuates with $n$ increasing, $\alpha$ must be less than 1. The specific values of these parameters can be estimated from simulations or computed numerically from the exact pdfs. Our extensive simulations show that even for only moderately large $H_n$, $f(r \mid H)$ has the following properties.

1. $\sigma_n \approx \sigma_{n-1}$, which means that the neighboring joint pdfs have similar spread.
Using property (1),

\[
d_n = \frac{m_{n+1}^2 - m_n^2 - 2\sigma_n^2 \ln \alpha}{2(m_{n+1} - m_n)}.
\]  

For large density \( \lambda \), property (5) is applicable, (9) simplifies to

\[
d_n = \frac{\sigma_n^2 m_{n+1} + \sigma_{n+1}^2 m_n}{\sigma_n^2 + \sigma_{n+1}^2}.
\]  

Applying property (1) to (11),

\[
d_n = \frac{m_n + m_{n+1}}{2}.
\]  

No matter which approximate solution we choose for \( d_n \), the decision rule is given by

\[
r \overset{n \geq n+1}{\sim} d_n.
\]  

In other words,

we decide \( \hat{n} \) if \( d_{n-1} < r \leq d_n \).

\[ (14) \]

2.3. Error performance analysis

For our decision rule, a decision error occurs only when the required number of hops is \( n \), but our decision \( \hat{n} \neq n \). Thus, the probability of error for a specific \( r \) is

\[
p(\epsilon \mid r) = \sum_{n \neq \hat{n}} f(H_n \mid r),
\]  

where \( f(H \mid r) \) is related to \( f(r \mid H_i) \) by the Bayesian rule. The total probability of error is obtained by integrating (15) over all possible \( r \),

\[
p(\epsilon) = \int p(\epsilon \mid r)f_r(r)dr.
\]  

According to property (4), only \( f(r \mid H = n - 1) \) and \( f(r \mid H = n + 1) \) could have outstanding value over the decision region \([d_{n-1}, d_n]\),

\[
p(\epsilon) = \sum_{n=2}^{\infty} \int_{d_{n-1}}^{d_n} f(r \mid H = n-1)p(H_{n-1}) + f(r \mid H = n+1)p(H_{n+1})dr
\]

\[
= \sum_{n=2}^{\infty} \alpha^{n-1} p(H_{n-1}) \left[ Q\left(\frac{d_{n-1} - m_{n-1}}{\sigma_{n-1}}\right) - Q\left(\frac{d_n - m_{n-1}}{\sigma_{n-1}}\right)\right]
\]

\[ (17) \]

\[
+ \alpha^{n+1} p(H_{n+1}) \left[ Q\left(\frac{m_{n+1} - d_n}{\sigma_{n+1}}\right) - Q\left(\frac{m_{n+1} - d_{n-1}}{\sigma_{n+1}}\right)\right].
\]  

\[ (17) \]
Note that
\[
\frac{d_n - m_{n-1}}{\sigma_{n-1}} - \frac{d_{n-1} - m_{n-1}}{\sigma_{n-1}} = \frac{d_n - d_{n-1}}{\sigma_{n-1}} \gg 1, \quad (18)
\]
therefore, \(Q((d_n - m_{n-1})/\sigma_{n-1})\) is negligible compared to \(Q((d_{n-1} - m_{n-1})/\sigma_{n-1})\). Similarly, \(Q((m_{n+1} - d_n)/\sigma_{n+1})\) is negligible. Equation (17) is approximated by
\[
p(\varepsilon) \approx \alpha^3 p(H_3) Q\left(\frac{m_3 - d_2}{\sigma_3}\right) + \sum_{n=3}^{\infty} \alpha^{n-1} p(H_{n-1}) Q\left(\frac{d_{n-1} - m_{n-1}}{\sigma_{n-1}}\right) + \alpha^{n+1} p(H_{n+1}) Q\left(\frac{m_{n+1} - d_n}{\sigma_{n+1}}\right)
\]
\[
= \alpha^2 p(H_2) Q\left(\frac{d_2 - m_3}{\sigma_2}\right) + \sum_{n=3}^{\infty} \alpha^n p(H_n) \left[Q\left(\frac{m_n - d_{n-1}}{\sigma_n}\right) + Q\left(\frac{d_n - m_n}{\sigma_n}\right)\right]. \quad (19)
\]
Substituting an appropriate solution of \(d_2\) into (19) would give us the probability of error within required accuracy. For example, if we choose (12),
\[
p(\varepsilon) \approx \alpha^2 p(H_2) Q\left(\frac{m_3 - d_2}{2\sigma_2}\right) + \sum_{n=3}^{\infty} \alpha^n p(H_n) \left[Q\left(\frac{m_n - d_{n-1}}{2\sigma_n}\right) + Q\left(\frac{d_n - m_n}{2\sigma_n}\right)\right]. \quad (20)
\]
Thanks to the Gaussian approximation, the error probability is given in forms of \(Q\) functions, which is tremendously simpler than the derivation from the original pdfs. This error process is general and applicable to other estimators. For example, even when we have to use a linear estimator due to limit of computation capacity, we can still use the above process to obtain the corresponding error probability.

3. APPLICATION EXAMPLES

We provide two application examples, latency and energy estimation, in this section. To emphasize the role of the number of hops in the estimation, we use general time and energy models. On how to derive the parameters such as \(T_{tx}, T_{rx}\) for a specific routing scheme, readers are referred to [16, 17].

3.1. Latency estimation

We use a simple time model, in which the latency increases linearly with the number of hops [18]. Suppose it takes \(T_{tx}, T_{rx}\) for a sensor node to process 1 bit of incoming and outgoing messages, respectively, and \(T_{pr}\) is the required time to transmit 1 bit of message through a band-limited channel. Therefore, the latency introduced for each hop is
\[
T_{hop} = T_{tx} + T_{pr} + T_{rx}. \quad (21)
\]

As shown in Figure 4, given the end-to-end distance \(r\), we can find the required number of hops \(n\) according to (13), thus, a good estimator of the total latency of an \(l\)-bit message is
\[
\ln T_{hop}. \quad (22)
\]

3.2. Energy consumption estimation

The following model is adopted from [19] where perfect power control is assumed. To transmit \(l\) bits over distance \(r\), the sender’s radio expends
\[
E_{tx}(l, r) = \begin{cases} 
I E_{elec} + I e_{fs} r^2, & r < d_0, \\
I E_{elec} + I e_{mp} r^4, & r \geq d_0,
\end{cases} \quad (23)
\]
and the receiver’s radio expends
\[
E_{rx}(l, r) = I E_{elec}. \quad (24)
\]
\(E_{elec}\) is the unit energy consumed by the electronics to process one bit of message, \(e_{fs}\) and \(e_{mp}\) are the amplifier factor for free-space and multipath models, respectively, and \(d_0\) is the reference distance to determine which model to use. In fact, the first branch of (23) assumes a free-space propagation and the second branch uses a path-loss exponent of 4. The values of these communication energy parameters are set as in Table 2.

![Figure 4: Time model.](image)

| Name  | Value   |
|-------|---------|
| \(r_0\) | 86.2 m  |
| \(E_{elec}\) | 50 nJ/bit |
| \(E_{DA}\) | 5 nJ/bit |
| \(e_{fs}\) | 10 pJ/bit/m\(^2\) |
| \(e_{mp}\) | 0.0013 pJ/bit/m\(^4\) |

Let \(s_n\) denote the single-hop distance from the \((n-1)\)th-hop to the \(n\)th-hop. Obviously, \(s_n \leq R\). In our experimental setting, \(R = 30m < d_0\) so that the free-space model is always used. This agrees well with most applications, in which multihop short-range transmission is preferred to avoid the exponential increase in energy consumption for long-range transmission. Naturally, the end-to-end energy consumption for sending \(l\) bit over distance \(r\) is given by
\[
E_{total}(l, r) = \sum_{i=1}^{\hat{m}} \{E_{tx}(l, r_i) + E_{rx}(l)\}, \quad (25)
\]
where $\hat{n}$ is the estimated number of hops for given $r$ and $r_1$ is the single-hop distance because the message is relayed hop by hop.

On the average,

$$E_{\text{total}}(l, r) = \hat{n}l(E_{\text{elec}} + \epsilon f_s E'[r_1^2] + E_{\text{elec}})$$

$$= \hat{n}l(2E_{\text{elec}} + \epsilon f_s (m_1^2 + \sigma_1^2)).$$

(26)

### 3.3. Simulation

We used the same scenario described in Section 2.1 and varied the node density $\lambda$ and transmission range $R$. In each simulation, the number of hops is estimated for each node using (11) and (13), and then the latency and energy consumption are estimated using (22) and (26), respectively. As comparison to our proposed statistic-based estimator, we choose a widely used linear estimator,

$$\text{linear estimator 1 } \hat{n} = \left\lfloor \frac{r}{R} \right\rfloor + 1,$$

$$\text{linear estimator 2 } \hat{n} = \left\lfloor \frac{r}{R} \right\rfloor + 2,$$

where $r$ is the given distance, $R$ the transmission range, and $\lfloor r/R \rfloor$ is the maximum integer less than $r/R$. We plot the average of latency and energy consumption in Figures 5(a) and 5(b) and the RMSE in Figures 6(a) and 6(b), respectively. The latency is plotted in units of $T_{\text{hop}}$ while the energy consumption in units of joules. The ripple shape of RMSE is due to the fact that decision errors occur more often in the overlapping zones of neighboring $f(r \mid H_i)$. Figure 5 shows that the linear estimator 1 performs well at the shorter range but suffers visibly at larger range, while the linear estimator does the opposite. The linear estimators, no matter what value their parameters take, may significantly underestimate or overestimate the latency and energy consumption as already pointed out in Section 2.1, while our statistic-based model keeps close to the actual latency and energy consumption at all ranges except for the border. This is also verified by Figure 6, which also shows that our model can reduce RMSE to at least half for both latency and energy consumption. These results show that linear models cannot identify network behavior accurately, as also confirmed by our extensive simulations for different settings of node density and transmission range, which is not shown here due to space constraints.

### 4. Conclusion

To address the fundamental problem “how many hops does it take for a packet to be relayed for a given distance,” we make both probabilistic and statistical studies. We proposed a Bayesian decision based on the conditional pdf of $f(r \mid H_i)$. Since $f(r \mid H_i)$ is computationally complex, we also proposed an attenuated Gaussian approximation for the conditional pdf, which visibly simplifies the decision process and the error analysis. This error analysis based on Gaussian approximation is also applicable to other estimators, including the linear ones. We also show that several linear models, though intuitively sound and widely used, may give significant bias error. Given as application examples, our approximation is also applied in the latency and energy consumption estimation in dense WSN. Simulations show that our approximation model can predict the latency and energy consumption.
Figure 6: Estimation RMSE: (a) latency; (b) energy consumption.

consumption with less than half RMSE, compared to the aforementioned linear models.

ACKNOWLEDGMENT

This work was supported by the US Office of Naval Research (ONR) Young Investigator Award under Grant N00014-03-1-0466.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” IEEE Communications Magazine, vol. 40, no. 8, pp. 102–114, 2002.
[2] H. Lim and J. C. Hou, “Localization for anisotropic sensor networks,” in Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05), vol. 1, pp. 138–149, Miami, Fla, USA, March 2005.
[3] A. Caruso, S. Chessa, S. De, and A. Urpi, “GPS free coordinate assignment and routing in wireless sensor networks,” in Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05), vol. 1, pp. 150–160, Miami, Fla, USA, March 2005.
[4] L. Fang, W. Du, and P. Ning, “A beacon-less location discovery scheme for wireless sensor networks,” in Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05), vol. 1, pp. 161–171, Miami, Fla, USA, March 2005.
[5] N. B. Priyantha, H. Balakrishnan, E. D. Demaine, and S. Teller, “Mobile-assisted localization in wireless sensor networks,” in Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05), vol. 1, pp. 172–183, Miami, Fla, USA, March 2005.
[6] R. Jain, A. Puri, and R. Sengupta, “Geographical routing using partial information for wireless ad hoc networks,” IEEE Personal Communications, vol. 8, no. 1, pp. 48–57, 2001.
[7] Y. Xu, J. Heidemann, and D. Estrin, “Geography-informed energy conservation for ad hoc routing,” in Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MOBICOM ’01), pp. 70–84, ACM Press, Rome, Italy, July 2001.
[8] M. Zorzi and R. R. Rao, “Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance,” IEEE Transactions on Mobile Computing, vol. 2, no. 4, pp. 337–348, 2003.
[9] Q. Huang, C. Lu, and G.-C. Roman, “Spatiotemporal multicast in sensor networks,” in Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys ’03), pp. 205–217, ACM Press, Los Angeles, Calif, USA, November 2003.
[10] T.-C. Hou and V. O. K. Li, “Transmission range control in multihop packet radio networks,” IEEE Transactions on Communications, vol. 34, no. 1, pp. 38–44, 1986.
[11] Y.-C. Cheng and T. G. Robertazzi, “Critical connectivity phenomena in multihop radio models,” IEEE Transactions on Communications, vol. 37, no. 7, pp. 770–777, 1989.
[12] S. Vural and E. Ekici, “Analysis of hop-distance relationship in spatially random sensor networks,” in Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC ’05), pp. 320–331, ACM Press, Urbana-Champaign, Ill, USA, May 2005.
[13] S. A. G. Chandler, “Calculation of number of relay hops required in randomly located radio network,” Electronics Letters, vol. 25, no. 24, pp. 1669–1671, 1989.
[14] S. Mukherjee and D. Avidor, “On the probability distribution of the minimal number of hops between any pair of nodes in a bounded wireless ad-hoc network subject to fading,” in
[15] G. Snedecor and W. Cochran, *Statistical Methods*, Iowa State University Press, Ames, Iowa, USA, 1989.

[16] M. Zorzi and R. R. Rao, “Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance,” *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 349–365, 2003.

[17] H. M. Ammari and S. K. Das, “Trade-off between energy savings and source-to-sink delay in data dissemination for wireless sensor networks,” in *Proceedings of the 8th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM ’05)*, pp. 126–133, ACM Press, Montreal, Quebec, Canada, October 2006.

[18] W. Ye, J. Heidemann, and D. Estrin, “Medium access control with coordinated adaptive sleeping for wireless sensor networks,” *IEEE/ACM Transactions on Networking*, vol. 12, no. 3, pp. 493–506, 2004.

[19] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “An application-specific protocol architecture for wireless microsensor networks,” *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.