

COM-LOC: A Distributed Range-Free Localization Algorithm in Wireless Networks

B.J.Dil¹, P.J.M.Havinga²

*Pervasive Systems, University of Twente
Zilverling building, 7522 NB Enschede, The Netherlands*

¹B.J.Dil@utwente.nl

²P.J.M.Havinga@ewi.utwente.nl

Ambient Systems

Colosseum 15d, 7521 PV Enschede, The Netherlands

¹Bram@ambient-systems.net

Abstract—This paper investigates distributed range-free localization in wireless networks using a communication protocol called sum-dist which is commonly employed by localization algorithms. With this protocol, the reference nodes flood the network in order to estimate the shortest distance between the reference and blind nodes. Existing localization algorithms that use this communication protocol only evaluate the shortest distance. Our approach is somewhat different in that we optimize the localization performance for this communication protocol. We present a new algorithm called COM-LOC which exploits a certain part of the information inherent in the protocol that other algorithms consider as redundant or false information. We show that the use of this additional part of information increases the performance compared to other range-free algorithms by 68% to 206%. Other comparisons with several RSS-based localization algorithms show that COM-LOC outperforms these algorithms under a wide range of conditions, while keeping the communication costs equal.

I. INTRODUCTION AND RELATED WORK

In the last few years, there is a growing interest in locating devices in wireless communication networks. In this paper, we focus on RSS-based localization. We distinguish three type of RSS-based localization algorithms, namely range-based, proximity-based and range-free localization.

Range-based localization algorithms assume that the signal strength decay over distance follows a distribution that is a priori known. This distribution is used for converting signal strength measurements into distance estimates. These distance estimates are then used for estimating the position (for example [13] and [16]).

Proximity-based localization algorithms assume that the signal strength decays over distance ([6] and [14]). The main difference with range-based algorithms is that they only use the order of RSS measurements. Therefore, proximity-based localization algorithms are not dependent on the goodness-of-fit of the RSS over distance distribution.

Range-free localization approaches use connectivity information ([5], [9], [11] and [12]). Existing localization algorithms based on connectivity assume that the transmission range is constant ([11]) or the deployment distribution is uniform and known a priori ([5], [9] and [12]). This means that the

performance depends on the difference between the expected and real values of the transmission range and deployment distribution.

Most existing localization algorithms in wireless networks are designed with the assumption that certain localization specific information is available. Afterwards, a communication protocol is designed to obtain this information. In this paper, we do it the other way around. We design a distributed range-free localization algorithm on the basis of sum-dist which is a communication protocol commonly employed by localization algorithms (as in [5], [9], [11], [12] and [15]). This means that we construct the Maximum Likelihood Estimator (MLE) for localization on the basis of the communication protocol. Theoretically, this maximum likelihood estimator should provide the best localization results for a given communication protocol. We present a new algorithm called COM-LOC using this new approach. Compared to other algorithms that use sum-dist, COM-LOC evaluates all information instead of only evaluating the shortest distance (as in [5], [9], [11], [12] and [15]). COM-LOC is a distributed range-free algorithm that adapts a grid-based Monte Carlo Localization (MCL) method, which has been successfully implemented in robotics localization (for example [3]) and in range-free and range-based localization in wireless sensor networks ([11] and [15]). This paper is organized as follows: after the problem formulation in Section II, Section III describes the model used for simulating the connectivity. Section IV shows how COM-LOC converts the information, obtained during the communication phase, into distance estimates and associated probabilities. Section V provides a description of COM-LOC. Section VI analyzes the localization performance of COM-LOC. In addition, this section also compares COM-LOC with DV-HOP([5]), ecolocation ([14]) and a modified version of the MLE described in [13]. Section VII presents the conclusion and future work.

II. PROBLEM FORMULATION

This section provides a formal description of the range-free localization problem. First consider a wireless network that

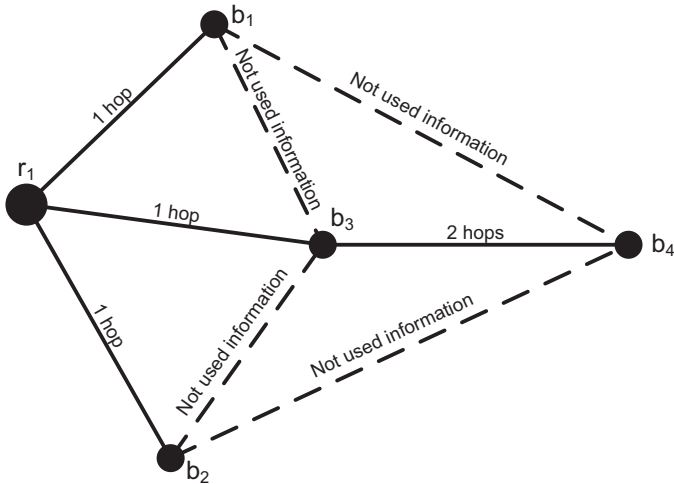


Fig. 1. Communication phase

consists of two type of nodes:

- *Reference nodes*: Reference nodes know their position in advance.
- *Blind nodes*: Blind nodes do not know their location in advance.

We address the problem of blind node localization on the basis of connectivity measurements using the following communication protocol (as in [11]). First, each reference node sends one message with its position and a hop distance set to one. Each receiving blind node stores the received reference position and hop count. Then the hop distance is increased by one and the message is forwarded. This is also the end of the communication phase. In general, this communication protocol is similar to sum-dist ([7]) with a time-to-live of two. We limit the time-to-live in order to keep the communication costs at the minimum and to allow localization functionality for mobile wireless networks. At the end of the communication phase, blind nodes have the following information components:

- 1 The blind nodes store a set of reference node positions that are one-hop-away. We represent this set by: $S \subseteq R$.
- 2 The blind nodes store a set of reference node positions that are two-hops-away. We represent this set by: $T \subseteq R$.
- 3 The blind nodes store the number of received messages from other blind nodes per reference node. We represent this number by: $nr \in r$, here $r \in R$.

Throughout this work, we abbreviate these information components by numbering them as in the previous enumeration ($\{1\}, \{2\}, \{3\}$).

Most existing distributed range-free localization algorithms use this communication protocol and only evaluate the shortest hop count for localization ([5], [9], [11] and [12]). This means that many received messages during the communication phase are considered useless and discarded. We use Figure 1 as an illustrative example in order to show this. Figure 1 represents a part of a wireless communication network. The black circles

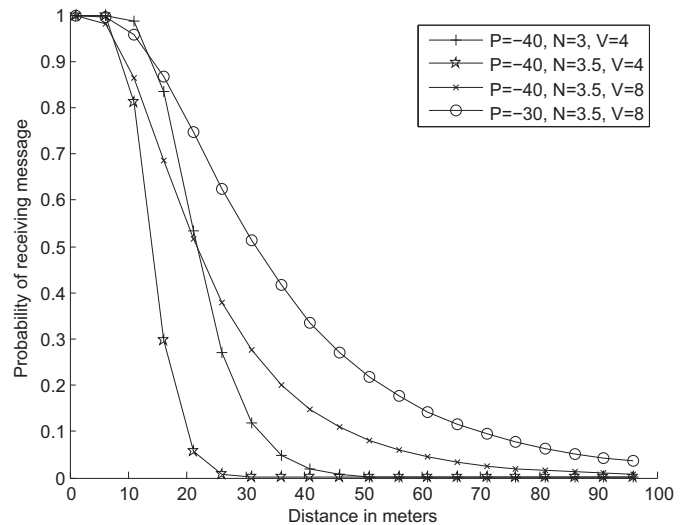


Fig. 2. Packet delivery rate as a function of the distance using the Log-Normal Shadowing Model

represent the nodes; r_1 represents a reference node and $b_1 \dots b_4$ represent the blind nodes. The solid and striped lines indicate that the nodes can communicate with each other. The text above the communication lines shows whether the received messages are processed by the existing algorithms:

- “1 hop” indicates that the blind node is one-hop-away from the reference node ($r_1 \in S$).
- “2 hops” indicates that the blind node is two-hops-away from the reference node ($r_1 \in T$).

We distinguish two type of messages that are not used by the existing algorithms:

- “redundant information”: The message from b_3 to b_4 indicate that b_4 is two-hops-away from reference node r_1 ($r_1 \in T$, information component 2). The messages from $\{b_1, b_2\}$ to b_4 also indicate this and are considered as redundant information. This means that information component 3 (if $nr > 1$) is considered as “redundant information” for reference nodes that are two-hops-away.
- “false information”: The message from r_1 to b_3 indicate that b_3 is one-hop-away from reference node r_1 ($r_1 \in S$, information component 1). The messages from $\{b_1, b_2\}$ to b_3 indicate that b_3 is two-hops-away from reference node r_1 and are considered as false information. This means that information component 3 (if $nr > 1$) is considered as “false information” for reference nodes that are one-hop-away.

The main differences with the existing distributed range-free localization algorithms is that COM-LOC processes these messages in order to increase the localization performance without increasing the communication costs. Section IV describes how this information is processed.

III. SIMULATION MODEL

This section describes the used model of the signal strength over distance distribution. We base COM-LOC on this model.

Moreover, the simulations at the end of this paper use this model (see Section VI).

Connectivity is defined by whether two nodes can communicate with each other or not. Existing literature on range-free localization models connectivity by using propagation models. Examples of used propagation models are [6] (Degree Of Irregularity) and [1] (Log-Normal Shadowing Model). In this study, we adopt the Log-Normal Shadowing Model (LNSM) for modeling the signal strength over distance distribution ([1]), because both theoretical and measurement-based studies support this model in indoor and outdoor environments ([2]). The LNSM describes the signal power decay over distance that suffers from shadowing effects. A log-normal distribution is a continuous distribution in which the logarithm of the variable follows a normal distribution. This means that:

- The average received signal strength decreases logarithmically over distance.
- The received signal power follows a normal distribution at a certain distance.

The following formula represents the LNSM:

$$P_d = P_{d_0} - 10 \cdot n \cdot \log_{10}\left(\frac{d}{d_0}\right) + X_{\sigma_{dBm}} = N(P_d, \sigma_{dBm}^2) \quad (1)$$

Here:

- P_d represents the received signal power in dBm at distance d .
- P_{d_0} represents the received signal power in dBm at reference distance d_0 . In general distance d_0 is relatively small. For simplicity, we assume that distance d_0 is 1 meter (see [2]).
- n represents the path loss exponent. The path loss exponent represents the rate at which the path loss increases with distance.
- X_{σ} represents the standard deviation of the received signal power due to shadowing effects invariant with the distance ([2]). X follows a zero-mean normal distribution with variance σ_{dBm}^2 :

$$X \sim N(0, \sigma_{dBm}^2) \quad (2)$$

- $N(P_d, \sigma_{dBm}^2)$ indicates that the received signal strength can also be represented as a normal distribution with mean P_d and variance σ_{dBm}^2 .

In this paper, we use the LNSM for estimating the connectivity probability as a function of the sender/receiver distance. We assume that the connectivity is determined by a RSS threshold (like in [13]). The following formula computes the connectivity probability as a function of the distance:

$$P(A \text{ hears } B|d) = 1 - F(RSS \leq thres|d) = 1 - \text{cdf}(thres, P_d, \sigma_{dBm}^2) \quad (3)$$

Here:

- $P(A \text{ hears } B|d)$ represents the probability that receiver A receives a message from sender B at distance d .

- $F(RSS \leq thres|d)$ represents the probability that receiver A measures a signal strength below threshold $thres$ at distance d .
- $\text{cdf}(thres, P_d, \sigma_{dBm}^2)$ represents the cumulative distribution function of the normal distribution that computes $F(RSS \leq thres|d)$.

Figure 2 shows the connectivity as a function of the distance using the LNSM with different parameters settings for $n = 3, 3.5$, $\sigma_{dBm} = V = 4, 8$ and $P_{d_0} = P = -40, -30$. Note that all parameters influence the connectivity. For simplicity, we assume that these parameters are known a priori (as in most range-free localization algorithms, for example [8], [11] and [13]). The values of these parameters can be determined by performing calibration measurements (as in [13]).

IV. ESTIMATING DISTANCES AND PROBABILITIES

This section shows how COM-LOC converts the information, obtained during the communication phase (see Section II), into distance estimates and associated probabilities using the model described in Section III.

A. One-Hop-Away Reference Nodes

The probability that reference node a communicates directly with blind node b as a function of the distance between reference node a and blind node b is defined as:

$$P(b \text{ hears } a|d_{a,b}). \quad (4)$$

This equation is equal to Equation 3 and is shown in Figure 2.

B. Two-Hops-Away Reference Nodes

The probability that blind node b did not receive a message from reference node a as a function of the distance between reference node a and blind node b is defined as:

$$1 - P(b \text{ hears } a|d_{a,b}) \quad (5)$$

C. Communication via Blind Nodes

The probability that reference node a can communicate indirectly with blind node b via nr blind nodes as a function of the distance between reference node a and blind node b is defined as:

$$P(b \text{ hears } a \text{ via } nr \text{ nodes}|d_{a,b}) \quad (6)$$

Before we provide a solution for Equation 6, we first solve Equation 6 for one blind node:

$$P(b \text{ hears } a \text{ via } 1 \text{ blind node}|d_{a,b}) \quad (7)$$

We approximate this probability by using Monte Carlo Simulations (MCS). The MCS first represents the position and distance distribution by drawing samples. We implement a grid-based sampling approach in order to ensure a uniform distribution.

- The blind node that forwards the message to node b lies within transmission range distance from node a . Therefore, we draw samples that lie within the transmission

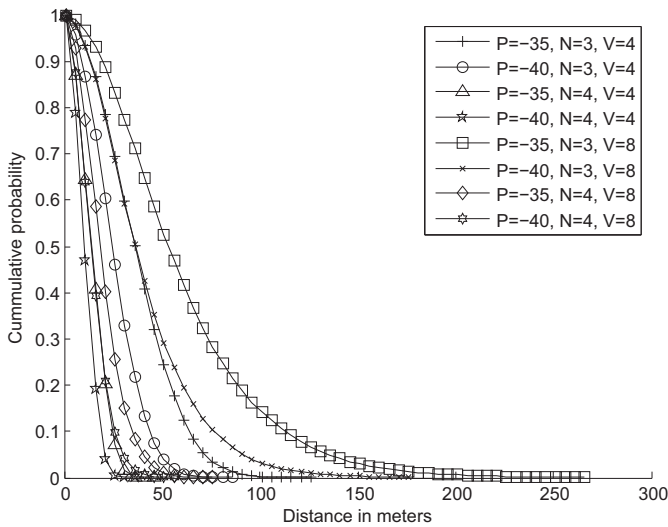


Fig. 3. Packet delivery rate via one blind node as a function of the distance using the Log-Normal Shadowing Model

range from node a . We represent this set of possible positions by: C .

- The distance between node a and node b lies in the following interval: $[0 \dots 2 \cdot tr]$. Therefore, we draw samples in a line that starts at a and has a length of $2 \cdot tr$. We represent this set of possible positions and distances by: B .

We use Equation 3 for estimating the probabilities between individual samples:

$$P(b \in B \text{ hears } a \text{ via } c \in C | d_{a,b}) = P(c \in C \text{ hears } a | d_{a,c}) \cdot P(b \in B \text{ hears } c \in C | d_{c,b}) \quad (8)$$

We use Equation 8 for estimating the probability that blind node b hears reference node a via one blind node (Equation 7):

$$P(b \in B \text{ hears } a \text{ via } C | d_{a,b}) = \prod_{c \in C} P(b \in B \text{ hears } a \text{ via } c \in C | d_{a,b}) \quad (9)$$

Figure 3 shows Equation 9 as a function of the distance between the reference node a and blind node b for parameter settings: $n = \{3, 3.5\}$, $\sigma_{dBm} = \{4, 8\}$ and $P_{d0} = \{-40, -30\}$. We use Equation 9 for estimating the probability that blind node b hears reference node a via nr blind nodes (Equation 6):

$$P(b \text{ hears } a \text{ via } nr \text{ nodes} | d_{a,b}) = \prod_{i=1}^{nr} P(b \in B \text{ hears } a \text{ via } C | d_{a,b}) \quad (10)$$

D. Final Probability over Distance Distribution

In this subsection, we are interested in the probability over distance distribution between one reference node and one blind node using the estimated probabilities in the previous sections.

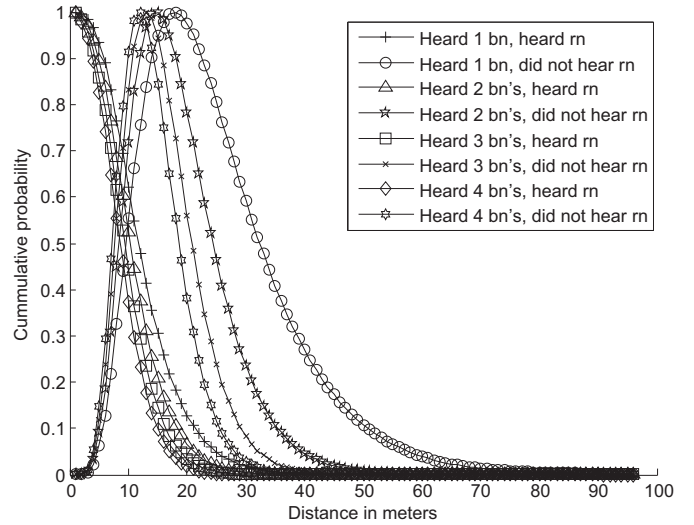


Fig. 4. Combined probabilities

We distinguish the probabilities in either one-hop-away or two-hops-away reference nodes:

- one-hop-away reference nodes ($s \in S$):

$$P(b \text{ hears } s | d_{s,b}) \cdot P(b \text{ hears } s \text{ via } nr \text{ nodes} | d_{s,b}) \quad (11)$$

- two-hops-away reference nodes ($t \in T$):

$$(1 - P(b \text{ hears } t | d_{t,b})) \cdot P(b \text{ hears } t \text{ via } nr \text{ nodes} | d_{t,b}) \quad (12)$$

Figure 4 shows the probabilities associated with Equations 11 and 12 using the following LNSM parameter settings: $n = 3.5$, $\sigma_{dBm} = 8$ and $P_{d0} = -40$. “bn” is an abbreviation for blind node and “rn” is an abbreviation for reference node.

E. Implementation and Computational costs

The computations described in the previous subsections are too expensive to run on a blind node. Therefore, we compute the outcome of the following equations before deployment:

- Equation 4 for distances in the interval $[0 \dots d_{suf}]$.
- Equation 9 for distances in the interval $[0 \dots 2 \cdot d_{suf}]$.

Here d_{suf} represents the distance that provides a sufficient packet delivery rate. The sufficient packet delivery rate is set by the user. Throughout this paper, we set the sufficient packet delivery rate to 1%. Blind nodes store the results of Equation 4 and 9 in a table with a user defined distance resolution. We represent the distance resolution by d_{res} . Throughout this paper, we implement a distance resolution of one meter ($d_{res} = 1$ meter). The stored results are later used for estimating Equation 11 and Equation 12. We interpolate probabilities associated with distance values that are not stored in the tables. The packet delivery rate at distances greater than d_{suf} are set to the probability associated with d_{suf} . The computational costs of estimating Equation 11 and Equation 12 are equal to:

$$nr \cdot \frac{d_{suf}}{d_{res}} \text{ multiplications per reference node} \quad (13)$$

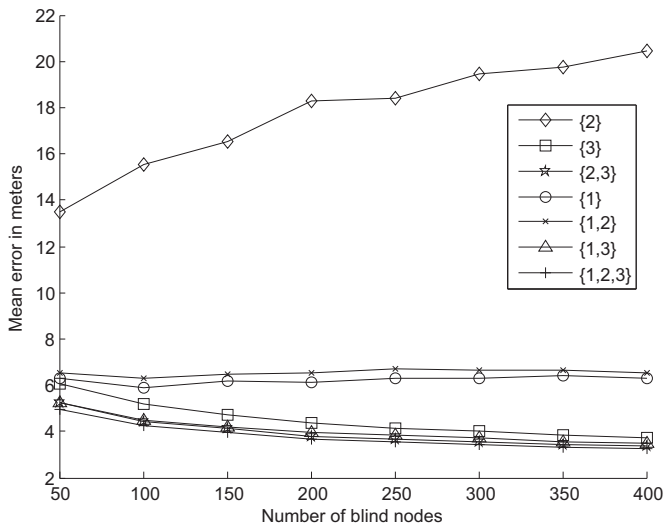


Fig. 5. Localization performance of individual and combined components

Equation 13 shows that the computational complexity is linear.

V. RANGE-FREE LOCALIZATION ALGORITHM

This section provides a description of COM-LOC for wireless networks using the results in Section IV. COM-LOC implements a grid-based Monte Carlo Localization approach. An overview of Sequential Monte Carlo methods can be found in [4]. For an example of a range-free MCL implementation, we refer to [11]. As in [11], COM-LOC consists of two phases:

- The “prediction phase” draws samples that represent the position distribution.
- The “filtering phase” weighs the samples drawn in the prediction phase according to the observations.

COM-LOC limits the x- and y-coordinates of the position distribution on the basis of the most dominant one-hop and two-hops-away reference nodes. We use this information to make a bounding box (as in [7]) and to keep the computational costs as low as possible. Throughout this paper we represent the position distribution by POS, an individual sample is represented by p ($p \in \text{POS}$). After the prediction phase we filter the samples by the MLE:

$$\begin{aligned} & \text{MAX}_{p \in \text{POS}} \\ & \left(\prod_{s \in S} P(p \text{ hears } s | d_{s,p}) \cdot \right. \\ & \left. P(p \text{ hears } s \text{ via } nr \text{ nodes} | d_{s,p}) \right) \cdot \\ & \left(\prod_{t \in T} (1 - P(p \text{ hears } t | d_{t,p})) \cdot \right. \\ & \left. P(p \text{ hears } t \text{ via } nr \text{ nodes} | d_{t,p}) \right) \quad (14) \end{aligned}$$

Here POS represents the position distribution. p represents the position estimate that maximizes the probability. Section II indicates that COM-LOC also evaluates “redundant information” and “false information” in comparison with other

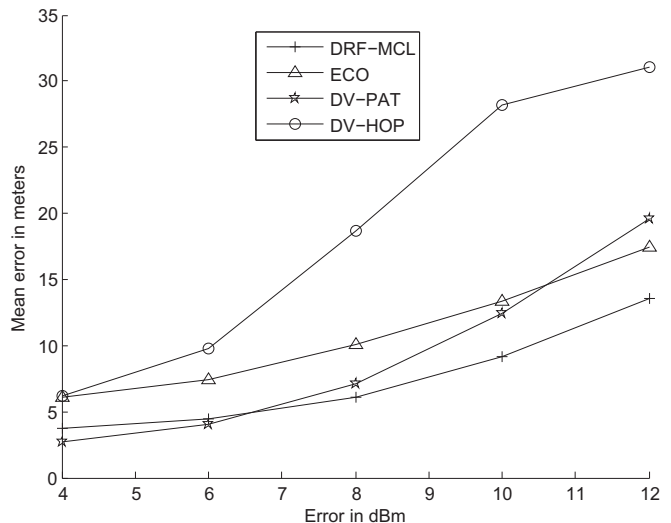


Fig. 6. Mean error of several algorithms

work in this field. This means that COM-LOC increases the computational costs by the number of computations defined in Equation 13.

VI. SIMULATIONS

This section analyzes the localization performance of COM-LOC. In addition, we compare COM-LOC with DV-HOP([5]), ecolocation ([14]) and a modified version of the MLE described in [13].

A. Set-up

Throughout this paper we use the same set-up, except when stated otherwise. The set-up parameters are:

- The surface area is 100×100 m.
- The simulations simulate the RSS by using the model described in Equation 1. In general, the following parameter values are used: $\{P_{d_0} = -40$ dBm, $n = 3.5$, $\sigma_{dBm} = 4\}$. See Figure 2 for the packet delivery rate over distance.
- 36 reference nodes are placed in a 6×6 grid over the surface area. This means that the distance between consecutive reference nodes are 20 meters.
- 200 blind nodes are randomly and uniformly placed over the surface area.
- The localization performance is given as the mean over 10 runs.

B. Influence of Information Components

This section analyzes the localization performance of the individual and combined information components (mentioned in Section II) as a function of the number of blind nodes. Section II includes three individual information components, this means that there are seven individual and combinations of information components ($\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$). Figure 5 shows the localization performance of these individual and combined information components. This figure shows that:

- The performance of component $\{3\}$ increases with the number of blind nodes. This is logical because component $\{3\}$ processes the number of heard blind nodes. More heard blind nodes means more localization information.
- Individual component $\{3\}$ provides the best localization results compared to other individual components (see $\{1\}$ and $\{2\}$). Component $\{3\}$ provides 3.4% to 70% better results than component $\{1\}$ and provides 122% to 449% better results than component $\{2\}$.
- Processing one extra information component (see $\{1,3\}$ and $\{2,3\}$) with information component $\{3\}$ increases the performance by 11% to 16%. The difference in performance decreases as the number of nodes increases.
- Processing one extra information component (see $\{1,2,3\}$) with information components $\{1,3\}$ or $\{2,3\}$ increases the performance only by 3% to 5%. The difference in performance decreases as the number of nodes increases.
- Processing information component $\{3\}$ (see $\{1,2,3\}$) increases the performance by 31% to 101% (compared to $\{1,2\}$). The difference in performance increases as the number of nodes increases.

The simulations and Figure 5 show that information component $\{3\}$ provides the most information about the blind node position and increases the performance by 31% to 101%.

C. Comparison with Other Localization Algorithms

This section compares COM-LOC with the following localization algorithms:

- DV-HOP is a range-free localization algorithm ([5]).
- A proximity based localization algorithm named “Ecolocation” ([14]), which only uses one-hop information.
- A range- and RSS-based localization algorithm described in [13]. We use the MLE described by [13] and the RSS obtained by the shortest distance (as in [7]). The MLE is estimated by a conjugate gradient method (as in [13]). We name this distributed algorithm DV-PAT.

It is known that the value of σ_{dBm} defines the performance of RSS-based localization algorithms (see [13]). Typical values of σ_{dBm} are between 6 and 12 dBm ([2]). Figure 6 shows the mean localization error as a function of these typical values of σ_{dBm} . This figure shows that COM-LOC always outperforms Ecolocation and DV-HOP. This figure also shows that COM-LOC outperforms DV-PAT with σ_{dBm} values higher than six. This implies that COM-LOC provides better results in most typical cases. Note that both Ecolocation and DV-PAT use RSS measurements, while COM-LOC only uses connectivity information like DV-HOP.

VII. CONCLUSION AND FUTURE WORK

In this paper we presented a novel localization design approach. This approach bases the MLE for localization on the used communication protocol. Theoretically, this maximum likelihood estimator should provide the best localization results given a communication protocol. We present a new distributed range-free localization algorithm using this new

approach: DRF-MCL. DRF-MCL optimizes the localization performance for a communication protocol commonly employed by localization algorithms. Simulations show that the use of this new approach increases the performance by 68% to 206% compared to other range-free algorithms using the same communication protocol. The comparative simulations of DRF-MCL with two RSS-based localization algorithms show that DRF-MCL performs better than other localization algorithms over a wide range of conditions.

In the future we would like to decrease the computational costs by implementing a smarter MCL algorithm or a iterative gradient search algorithm. We also plan to study the effect of wrongly guessing the parameters of the Log-Normal Shadowing Model on the localization performance.

REFERENCES

- [1] Hashemi H.: The indoor radio propagation channel, Proc. IEEE, July 1993, pp. 943- 996.
- [2] Rappaport T.S., Wireless Communication: Principles and Practice, Prentice Hall, ISBN 013 3755633, 1996.
- [3] F.Dellaert, D.Fox, W.Burgard, S.Thrun: Monte Carlo Localization for Mobile Robots. IEEE International Conference on Robotics and Automation (ICRA). May 1999.
- [4] A.Doucet, S.Godsill, C.Andrieu: On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. Statistics and Computing. Volum 10, pp. 197-208. 2000.
- [5] D.Niculescu, B.Nath: Ad hoc positioning systems. In: IEEE Globecom 2001, San Antonio. 2001.
- [6] T.He, C.Huang, B.M.Blum, J.Stankovic, T.Abdelzaher: Range-free localization schemes for large scale sensor networks. MobiCom, San Diego, California, September 2003, pp. 81-95.
- [7] Koen Langendoen and Niels Reijers: Distributed localization in wireless sensor networks: A quantitative comparison. In Computer Networks (Elsevier), special issue on Wireless Sensor Networks, 2003.
- [8] Y. Shang, W. Ruml, Y. Zhang and M. Fromherz: Localization From Mere Connectivity. MobiHoc’03, Annapolis, Maryland, June 2003.
- [9] R.Nagpal, H.Shrobe, J.Bachrach: Organizing a Global Coordinate System from Local Information on an Ad Hoc Sensor Network. 2nd International Workshop on Information Processing in Sensor Networks (IPSN). April 2003.
- [10] Yi Shang and Wheeler Ruml: Improved MDS-based localization. In Infocom 2004
- [11] L.Hu, D.Evans: Localization for Mobile Sensor Networks. Tenth Annual International Conference on Mobile Computing and Networking (MobiCom 2004), USA. 2004.
- [12] S.Dulman, P.Havinga: Statistically enhanced localization schemes for randomly deployed wireless sensor networks. DEST International Workshop on Signal Processing for Sensor Networks, Australia. 2004.
- [13] N.Patwari: Location estimation in sensor networks. Thesis of Neal Patwari at University of Michigan, 2005.
- [14] K.Yedavalli, B.Krishnamachari, S.Ravula, and B.Srinivasan: Ecolocation: A sequence based technique for RF-only localization in wireless sensor networks. In IEEE IPSN 2005, April 2005.
- [15] B.Dil, S.Dulman, and P.J.M.Havinga: Range-based localization in mobile sensor networks. In Third European Workshop on Wireless Sensor Networks, volume 3868 of Lecture Notes in Computer Science, pp. 164 179, Zurich, Switzerland, Feb. 2006. Springer.
- [16] Rong Peng, Mihail L. Sichitiu: Probabilistic Localization for Outdoor Wireless Sensor Networks. ACM SIGMOBILE Mobile Computing and Communications, Volume 11, Issue 1, January 2007, pp. 53-64.