

# Locating Sensors in the Forest: A Case Study in GreenOrbs

Cheng Bo\*, Danping Ren<sup>†</sup>, Shaojie Tang\*, Xiang-Yang Li<sup>\*§</sup>, Xufei Mao<sup>†</sup>,  
Qiuyuan Huang\*, Lufeng Mo<sup>‡</sup>, Zhiping Jiang<sup>‡</sup>, Yongmei Sun<sup>†</sup>, Yunhao Liu<sup>§</sup>

<sup>\*</sup>Illinois Institute of Technology <sup>†</sup>Beijing University of Posts and Telecommunications

<sup>‡</sup>Xi'an Jiaotong University <sup>§</sup>Tsinghua National Lab for Information Science and Technology, Tsinghua University

**Abstract**—As a large scale real sensor network system, GreenOrbs reveals that locating sensor nodes in the forest still faces great challenges because of volatile and fluctuating environmental factors. In this paper, we present a novel localization scheme, EARL, which provides accurate reference nodes and good ranging quality. We exam the range quality along routing paths by taking complex environmental factors into account, such as forest density, temperature and humidity. To improve localization accuracy, we use power scanning technique to judge the accuracy reference nodes and further calibrate the bad nodes through reverse-localization. To overcome the error propagation, we assign different weights to the range measurement according to the ranging quality. We implemented our localization scheme in GreenOrbs testbed, and evaluate through extensive experiments. The results demonstrate that EARL outperforms the current localization approaches with better accuracy. The localization accuracy achieved by our method is around 20% higher than best existing methods.

## I. INTRODUCTION

Localization, as a fundamental service in WSNs, has attracted more attention in recent years. Although the Global Positioning System (GPS) is widely used in outdoor localization, it still works poorly in indoor [1], underground [2] and the forest [3]. Our work is motivated by the project of GreenOrbs [4], one of the world's largest wireless sensor networks. In order to monitor the forest condition and carbon emission, thousands of sensor nodes are deployed to collect various data including temperature, humidity, illumination, and carbon dioxide. The potential applications of canopy closure [4] calculation, climate change observation, and search and rescue in the forest need the location information of sensor nodes in the wild to make the data useful and meaningful. And the experiences of GreenOrbs indicate that locating sensor nodes in the forest still faces difficulties. Especially, environmental noises, such as illustrates, temperature, humidity and canopy closure, affect the accuracy of the location results. For traditional range-based localization, RSSI is used as the main method to estimate the distance between two sensor nodes. However, the interfering factors, as well as the complex terrain and obstacles (shrubs and tree trunks) in the forest badly affect the signal propagation.

The accuracy of localization is determined by the ranging quality. Most methods focus on the accuracy of either range measurements or reference nodes, neglecting the relationship between two sensor nodes, environmental influence, and the nodes density in extent area. Some range-free approaches are

more likely to bring about large errors. Sensor nodes may locate on the midperpendicular of two anchors when it has the same hop-counts to two anchors. Nevertheless, such node could be closer to one anchor because the obstacle between them hinder the wireless communication quality.

In this paper, we propose an environmental aware localization approach, EARL, which inherits the advantages of range-free localization methods, and obtains better metric in the localization process. Meanwhile, environmental noises in the forest are comprehensively investigated to gain the good ranging quality and the influence caused by the obstacles, such as trees and shrubs. We implement our method in GreenOrbs testbed, and compare the performance with DV-Hop and CDL [3]. The result shows that our scheme outperforms existing approaches. The main contributions of this work are as follows.

(1) We propose a novel localization metric called Joint Neighbor Distance (JND), which measure the distance between nodes while taking environmental noises and factors into account. Using JND, the preliminary estimated node locations are more accurate than other localization methods.

(2) In order to obtain better ranging quality, we design a neighbor node relation verification technology through power scanning and neighboring node comparison. Such process identifies nodes with good location accuracy (called good nodes) from existing nodes with high accuracy.

(3) We present a two-phases location calibration to rectify node locations. We first calibrate the nodes on the boundary of deployment region, and verify the good nodes through mutual node relation technology. Then, we improve the location accuracy by good nodes.

(4) We implement EARL in GreenOrbs, and evaluate its performance with extensive experiments. The results demonstrate that our method outperforms other existing schemes, and the mean error is roughly 20% better than that of CDL which is one of the best scheme proposed very recently.

## II. RELATED WORKS

Many approaches have been proposed to determine the location of sensors in WSNs. Existing work mainly falls into two categories: range-based and range-free localization. Range-based approaches measure the distance and/or the angles of neighbor sensor nodes which are randomly deployed in the field through multiple skills. Although Time of Arrival

(ToA) [5], Time Difference of Arrival (TDoA) [6] [7] and Angle of Arrival (AOA) [1] [8], being the main techniques in range-based localization, could lead to accuracy results, the costs of them would be high, and such methods could not be spread to large-scale system. An alternative range-based approach is RSSI-based ranging measurement, which is popular in practical deployment. Researchers have tried to convert RSSI value to relative distance in reality. However, empirical studies have demonstrated that RSSI is sensitive to environmental noises as well as the hardware difference of sensor nodes. Although RSSI could indeed reflect the communication quality to some extent, it is still not a good choice for accuracy ranging measurement in the outdoor. Range-free approaches mainly depend on simple sensing, such as network connectivity, anchor proximity or event detection. Centroid [9], APIT [10], DV-Hop [11], MDS-MAP [12], RPA [13] which depend on connectivity measurements are proposed with low system cost in recent years. One of the advantages is that only small amount of anchors are needed to establish global coordinates. To achieve high accuracy, however, more anchors are necessary, which is proved impractical in large-scale networks, especially in the wild. Another range-free approach is beyond connectivity [14]. Based on the experimental observation, the radio signal strength weakens approximately monotonically with the physical distance, a neighbor sequence acquired through RSSI is used to capture the relative distance between 1-hop neighboring nodes. Then, the rough relationship between sensor nodes is gained through comparing the neighbor sequence. CDL [3] is proposed very recently. It is motivated by the need for accurate localization in GreenOrbs, a large-scale forest based sensor network system, is a combination of range-based and range-free scheme. It pursues better ranging quality through the localization process. It designs the virtual-hop algorithm according to their observation to address the non-uniform deployment problem. Furthermore, the local filtration and ranging-quality aware calibration are used to gain better ranging quality. However, this work does not mention the influence brought about by the natural vegetation and environmental obstacles.

### III. PRELIMINARY RESULTS AND MOTIVATION

This work is motivated by the project of GreenOrbs, one of the largest wireless sensor networks in the world, aiming at establishing a long-term monitoring system in the forest. One of GreenOrbs systems locates in a campus forest in Zhejiang Forestry University. As shown in Figure 1. It consists of more than 200 TelosB sensor nodes having MSP430 MCU and CC2420 RFIC integrated on board. The running software is based on TinyOS 2.1.1. The purpose of this system is to monitor the forest by collecting forest factors.

#### A. Preliminary Experiments

Our statistical experimental data shows that RSSI is quite susceptible to environment. Figure 2 shows the RSSI sensing results from TelosB nodes in outdoor experiments conducted in 24 hours. The transmission power is set to be 31. We

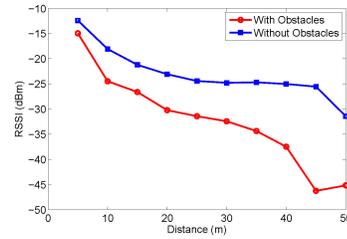


Fig. 4. RSSI value in two different environments

collect the temperature, humidity and illumination information as well. Generally, environmental factors such as temperature and humidity do affect the wireless transmission in some extent. We calculate the Correlation Coefficients according to Equation 1 to present the linear relationship between environmental factors and RSSI value. In this Equation,  $Cov(RSSI, P)$  means the Correlation Coefficients between RSSI value and the Parameter (temperature, humidity and illumination). The  $rss_i$  and  $p$  in the equation stands for the RSSI value and its relative parameter in temperature, humidity and illumination in every sample points. The correlation score is 0.0613 in temperature to RSSI value. The same results are 0.0907 and 0.1325 respectively in humidity to RSSI value, and illumination to RSSI value. In addition, we exam the RSSI value in different temperature, humidity and illumination respectively. Figure 3 illustrates the relationship between RSSI value and other three parameters separately. The blue square, cross and circle in all three figures indicate the raw RSSI value (taken in 24 hours) according to different temperature, humidity and illumination. We wire the value of every RSSI value with red line, and formulate the curve with green lines to fit the original value. Empirically, the relationship between RSSI value and three main parameters in the forest is quite hard to capture. Therefore, taking temperature, humidity, illumination and RSSI into account, it is quite difficult to estimate the distance between nodes.

$$Cov(RSSI, P) = \frac{\sum rssi \times p - \frac{\sum rssi \sum p}{N}}{\sqrt{(\sum rssi^2 - \frac{(\sum rssi)^2}{N})(\sum p^2 - \frac{(\sum p)^2}{N})}} \quad (1)$$

On the other hand, the monotonic RSSI-distance relationship of sensor node indeed present the location information to some extent. In [14], the authors indicate that any single node's RSSI sensing results for its neighboring nodes can be used as an indicator for the relative "near-far" relationship among neighbors. However, such view can only be accepted under one constraint, which is that sensor nodes deployed evenly in open environment without any obstacles. Figure 4 illustrates the RSSI result of two sensor nodes. Those two sensor nodes are deployed in different distances apart with and without any obstacles between them. As expected, we find that the obstacles in between those two nodes play important roles in affecting wireless transmission. Wireless signal could be absorbed and reflected by obstacles, and under the effect

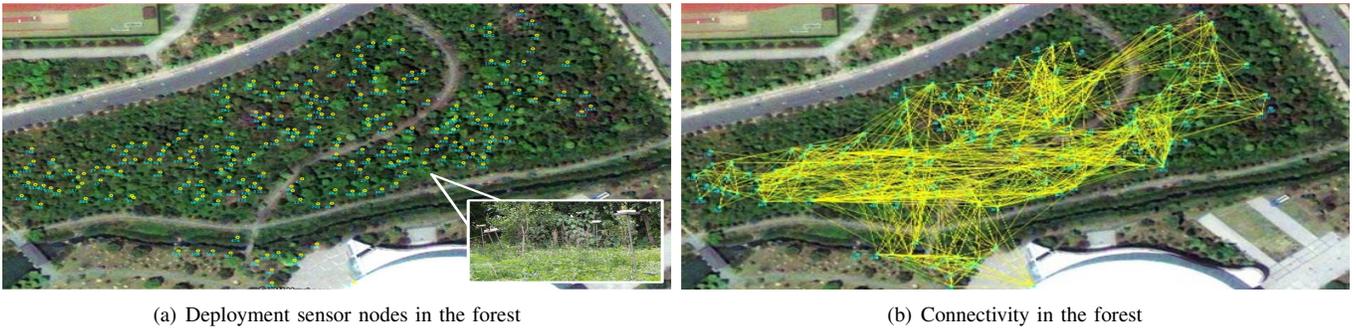


Fig. 1. GreenOrbs deployment in the campus forest

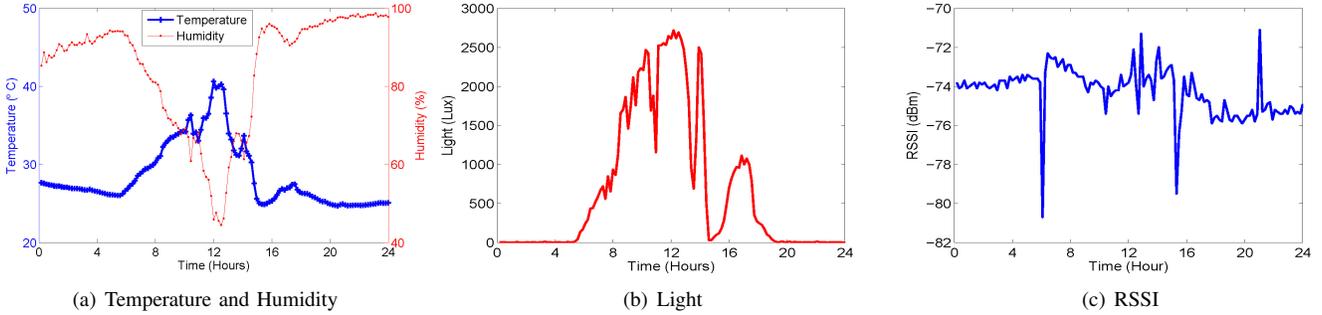


Fig. 2. RSSI in different RF output power level

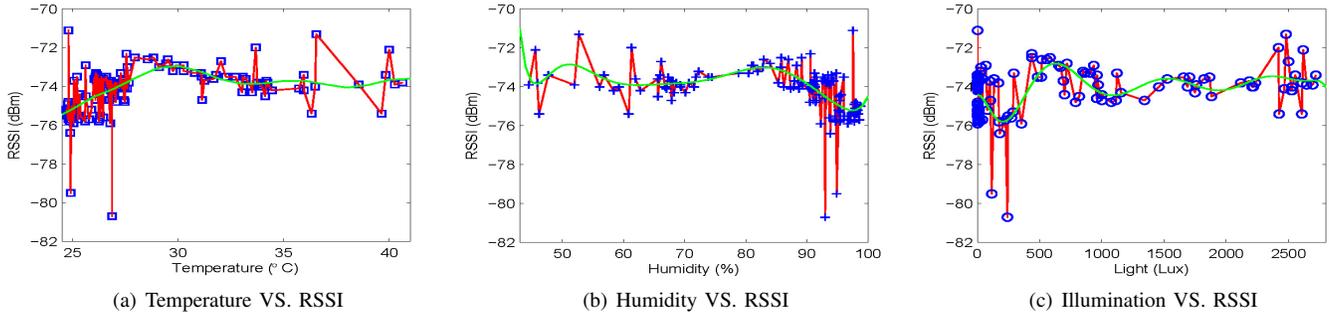


Fig. 3. RSSI in different RF output power level

of multipath effect, the signal strength will be attenuated dramatically. In the presence of obstacles, the RSSI value increase from distance 45 to 50 because of experiment error, and we demonstrate it to be normal. Open area will lead to better wireless communication quality.

### B. Experiments In The Wild

We then conducted a large scale outdoor experiment under three different environments to reveal the relationship between distance, power level and RSSI. The first experiment is conducted in campus forest, woods with thin tree density and grass with few trees. We put two TelosB sensor nodes in these three environment, the distance between them is 10 meters. Figure 5 displays the empirical data obtained from the three priori experiments. The horizontal axis presents the RF output power level, and the vertical axis is the RSSI value. We evaluate the RSSI value in different RF power level in ten

times, and the blue dots in the figures refer to the value in different power level (neglecting the repeat value). According to the blue dots, the RSSI value may vary dramatically for identical RF output power level. For example, as shown in Figure 5(a), RSSI value ranges from -40dBm to -35dBm when the power level is 4. On the other hand, a single RSSI value may correspond to a wide range of RF output power level. For instance, in the same figure, -29dBm could range from the 9th to 14th RF power level. Even worse, -21dBm covers 17th to 26th RF power level. Hence, we wire the mean value of RSSI in every RF power level in red line, and we formulate the curve (green lines) to fit the original value. Hence we utilize this kind of curve to present the changing situation of RSSI in different RF output power level.

In the second experiment, we deploy 11 sensor nodes in the same three areas to measure the changing pattern of RSSI according to different RF power level and pairwise distance.

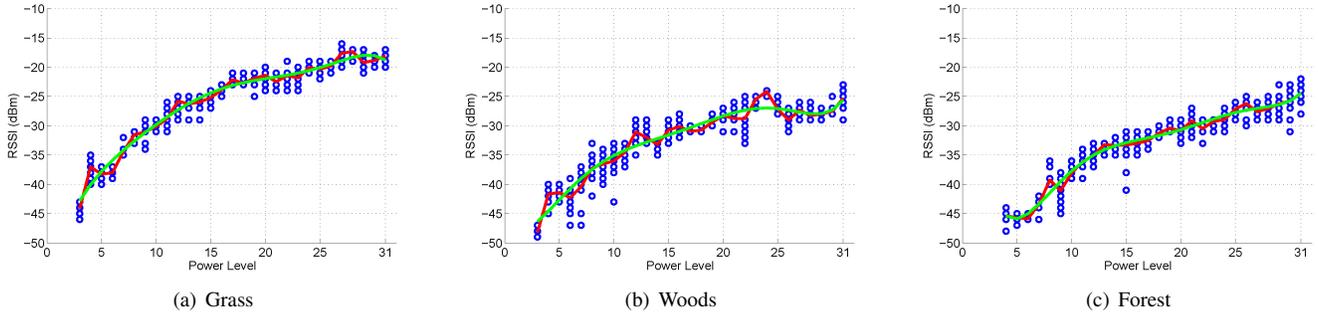


Fig. 5. RSSI in different RF output power level

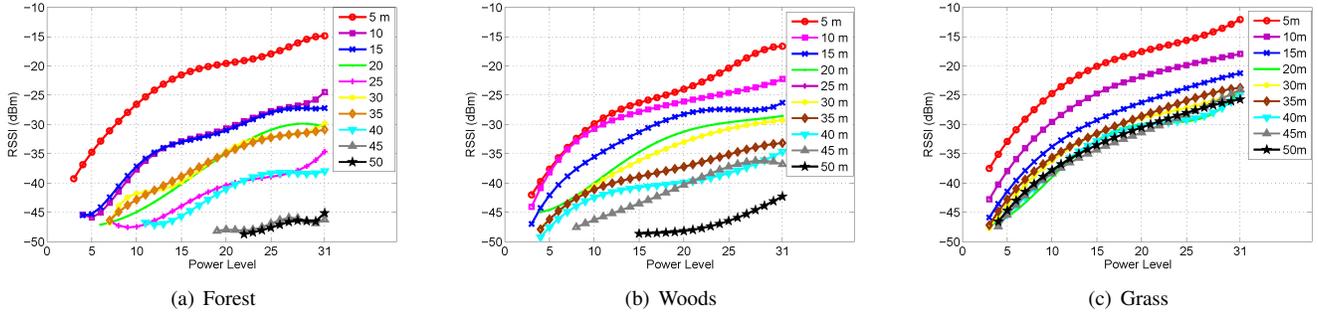


Fig. 6. RSSI value according to RF power level in three environments

The anchor node is placed in the center of an area, and the rest 10 nodes are deployed around the anchor. The distance between the sensor node the anchor is in every 5 meters, ranging from 5 meters to 50 meters. The anchor node transmits a beacon packet in every second, and the 10 nodes record the RSSI value when they receive the beacon packet. The initial RF power level is set to 1, and the power level increases 1 at a time with every beacon until 31. We fit the recorded RSSI values into corresponding curves, and plot them in Figure 6.

From this experiment, first of all, it is quite clear that when the RF power level increases, anchor node could reach more neighboring nodes. In the open area, all the neighboring sensor nodes could be reached in the third or the fourth RF power level. When it comes to the woods with thin tree density, although there are some trees in the woods, the space between many sensor nodes and the sink is quite clear, having no trees in the line. Hence the effect of trees is greatly reduced. In the figure, we could see that most of the nodes could be reached in less than the fifth RF power level. While in the forest, with the same distance between two nodes, packet should be transmitted in a much higher transmission power level than that of in the woods and open area. Assume that trees are planted randomly but evenly in the campus forest, the "near-far" relationship between neighboring nodes could be obtained according to the sequence of the sensor reached with the RF power level increasing.

Secondly, when the RF power level increases, anchor nodes could reach more neighboring nodes. In Figure 7, 10 sensor nodes are located around the anchor nodes in each environment, with the same deployment. In the forest, only one

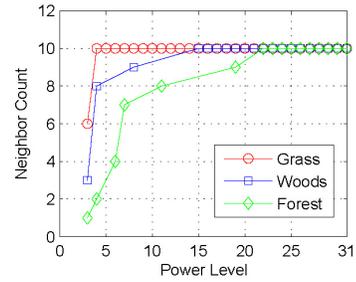


Fig. 7. The Neighboring Nodes Count when RF Power Level Increase

neighbor node could be reached in the third RF power level, because of the thick tree density. In comparison, three and six neighboring nodes are reached under the same RF power level in the woods and grass respectively. The furthest node is deployed in 50 meters away to the anchor node, when the RF power level turns to 4, all the neighboring nodes around the anchor in the grass are reached, and eight nodes are sensed in the woods. However, only two report their existence. In this experiment, the minimum RF power level to reach all the nodes is 15 in the woods and 22 in the forest. Therefore, it is undoubtedly that in wireless sensor network, where sensor nodes are deployed evenly, the more neighboring nodes being reached in maximum RF power level, the better wireless signal transmission quality, which consequently leads to the conclusion that the density in this area is much lower.

Moreover, in the forest, many curves share the similar RSSI according to the different power level, hence the curves can

be categorized into many clusters. After we check the location of these sensor nodes in the forest, however, we discover that nodes, sharing the similar RSSI curves, are located in the same area, these wireless signals are all blocked by the same trees or obstacles. Besides, if the two nodes are located with the same distance to the anchor node, whether or not there are trees or obstacles between them, both of these two nodes still share the similar RSSI changing situation curves. Consequently, the final calibration of localization system is required after the initial localization result.

#### IV. EARL SYSTEM DESIGN

This section presents the main idea of a range-free localization scheme. Traditionally, average hop distance is calculated in the network to measure the real distance between the nodes and the landmarks. In the project of GreenOrbs, sensor nodes are randomly deployed, we consider EARL in this network by three steps: neighboring node relation estimation, JND-hop localization and error aware calibration.

In neighboring node relation estimation phase, each node initially senses their neighbor nodes' absolute near-far relation. Sorting the neighbor sequence of one node through RF power scanning will compensate particularly for the error of RSSI value alone.

Subsequently, EARL locates the nodes' location using a range-free approach. In order to calculate the distance between nodes and landmarks, it counts the JND-hops (Joint Neighbor Distance hops) instead of DV-hops, taking the transmission in the forest when facing obstacles and trees into account.

EARL executes the calibration in the next process. We verify the nodes which are located on the boundary, and calibrate them in the first place. The nodes with reasonable localization accuracy (called good nodes) are selected through reverse-localization, and neighbor sequence is used as a metric to calibrate the bad nodes (with large localization error) with help of good nodes. After carefully calibration, the mean error would decline greatly.

In the next subsections, we will elaborate on the design of these three phases respectively.

##### A. Neighbor Relationship

In previous range-free localization works, a node can obtain their neighboring nodes according to sensed RSSI results. The near-far relationship of neighboring nodes are sorted according to the RSSI value in decreasing order. A curve is concluded by observing the relationship between RSSI and distance as  $PL(d) = PL(d_0) - 10 \times \eta \times \log(\frac{d}{d_0}) + X_\sigma$  where  $PL(d)$  represents the reduction in received signal strength after propagating through a distance  $d$ .  $PL(d_0)$  denotes the path loss at a short reference distance  $d_0$ , and  $\eta$  is the path loss factor or signal propagation constant. In addition,  $X_\sigma$  is a random environment noise, which follows  $X \sim N(0, \sigma_{X^2})$ .

However, the mapping of RSSI distance is in fact not reliable enough. We collect the RSSI values from 10 sensor nodes in the above experiment in one time-instance, and obtain a decreasing order of neighboring nodes sequence. We

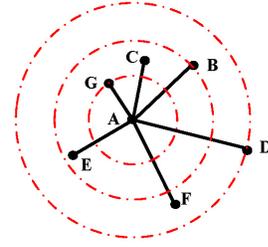


Fig. 8. Nodes Sensed by Different RF Power Level

discover that in real scenario, the RSSI value changes a lot, and cannot be used as a reliable metric to estimate the distance.

The preliminary experiment records the RSSI value in different RF power level. In the forest, trees are somehow evenly distributed. With the RF power level increases, more neighboring nodes can be reached. Therefore, a neighborhood ordering of a node can be obtained with the sequence of appearance of neighboring nodes according to the increasing RF power level. A simple example is shown in Figure 8. In this figure, when the RF power level is 3, only node  $G$  is reached by a packet broadcast by the anchor node  $A$ . When the power level increases to 4, node  $C$ ,  $E$  and  $B$  is reached. Similarly, node  $F$  and  $D$  are reached when the power level turn to 5. Therefore, a near-far relationship information of neighboring sensor nodes around a certain node can be obtained as  $\{G, C, E, B, F, D\}$ .

##### B. Testbed In The Woods

In order to confirm the observation from previous section, we construct a 2-dimensional random-deployed network testbed in the woods to evaluate the neighboring nodes information.

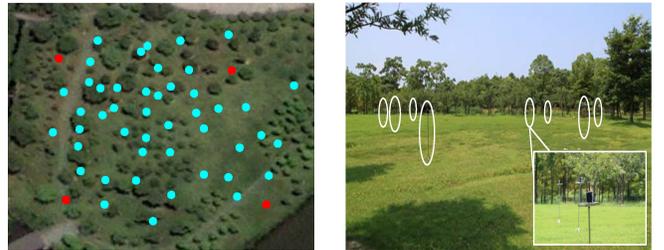


Fig. 9. Testbed in the woods

1) *Experiment Setup*: Figure 9 shows the experiments scenario of a large woods with low tree density in a campus of Zhejiang Agriculture and Forestry University. We deployed 50 TelosB sensor nodes randomly and the network covers an area of roughly  $100 \times 100$  square meters. Each sensor node, with antennas pointing to the sky, is placed 1.3 meters above the earth by a stainless steel frame.

In the woods scenario, the difference in the density of trees will lead to the attenuation of wireless signal strength, and obstacles in the wild heavily affect the wireless transmission radius. Therefore, under the same RF power level, different numbers of neighboring nodes could be reached. Suppose

sensors  $A$ ,  $B$  and  $C$  are the anchor nodes in the testbed, all the packets in other nodes could be transmitted to the anchor nodes through multi-hop transmission, and the hop distance in each hop is unequal.

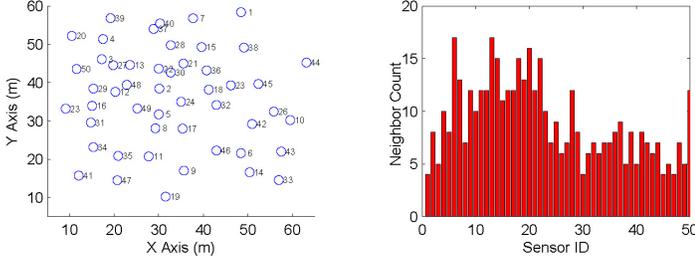


Fig. 10. Neighbor Count of nodes in the testbed

2) *Performance*: Figure 10 plots the neighborhood size of each node.  $X$ -axis lists ID of each node and  $Y$ -axis indicates the count of neighboring nodes in one-hop distance at that position. This figure verifies that nodes located in open area have more 1-hop neighbors than those which are blocked by the trees, and the boundary nodes have smaller neighbor nodes size.

### C. Mutual Neighbor Information

In the previous test, we find that sensor node could reach more neighboring nodes if they are deployed in a an open environment, such as low tree densities and less obstacles in the area. As a result, the transmission radius and the one-hop distance are both increased. In GreenOrbs, the sensor nodes are deployed randomly at uniform. The trees therefore play an important role in affecting wireless transmission among sensor nodes. In the subarea with high tree densities, the transmission radius is much smaller, therefore, the neighbor count will be relatively low. We find that the neighbor count of nodes, in some extent, reflects the transmit radius. Many interesting approaches have been proposed for localization with mere connectivity information. Considering the limited communication range of each sensor node, it only knows which nodes are nearby under its local communication range. Unfortunately, according to the irregular RSSI value alone, it does not know how far and which direction its neighbors are. Here, we use the proximity distances and connectivity information to estimate the location of sensor nodes in the forest. First, a few nodes with known position deployed in the network, are used as landmarks. Then, in order to estimate the distance from the nodes to the landmarks, we propose **Joint Neighbor Distance (JND)** to estimate the distance of each pair of nodes. JND calculates the neighboring nodes count in one-hop transmission radius along this path, by taking environmental factors, such as tree density, obstacles, canopy closure, into account. We use JND, which is defined in Equation 2, to quantify the impact from surrounding environment on the distance measures.

$$JND(X_i, X_j) = NC(X_i, X_j) \cup NC(X_j, X_i) \quad (2)$$

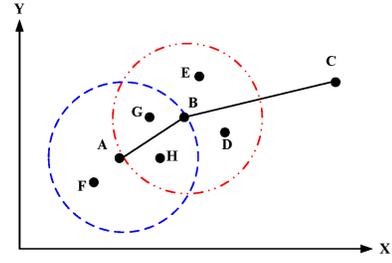


Fig. 11. Example of Joint Neighbor Count

Where  $NC(X_i, X_j)$  is the Neighbor Count (the number of neighbors) of  $X_j$  with respect to  $X_i$ , when the transmission radius of sensor  $X_i$  just touches sensor  $X_j$ . The JND between any pair of 1-hop neighbors is the sum of the mutual neighbor count of these two nodes.

Take an example in Figure 11. With the power level growing, the sequence of neighboring nodes being sensed can be obtained. Ideally, if a node can reach a farther node with larger RF power level, it can also cover the nearer ones. The transmission radius could reach sensor B, it could cover sensor F, G and H as well. Similarly, sensor B could cover sensor D, E, G and H when the transmission radius is  $D_{AB}$ , which is the Euclidean distance between A and B. Therefore,  $NC(A,B)=4$ , and covers  $\{B, F, G, H\}$ , while  $NC(B,A)=5$ , covering  $\{A, H, G, E, D\}$ . Therefore,

$$JND(A, B) = NC(A, B) \cup NC(B, A) = 4 + 5 - 2 = 7 \quad (3)$$

The reason of subtracting 2 is because sensor G and H both appear in  $NC(A, B)$  and  $NC(B, A)$ . The environmental-aware localization initially estimates nodes locations using a range-free method. Instead of DV-hops, it utilizes the JND to calculate the estimated distances along the path between a pair of sensor nodes. Similar to DV-hop, the relative distance turns to the smallest accumulated JND instead of shortest path-hops. With the help of the accurate location information from beacon nodes, the expected physical distance for 1 unit JDN is given by Equation 4:

$$\begin{aligned} JND_{unit} &= \frac{\sum_{i \neq j} Distance(R_k, R_j)}{\sum_{i \neq j} JND(R_k, R_j)} \\ &= \frac{\sum \sqrt{(X_k - X_j)^2 + (Y_k - Y_j)^2}}{\sum_{i \neq j} JND(R_k, R_j)} \quad (4) \end{aligned}$$

Where  $Distance(R_k, R_j)$  is the Euclidean distance between landmarks  $R_k$  and  $R_j$ . Then, each node computes its distance to the landmarks as:  $D_{i,k} = JND_{unit} \bullet JND(v_i, R_k)$  where  $D_{i,k}$  is the distance from node  $v_i$  to the landmark  $R_k$ . After estimating the distances to the beacon nodes, similar to DV-hop, each node computes its coordinates based on trilateration using Least Square Estimation.

### D. Calibration

After calculating the initial estimated locations of sensor nodes, calibrations should be processed. The calibration process

consists of three steps: boundary nodes detection, good nodes verification and bad nodes calibration.

In the localization process, the neighbor nodes information is utilized as a metric to measure the distance between landmarks and nodes. Empirically, boundary nodes have small neighbor count, which will lead to the great error of locations of these nodes. In order to verify the boundary nodes, we select one node as a root from the located nodes to establish a tree, and record all the leaf node of this tree. These leaf nodes could be regarded as possible boundary nodes. Every located nodes should be chosen as root node to form a tree, and we could compute a weight for each node:  $P_i = \frac{N_i}{N_{node} - N_{landmark}}$ .  $P_i$  represents the possibility for that node to be a boundary node.  $N_i$  is the number of sensor  $i$  being estimated to be a boundary node.  $N_{node}$  and  $N_{landmark}$  represent the amount sensors in the filed and the landmarks respectively.

Those nodes which are located in the center have low weight, thus we select a threshold to filter these absolute none-boundary nodes. Then, we compute the Virtual Neighbor Count (VNC) by  $VNC(i, j) = NC(j, i) \times P_i$  where  $VNC(i, j)$  stands for the virtual neighbor count of sensor  $i$  according to sensor  $j$  and  $j$  is the nearest neighbor of sensor  $i$ . The  $NC(j, i)$  denotes as sensed neighbor count of sensor  $j$  according to sensor  $i$ . The maximum of  $VNC(i, j)$  and  $NC(i, j)$  will be selected as the Calibrated Neighbor Count (CNC). CNC will replace the original NC to re-locate these possible boundary node again through CNC. In this case, the error of boundary nodes will decrease.

In addition, according to the JND scheme, we could obtain the set of nodes which are located on the boundary of the area by the coordinate of each nodes. Empirically, the JND of these nodes will be different from the ground truth, experiencing great errors, because of the small neighbor count. We firstly calibrate the nodes located on the boundary of the area. We check the neighbor sequence of every boundary node, and choose the neighbor count of nearest neighbor nodes to be the neighbor count of related boundary node.

Generally, in open area, RSSI value between two nodes decreases monotonically as the distance between them increases. But in the wild scenarios, such as forest, RSSI value is heavily affected by the trees and other surrounding obstacles, which is not reliable when obtaining neighbor sequence. In the process of the neighbor scanning, the near-far relationship of neighboring nodes can be obtained according to the appearance order when the RF power level increase. However, every RF power level increases, the transmission radius increases none-linearly, and more than one neighboring nodes may be added into the neighbor sequence. In order to get more correct neighboring node sequence, we propose a two step process. First of all, when the RF power increases one level, we put the nodes which appears in this level into one group. As shown in Figure 8, when the power level is 4, sensor  $G$  appears, and put into first group. When the power level climb to 6, sensor  $B$ ,  $C$  and  $E$ , while  $F$  and  $D$  are sensed when the power level grow to 8. Hence the neighboring nodes are divided into three groups,  $((A), (G), (B, C, E), (D, F))$ . In the certain subarea, where the

neighboring nodes are being sensed in some specific RF power level, the near-far relationship according to certain anchor can be obtained through RSSI. Secondly, the RSSI value of sensor nodes in every group are about to be measured. Therefore, we denote the neighboring nodes sequence around sensor  $i$  to be  $S_i$ , e.g.,  $S_A$  can be sorted as  $(A, G, C, E, B, F, D)$ .

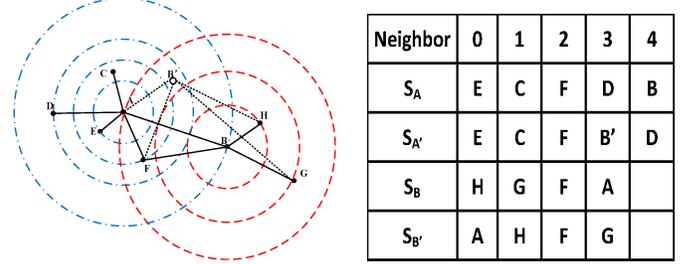


Fig. 12. Neighbor sequence in calibration

As shown in Figure 12, sensor  $B$  is the next hop of sensor  $A$ , the neighboring nodes sequence of sensor  $A$  is presented as  $S_A = (E, C, F, D, B)$ . Similarly,  $S_B = (H, G, F, A)$ . According to the JND localization scheme, node  $A$  initially sorts its neighbors in the ascending order with respect to the estimation distance to them, generating the second neighbor nodes sequence, denoted as  $S_{A'}$ . Theoretically, the sequence  $S_A$  and  $S_{A'}$  should be identical if the estimation is correct. However, if there is a significant mismatch between them, the node's estimated location would be wrong. We also take Figure 12 as an example. Sensor  $B$  is the ground truth node in real deployment, while the  $B'$  is the estimation location of sensor  $B$ . Then  $S_{A'}$  is  $(E, C, F, B', D)$ . Comparing with  $S_A$ , where there is light mismatch between  $S_A$  and  $S_{A'}$ , the only obvious mismatch occurs in sensor  $D$  and sensor  $B$ . The difference between  $S_B$  and  $S_{B'}$  is metric of localization error of sensor  $B$ . In this case, we introduce the ratio of longest common subsequence of neighbor nodes, denoting as  $\eta_A$  to filter good nodes from bad nodes. Empirically, the ratio of longest common subsequence of a good node is higher than that of a bad node, hence we set a threshold to filter the good nodes from bad nodes.

Unfortunately, some pre-selected good nodes still suffer from relatively high localization errors. Therefore, we have to calibrate these good nodes. In this phase, we propose a novel calibration scheme, called reverse-localization. We randomly select four good nodes to be reverse-landmarks, and pretend the original four landmarks to be unknown. We recompute the localization scheme to locate the four original landmarks, and record the error. If the results are within an acceptable boundaries, these four nodes can be proved to be good nodes. After iterating reverse the four original landmarks and recording the error, we select one set of four nodes to be the absolute good nodes if the error of these four nodes are minimum. To some extent, we have eight landmarks then, and we calibrate other relative good nodes by these eight landmarks, which will decrease the mean error of localization dramatically. Figure 13 shows the results after selecting four

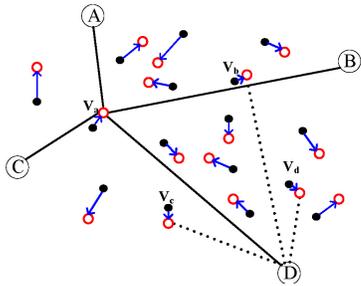


Fig. 13. Verify good nodes

good nodes:  $V_a$ ,  $V_b$ ,  $V_c$  and  $V_d$ . The black dots are the ground truth of each node's position, and the red circles are the calculated locations of each node.

## V. PERFORMANCE EVALUATION

We implement EARL in GreenOrbs to locate all the sensors in the forest. The performance of EARL is evaluated through testbed evaluation and large scale experiments. We also implement DV-hop, and CDL [3] scheme for performance comparison.

### A. Experiments on Testbed

In the first experiment, we setup our testbed in the campus woods near the laboratory, the deployment is illustrated in Figure 9. 50 sensor nodes are randomly deployed, among which some nodes are located behind trees, while some are in the open area without any obstacles. We set four sensor nodes, which are marked by red circles, located in the corner of area as landmarks. The result of EARL is plotted in Figure 14(a), the black dot is the ground truth of the nodes, and the blue squares are the estimated location of each nodes. It is quite obvious that those located in the open area have lower error, and others located behind trees have relatively higher error. Meanwhile, we also localize these 50 nodes with DV-hop and CDL, and the performance comparison are shown in Figure 14(b). The mean error resulted from our method is no more than 5 meters, CDL takes 9, and DV-hop's is larger. Figure 14(c) shows the CDF of DV-hop, CDL and EARL. From experimental results, we could see that the performance of DV-Hop is the worst.

### B. Large-scale experiment

Then, we take a large-scale experiment, with more than 200 sensor nodes are deployed randomly in the forest near the laboratory. We record the neighbor count of each nodes, and calculate JND of each pair of nodes. The natural vegetation in the forest is quite unexpected and irregular. Some places are covered by shrubs and some by trees, therefore the wireless transmission radius is effected greatly, as shown in Figure 1.

Corresponding to the deployment, GreenOrbs consists of near 230 sensor nodes in a rectangular coordinate. Sensor nodes are placed roughly 1 meter high above the earth by a stainless steel frame. Some of the nodes are covered by shrubs, some are located behind big trees, and some are above the open ground. Environmental data is transmitted to sink

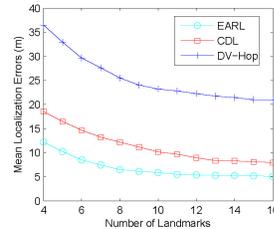


Fig. 16. Impact of the number of landmarks

through multi-hop. Four nodes positioned near the boundary of the deployment are chosen as landmarks. We have tested the localization results of three approaches and the results are shown in Figure 15. In Figure 15(a), the black dot is the ground truth of every nodes the forest, the red circles are the estimated location of every sensor nodes, and the blue square is the location computed by EARL. We connect every ground truth of nodes to its calculated location through CDL and EARL respectively, the length of the line represents the error of localization. It is clear that most of the nodes have less error by EARL than by CDL. We choose 100 sensor nodes randomly from the deployment, and compare the error between CDL and EARL as shown in Figure 15(b). The blue cross stands for the error of EARL while the red square means the CDL. Generally, most of blue crosses are below red squares, which demonstrates that the performance of EARL is better than CDL.

Figure 15(c) shows the cumulative distribution of location errors of three localization schemes. It is easy to see that JND achieves high localization accuracy than DV-hop and CDL. Owing to the complex environment in the forest, the hop distance is quite unexpected. Hence the DV-Hop performs the worst, it does not taking the difference of transmission radius facing obstacles or trees, and the mean error is around 40 meters. The CDL works much better, but when facing larger scale networks, the mean error is roughly 12 meters. The JND undergoes approximately 9 meters error in the forest. The results of three localization methods are plotted in Figure 15(c). We take half of nodes in this figure.

Due to the complicated deployment environment and localization schemes in the forest, the number of landmarks has great impact on the localization accuracy, especially in the wild. In our experiment, we tune the number of landmarks from 4 to 16, and compare the results with CDL and DV-Hop. Figure 16 plots that more landmarks help improve the localization accuracy for all approaches. It is worth mentioning that localization error of EARL reduced constantly comparing CDL and DV-hop. Through observation, when utilizing JND, over 90% of sensor nodes experience relatively lower localization errors than DV-hop and CDL.

## VI. CONCLUSION

In this work, we present environmental aware localization scheme EARL, which takes the joint neighbor count to measure the distance between two nodes. This paper presents

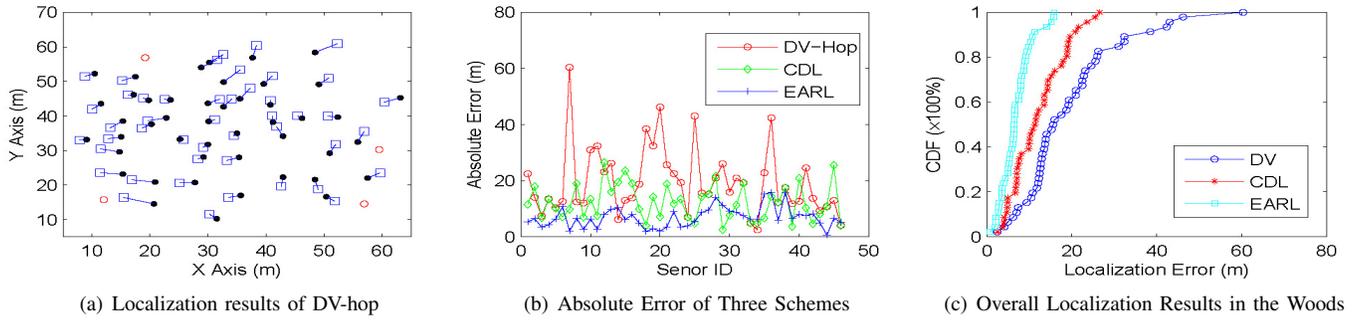


Fig. 14. Localization in the Woods

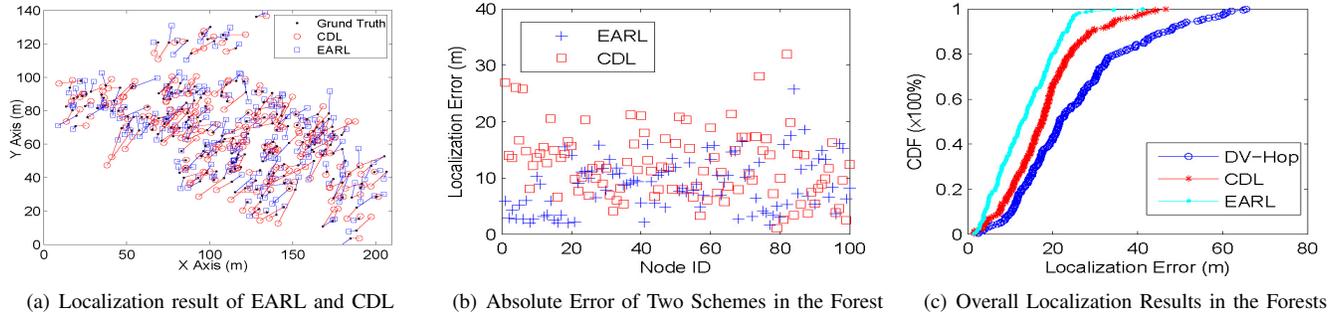


Fig. 15. Localization in GreenOrbs

our experience from establishing large-scale wireless sensor networks. We find that the network is highly affected by the complex environment factors, which will further affect the localization accuracy. Therefore, we propose Joint Neighbor Distance as a metric to depict the distance between node pairs. Our extensive experimental results demonstrate that EARL outperforms existing approaches in terms high accuracy and efficiency in the forest.

#### ACKNOWLEDGMENT

The research of authors is partially supported by NSF CNS-0832120, NSF CNS-1035894, National Basic Research Program of China (973 Program) under grant No. 2011CB302702 and No. 2011CB302705, National Natural Science Foundation of China under Grant No. 61170216, program for Zhejiang Provincial Key Innovative Research Team, program for Zhejiang Provincial Overseas High-Level Talents (One-hundred Talents Program), and Funded by Tsinghua National Laboratory for Information Science and Technology (TNList).

#### REFERENCES

- [1] H. Chang et al. Spinning beacons for precise indoor localization. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 127–140. ACM, 2008.
- [2] M. Li and Y. Liu. Underground structure monitoring with wireless sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks*, pages 69–78. ACM, 2007.
- [3] W. Xi, Y. He, Y. Liu, J. Zhao, L. Mo, Z. Yang, J. Wang, and X. Li. Locating sensors in the wild: pursuit of ranging quality. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 295–308. ACM, 2010.
- [4] L. Mo, Y. He, Y. Liu, J. Zhao, S.J. Tang, X.Y. Li, and G. Dai. Canopy closure estimates with GreenOrbs: sustainable sensing in the forest. In *SenSys*, pages 99–112. ACM, 2009.
- [5] S. Lanzisera, D.T. Lin, and K.S.J. Pister. Rf time of flight ranging for wireless sensor network localization. In *Intelligent Solutions in Embedded Systems, 2006 International Workshop on*, pages 1–12. IEEE, 2006.
- [6] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 1–14. ACM, 2007.
- [7] A. Savvides, C.C. Han, and M.B. Strivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 166–179. ACM, 2001.
- [8] D. Niculescu and B. Nath. Ad hoc positioning system (aps) using aoa. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pages 1734–1743. Ieee, 2003.
- [9] N. Bulusu, J. Heidemann, and D. Estrin. Gps-less low-cost outdoor localization for very small devices. *Personal Communications, IEEE*, 7(5):28–34, 2000.
- [10] T. He, C. Huang, B.M. Blum, J.A. Stankovic, T. Abdelzaher, and VIRGINIA UNIV CHARLOTTESVILLE DEPT OF COMPUTER SCIENCE. *Range-free localization schemes for large scale sensor networks*. Citeseer, 2003.
- [11] D. Niculescu and B. Nath. Dv based positioning in ad hoc networks. *Telecommunication Systems*, 22(1):267–280, 2003.
- [12] Y. Shang, W. Ruml, Y. Zhang, and M.P.J. Fromherz. Localization from mere connectivity. In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, pages 201–212. ACM, 2003.
- [13] C. Savarese, J. Rabaey, and K. Langendoen. Robust positioning algorithms for distributed ad-hoc wireless sensor networks. In *USENIX technical annual conference*, volume 2, 2002.
- [14] Z. Zhong and T. He. Achieving range-free localization beyond connectivity. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 281–294. ACM, 2009.