DENSITY-AWARE HOP-COUNT LOCALIZATION (DHL) ALGORITHM IN UNEVENLY DISTRIBUTED WIRELESS SENSOR NETWORKS

WONG SAU YEE

(B.Eng (Hons), MMU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTR OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2005

Acknowledgments

I would like to acknowledge the contributions of a number of people. I deeply appreciate their valuable contributions. My research would be stuck in the wilderness without the guidance of my supervisor, Dr. Winston Seah, who provided the initial stimulating ideas and subsequent perspectives on the research direction. I am particularly grateful to Mr. S.V. Rao and Mr. Lim Joo Ghee, who volunteered many hours of their valuable time to give advices to refine the algorithm. The TARANTULAS ("The All-terrain Advanced Network of Ubiquitous Mobile Asynchronous Systems") project members also contributed substantially to the development of this work. Their comments and discussions had a direct impact on the quality of this thesis. I would like to thank Institute for Infocomm Research (I^2R) for the scholarship given under TARANTULAS research grant. TARANTULAS project is funded under Embedded and Hybrid Systems (EHS) Thematic Strategic Research program by the Agency for Science, Technology and Research (A*STAR). I am very grateful to Mr. Jean-Luc Lebrun for giving a lot of useful tips and comments in writing a good thesis. I have the pleasure to interact with several students who are also doing their graduation works and have been beneficial from their inputs for the final presentation of this thesis. I feel a deep sense of gratitude for my family and friends who provide persistent inspiration and support to me all along.

Table of Contents

Summary

Wireless sensor networks are data-centric networks that have direct interaction with physical environment. In these networks, micro-sensors collaborate to feed the network administrator with desired information related to the monitored physical environment. In order to extract meaningful information from the network, some sensing data need to be stamped along with position information. However, localization is not an easy task due to challenges in the sensor networks such as cost, sensor size, resource shortage, and energy limitation.

Hop-count based localization algorithms offer a feasible solution despite these network constraints. Positioning based on hop-count is simple and distributed. In multi-hop sensor networks, the distance progressed by a broadcast is almost equivalent to the transmission range of the transmitting node. Thus, counting the minimum number of packet broadcast, i.e., hop-counts, between two nodes can be used to approximate the distance between the two communicating nodes. Besides, sensors usually have low mobility. During the period between hop-counts are disseminated and hop-counts are obtained by each node, the node positions do not change considerably. Thus, the linear relationship between hop-count and distance is consistent over time. Therefore, hop-count technique is suitable for localization in multi-hop and low-mobility wireless sensor networks. However, there are issues to be solved before they can be applied extensively in different sensor network scenarios.

We identify two potential issues with conventional hop-count localization algorithms. Firstly, localization accuracy is not guaranteed for non-uniform and sparse

networks. Localization are usually designed based on the assumption that the network distribution is uniform and dense. In such scenario, the distance progressed by one hop (i.e., hop-distance) can be associated with a constant range. However, in non-uniform networks, if constant hop-distance is used, the accuracy of distance estimation tends to degrade. This is because the actual hop-distance tends to be variable from one hop to another hop. We call this first issue as *density issue*.

Secondly, error in distance estimation tends to accumulate with the increase of hop-counts. By advancing one hop, the actual progressed distance is either less than or equal to transmission range. This disparity is accumulated with the increase of hop-count. Besides, with the increase of propagation path length, the probability of achieving a straight and direct end-to-end propagation path decreases. A winding path tends to accumulate more hop-counts. Thus, a node that is positioned far from a reference point tends to accumulate more errors. This issue is called *path length issue* in this thesis.

Realizing that these two issues have not received much research attention, a novel Density-aware Hop-count Localization (DHL) algorithm is proposed. In our algorithm, the distance advanced by each hop is not necessarily linearly proportional to one hop-count. Instead, a range ratio parameter, which is based on the surrounding density of a transmitting node, is used to estimate the hop-distance from the node. This effectively reduces distance overestimation. In addition, a 'Confidence Level' is associated with each estimated distance. If more hop-counts is accumulated in hop-count propagation, the corresponding estimated distance is associated with a lower confidence rating. Then, a node can select the estimated distances with high confidence levels to compute its position by method like triangulation [31].

List of Tables

List of Figures

List of Symbols

D_{avg}	Average Hop-distance
HC	Hop-count
N_{ngbr}	Number of neighboring nodes
R	Transmission range
N	Total number of nodes
K	Total number of reference nodes

Chapter 1 Introduction

1.1 Localization Challenges in Wireless Sensor Networks

In ad hoc wireless sensor networks [1][9][12][13], hundreds or thousands of tiny sensors are scattered randomly over an area to perform coordinated surveillance or to monitor environmental phenomenon [7], such as temperature, humidity, pressure and many others. In many cases, in order to extract meaningful information, gathering sensing data alone is not sufficient; this collected data needs to be complemented with position information. For example, position information is essential in acquiring the origins of events, to assist querying of sensors, to discover network coverage and to track target movements. However, the inherent characteristics of wireless sensor networks make acquiring this position information a challenging issue.

Position estimation in wireless sensor networks is not an easy task due to network constraints like lack of infrastructure, cost, form factor, limited computation and communication capabilities, and finite energy supply. In designing a localization algorithm, some influencing factors need to be taken into consideration. A localization algorithm should be (a) distributed (i.e., does not rely on some powerful nodes to do centralized computation), (b) self-organizing (i.e., does not rely on preinstalled infrastructure or set up), (c) robust (i.e., tolerant to network dynamisms like node failure), (d) energy-efficient (i.e., does not incur large computation and communication overheads), and (e) scalable (i.e., practical for large number of nodes). Given these design objectives, hop-count based localization fits into the picture since it meets these requirements. Thus, hop-count based localization can offer a feasible solution to wireless sensor networks. However, there are issues that need to be resolved before hop-count based localization can be applied widely in the ad hoc sensor networks.

For conventional hop-count based localization, the major concern comes from the need for a *dense* and *uniformly distributed* network (i.e., each node has high and similar number of neighbors). If this network requirement is fulfilled, distance propagated by one hop is consistent and approximately equals to transmission range. Thus, hop-counts can be used to gauge the distance between two nodes. However, in a sparse or non-uniformly distributed network, the distance progressed by each propagation is not consistent. Thus, the relationship of hop-count being linearly mapped to progressed distance is not always true in sparse or non-uniform networks. In this thesis, the problem in localization caused by non-uniform node distribution is referred to as *density issue.* To address this issue, it calls for the consideration for density awareness in hop-count based localization.

Second issue of concern is error accumulation over long propagation path (henceforth referred to as *path length issue*). Error accumulates when hop-count is incremented over multiple hops. This error arises because each hop-distance is considered as equivalent to one transmission range, but commonly, the actual hop-distance is less than that, i.e., the distance advanced by propagating one hop is not exactly equivalent to one transmission range. Over a long propagation path, the disparity accumulates and the cumulative error becomes increasingly significant. Besides, the probability of finding a straight and direct propagation path over a long path diminishes. A winding path tends to accumulate more unnecessary hop-counts than a direct path. Consequently, a sensor node that is positioned far from a reference point tends to pile up more errors.

In addressing these two primary issues, our algorithm has two phases. The first phase, *Density-aware Phase*, deals with *density issue* where the algorithm strives to integrate density-awareness while propagating hop-counts throughout the network. The hop-count increment incorporates the parameter of local density, i.e., a sensor node's connectivity per unit transmission coverage. The second phase, *Path length-aware Phase*, deals with *path length issue* where each estimated distance is associated with a confidence rating. When a node computes its position using methodologies like triangulation, it selects those estimated distances with high confidence, i.e., distances that are computed from less hop-counts.

The driving design factor of the algorithm is to address the two above mentioned issues and to deliver reliable estimated positions to sensor nodes in sparse and non-uniform networks.

1.2 Conventions Used in Thesis

To ease explanation, some variables are represented in specific terminologies or annotations in this thesis. This section explains and clarifies the meaning of these representations and symbols.

Localization may be defined as the process of determining an object's position relative to a particular coordinate system. It can also be regarded as the process of discovering spatial relationship among objects. Localization has also been referred to as locationing, positioning, location estimation, position estimation, location discovery and position discovery in the literature.

In wireless sensor networks, localization can leverage on a few specific nodes with a priori known positions, henceforth known as *reference nodes*, to jump-start the position discovery process. These nodes are readily equipped with location information at the beginning of network deployment. The location information can be pre-programmed or pre-coded into the memory of these nodes. Alternatively, special hardware can be attached to the nodes. Another method is to place the reference nodes deliberately at specific positions. Reference nodes are also known as beacons [35][37], GPS nodes, seed nodes [26], landmarks [27][29], or anchors [34] in the literature.

The rest of the nodes that do not have a priori knowledge of their locations are simply known as “sensor nodes”, “sensors” or “nodes”. The sensor nodes can compute their positions with respect to the reference nodes in a certain global coordinate system or an independent relative coordinate system.

To characterize a network, the following annotations are used in algorithm description.

- Hop-counts, HC
- Number of neighbors of a node, $N_{ngb\mathrm{r}}$
- Radio transmission range of a node, R
- Total number of nodes in the network, N
- Total number of reference nodes in the network, K

1.3 Objectives and Contributions

Wireless sensor networks are data-centric networks. In these networks, sensors collaboratively feed the network administrator with desired information related to the monitored environment. In order to extract meaningful information from the network, some sensing data need to be stamped along with position information. However, traditional localization algorithms do not provide straightforward solutions due to constraints such as cost, sensor size, resource shortage, and energy limitation. Hop-count based localization algorithm offers a feasible solution despite these network constraints; however, there are issues to be solved before hop-count localization can be applied extensively in different network scenarios.

The principal objective of this work is to develop a hop-count based localization algorithm that is capable of providing position estimations to nodes in ad hoc wireless sensor networks even though the node distribution is non-uniform or the node density is low [39]. Also, we seek to reduce errors in position estimation introduced by long propagation path [40]. Besides achieving these main goals, we seek to develop a localization algorithm that fulfills the criteria of being simple, distributed, robust and energy-efficient.

The main contribution of our work [39][40] is to identify two potential issues that have not received substantial research attention but have great impacts on conventional hop-count based localization algorithms that are designed for ad hoc sensor networks. The issues are listed as follows:

(i) *Density issue*: Localization accuracy is not guaranteed for non-uniform and sparse networks;

(ii) *Path length issue*: Cumulative error in distance estimation becomes significant for long hop-count propagation path (especially common in large networks with small number of reference nodes).

We develop a localization algorithm [39][40] that provides better position estimation for sensor nodes when the node distribution is sparse or non-uniform. We also improve the accuracy of position estimation for sensor nodes that are located far away from the reference nodes.

1.4 Scope and Outline

The rest of the thesis is organized as follows. Chapter 2 covers the introductory background of wireless sensor networks and localization algorithms that are commonly used in ad hoc networks. Some common position computation methodologies are also explained. Chapter 3 analyzes localization issues caused by non-uniform node distribution and long hop-count propagation path. It reviews the factors that can cause non-uniformity in network distribution. It also investigates the impacts of network non-uniformity and long path on localization accuracy. It presents and explains the Density-aware Hop-count Localization (DHL) algorithm that has been developed. Subsequently, Chapter 4 reports and interprets the experimentation performed to verify the algorithm presented in Chapter 3. Chapter 5 concludes the work with discussions on possible future works.

Chapter 2 Background and Related Works

This chapter gives an introductory background on ad hoc wireless sensor networks and related works on localization. Section 2.1 discusses the applications of sensor networks as well as the differences between sensor networks and ad hoc networks. Section 2.2 provides an overview of localization in wireless sensor networks, examines the constraints related to localization algorithm design, as well as studies the common techniques used in position computation. Subsequently, Section 2.3 includes a study on conventional localization schemes in wireless sensor networks. Some of the prominent and representative works are presented. The last part of this chapter coves some theoretical methods to compute sensor positions.

2.1 Wireless Sensor Networks

The maturing of microelectromechanical systems (MEMS), integration of digital circuitry, and wireless communication technology have contributed to the emergence of wireless sensor networks [1][9][12][13]. These underlying advancements in technology have made it possible to design small, inexpensive and autonomous smart sensors, e.g. Smart Dust [3], which are capable of wireless communication. A collection of these sensors can collaborate and perform much larger missions by distributed sensing.

Wireless sensor networks are task-based networks that hold the promise in the area of continuous unmanned surveillance and monitoring. Hundreds or thousands of sensors form a wireless network to perform coordinated tasks. Wireless sensor networks in hazardous environments such as remote terrain, disaster areas, toxic regions and battlefields are particularly useful. Applications include toxic leak detection, outdoor surveillance, intrusion detection, target tracking, search and rescue, obtaining micro-level information and many others. The sensing data can include the readings of surrounding temperature, humidity, light, airflow, pressure, etc. Then, the collected sensing data is transported back hop-by-hop to the sink node, where the network information is retrieved.

Some unique features distinguish a wireless sensor network from an ad hoc wireless network. Firstly, it is a sensor and actuator-based network that usually has direct interaction with physical environment. An assigned task is accomplished by collaborative effort of a group of sensors. These sensors are small, cheap, and untethered. They have modest computation and communication capabilities, as well as limited energy supply. Comparatively, an ad hoc network usually comprises of devices like handheld, laptop, etc that are larger in size, better in computation capability, improved energy supply and more costly. Besides, ad hoc devices usually have human users instead of having interaction with physical environment. In addition, the number of nodes deployed in a sensor network can be several orders of magnitude higher than an ad hoc network. The topology of a wireless sensor network changes due to node failure while that of an ad hoc network changes due to node mobility. Once deployed, the network operates unattended with minimal external management or configuration.

After a general discussion of wireless sensor network, a more specific aspect of wireless sensor network, i.e., localization, is presented next.

2.2 Localization in wireless sensor networks

Localization may be defined as the problem of determining the spatial relationship among nodes in a specific coordinate system that can be a global coordinate system or an independent local coordinate system. Localization is fundamental to wireless sensor networks since the usefulness of sensing data is inherently associated with the location where the data is derived from in the physical world. However, localization in wireless sensor networks poses significant design challenges.

From the perspective of the volume of sensors to be deployed, it is prohibitive for a network administrator to place each sensor node individually at its intended position. In many cases, wireless sensor nodes are expected to be deployed in an ad hoc manner. One common method is to airdrop and scatter the sensor nodes over an unknown region. With ad hoc deployment, one is unable to arrange or predefine the positions of the sensors beforehand. Therefore, some robust localization algorithms need to be devised for wireless sensor networks.

Some existing localization systems such as Global Positioning System (GPS) [31] can be embedded in wireless devices. However, GPS is unable to meet the constraints in wireless sensor networks in terms of cost and operational requirements, i.e., low cost and low energy consumption. In the following section, applications that demand localization information are discussed.

2.2.1 Applications of Localization in Wireless Sensor Networks

Spatial localization is of paramount importance to wireless sensor networks applications. The location information is useful for target velocity computation, data aggregation, sensor query, origins of events identification, and position-based routing. Some examples of application are elaborated below.

In habitat monitoring [7], location information is essential in determining a target's velocity. Whenever a target enters a sensor's detection range, the sink node is updated. The sensor updates the sink node with the target detection time as well as the sensor's own physical location. The sink node is then able to compute the target velocity by knowing how rapidly the target reaches different points in the network.

In a network with vast number of nodes, localization can be employed to substantially reduce the overheads of data forwarding to sink node. Data aggregation [20] is used to combine redundant data, thus reducing the volume of data sent back to the sink node. This can effectively reduce the network power consumption caused by broadcasting. Intermediate nodes require sensors' location to decide which data that are derived from different nodes can be combined. This is because the intermediate nodes need to identify the sets of data collected in the same vicinity since these data have higher probability of being similar.

In addition, with localization capability, sensors are able to decide whether they should respond to a query. For example, in a network employing Directed Diffusion [19], when an attribute-value query "Location = Region χ" is broadcasted, all nodes with matching location are expected to respond to the query and take subsequent actions. If sensors fail to respond due to false location information, this can lead to the failure of a critical mission.

Location information also plays a significant part in assisting position-based ad hoc routing protocols, such as GPSR [21] and LAR [22]. The next forwarding node is selected based on its position so that a packet can be sent to the intended destination node by as few hops as possible. This type of routing protocol routes a packet based on a node's geographical position instead of its node ID or other factors. This significantly reduces energy consumption and communication overheads.

Another example of the applications of location information is to identify the origin of an event. This is particularly useful in disaster rescue and relief operations, for example, the sensors can help to provide the location of an earthquake victim buried underneath the rubble. Thus, each sensor should possess localization capability to provide the desired location information whenever necessary.

2.2.2 Localization Constraints in Wireless Sensor Networks

Since sensors are usually unattended after deployment, localization algorithms should be robust and function with minimum configuration even when there are network constraints. Some significant network constraints are discussed below.

The major challenge in localization of wireless sensor networks is to deal with stringent constraint on energy supply. Usually, the battery energy of a sensor is not replenished once depleted. Thus, the battery energy should be preserved and a localization algorithm should minimize energy consumption. Depending on specific applications of a sensor network, sometimes coarse location estimation is sufficient. In this case, the algorithm should not be too complex at the expense of energy resources to obtain location to fine precision. Another alternative is to obtain coarse location information initially and then apply some refinement methods to reduce the error in location estimation.

Also, a sensor node may have modest communication and computation capabilities. The limited transmission power enables a node to communicate only within a short range. The limited processing power may prohibit a node from handling complex computation. Thus, an algorithm for a sensor network should be simple to implement.

Localization algorithm should not incur high cost since the sensors are supposed to be inexpensive and disposable. Besides, form factor should also be taken into consideration since miniaturization of sensor nodes has become an inevitable trend. This instantly precludes the installation of expensive, complex and bulky hardware. Currently, GPS [31] is not a suitable solution due to cost and energy consumption concerns.

Due to the unpredictable nature of physical environment, a localization algorithm should not be tightly coupled to particular environmental conditions. Instead, it should be applicable in different environments or network setting.

Another constraint to deal with is the radio range irregularity and asymmetric wireless link. Currently, ranging techniques do not offer reliable measurement. The accuracy of range measurement largely depends on the condition of transmission medium and surrounding environment. Depending on whether range measurement is needed, there are two broad classes of localization algorithm, i.e., range-based (e.g. [6],[30],[36]) and range-free localization (e.g. [17],[27],[34]). Range-based localization requires point-to-point distance to be known and these algorithms always make the assumption that the distance can be determined via methods like Time-of-Arrival (ToA) or Received Signal Strength Indicator (RSSI). The accuracy of range-based localization algorithms largely depends on the accuracy of range estimation techniques. Comparatively, range-free localization algorithms may provide a coarser estimation but are not affected by the current ranging technology.

In short, a robust localization system should be able to provide good location estimation despite the above mentioned constraints like finite energy supply, limited communication and computation resources, cost, and unreliable range estimation techniques. Thus, a localization algorithm should be distributed, simple, and scalable.

2.2.3 Localization Techniques in Wireless Sensor Networks

Position computation methodologies typically require distance or angle measurement between a node and a set of reference nodes in order to discover the node's specific location. In conventional wireless networks, these distance or angle measurements can be determined by techniques such as Time of Arrival (ToA), Time Difference of Arrival (TDoA), Angle of Arrival (AoA) and Received Signal Strength Indicator (RSSI). However, none of these techniques fit wireless sensor networks perfectly due to the inherent network constraints. The merits and drawbacks of these techniques are discussed below.

The ToA technique is capable of estimating the distance between two nodes by measuring the time taken by a signal with known speed to travel from a sending node to a receiving node. However, synchronization between these two communicating nodes is required to compute the time lapsed between signal transmission and reception. Synchronization among nodes could consume a lot of network's scarce power and bandwidth resources. One example that makes use of TOA is the GPS system [31]. GPS requires costly and energy-consuming devices to precisely synchronize a node with the satellite's clock. Like TOA technology, TDOA also relies on extensive hardware. In wireless sensor network, nodes are usually separated with short distances, ToA or TDoA requires a signal that has slower propagation speed than radio signal, such as ultrasound, to measure the time-of-flight. However, sensor nodes need to be installed with specific hardware to receive the ultrasound signals. A range estimation algorithm using this technique is proposed by Girod and Estrin [15].

To detect AoA, costly and bulky detecting component such as a directional antenna or an array of antennas needs to be attached to the sensors to measure the angle at which a signal arrives. It is not viable since sensors are small in size, disposable and low cost. Another drawback of this technique is the possibility of error introduced by multipath reflections. A localization example using AoA is a scheme proposed by Niculecu and Nath [28].

The RSSI technique is capable of translating signal strength into distance estimation since radio signal attenuates exponentially with distance. However, RSSI measurement may not be reliable due to problems such as multi-path fading, background interference, shadowing and irregular signal propagation characteristic. Some researchers propose to use averaging, smoothing and other techniques to reduce the ranging error. An example of localization based on RSSI is RADAR [2].

The drawbacks of these techniques have motivated researchers to come up with new techniques that fit well with wireless sensor networks, and one of these is the hop-count technique. The special multi-hop nature of sensor network and the vast quantity of low-mobility sensors are two major factors that enable the use of the hop-count technique.

Hop-count based localization is a range-free technique. It does not require the knowledge of absolute distance between two neighboring nodes, making it simple and appealing. Hop-count based localization is a distributed algorithm that exploits the inherent multi-hop feature of sensor networks. There is no requirement for special hardware installation or infrastructure setup to implement hop-count localization. Hop-counts can be easily obtained by network broadcasting. Since hop-count is the only essential information in distance estimation, the packet size is small and consistent. Each node only needs to communicate with its local neighbors. Some well-known hop-count based localization schemes in wireless ad hoc networks are Ad Hoc Positioning (APS) [27][29], Robust Positioning [34] and *N*-hop multilateration [36].

2.3 Related Works

There are many works done for localization in wireless and mobile networks. An analysis by Tseng et al. [38] reviews the importance and applications of location awareness in ad hoc wireless mobile networks. In another study, Hightower and Borriello [18] survey the existing research in location system for mobile computing applications.

Some localization approaches require a single and centralized node to solve the location discovery problem. For example, in the approach proposed by Doherty et al. [10], a set of geometric constraints are formed based on nodes connectivity. The constraints are solved using convex optimization by a single powerful node. In some other research proposals, particular set-up is required. For example, in GPS-less system by Bulusu et al. [4], reference nodes are required to be placed in a regular mesh pattern and separated by a constant distance. In comparison, hop-count based localization is capable of offering simple and distributed localization solution in wireless sensor networks. In the following sections, some prominent and representative localization works that make use of hop-count techniques are discussed in details.

2.3.1 Ad hoc Positioning System (APS)

Niculescu and Nath [27][29] propose a distance-vector based ad hoc localization algorithm, Ad Hoc Positioning System (APS). This algorithm uses hop-by-hop propagation capability of the network to forward distance information from the reference nodes (RN). There are four methods in measuring the distance from the reference nodes, i.e., DV-Hop, DV-Distance, Euclidean, and DV-Coordinate. Among these four methods, DV-Hop is the only method that uses hop-count information without requiring range or angle measurements.

DV-Hop comprises of three stages. In the first stage, the flooding process enables each node to obtain hop-counts from reference nodes. The process starts with the broadcast from one of the reference nodes, RN_i. Nodes that hear the broadcast discover that they are within one hop distance from RN_i. Thus, they maintain a hop-count, HC_{RN_i} =1 from RN_i and then forward this hop-count value to their neighbors. Their neighbors then increment and forward the hop-counts to their subsequent one-hop neighbors. The process is repeated successively. If the newly received hop-count is larger than a previously received value, a node simply discards the received packet. This process continues until all the RNs have broadcasted and each node has obtained minimum hop-counts from at least three reference nodes.

In the second stage, after each reference node (X_i, Y_i), accumulates hop-counts from all other reference nodes, HC_{RN_j}, (where $j=1, \ldots, K, j \neq i$, and K is the total number of reference nodes), it computes average distance per hop-count, D_{avg}. This D_{avg} is the average size of each hop. D_{avg} can be calculated since the locations of each of the other reference nodes (X_j, Y_j) are known.

$$D_{avg} = \frac{\sum \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}}{\sum HC_{RN_j}}, j \neq i, \quad j = 1....K \tag{2.1}$$

In the third stage, each of the RNs distributes its computed D_{avg} through *controlled flooding*. This means that once a node gets and forwards a D_{avg}, it will ignore the subsequent ones. Thus, most nodes will receive only one D_{avg}, and usually from the closest reference node. Subsequently, each node translates hop-counts to distances by computing the product of D_{avg} and HC. These estimated distances from three or more non-collinear reference nodes can be used to compute a node's physical location by methods such as triangulation [31]. In terms of transmission overheads for DV-Hop, the total transmission overheads can be computed by the total number of transmissions in the *first* and the *third* stage.

DV-Hop is a simple method. It is independent of errors caused by inter-node range estimation. However, according to the authors [27][29], "it only works for isotropic networks, that is, when the properties of the graph are the same in all directions". In dealing with non-uniform networks, the authors have proposed another method, namely the Euclidean method. This method is based on geometry computation. Fig. 2.1 illustrates how a node *A* estimates its distance to reference node, node *L*, by using Euclidean method. Initially, node *A* measures ranges to its two neighbors, nodes *B* and *C*. Then it

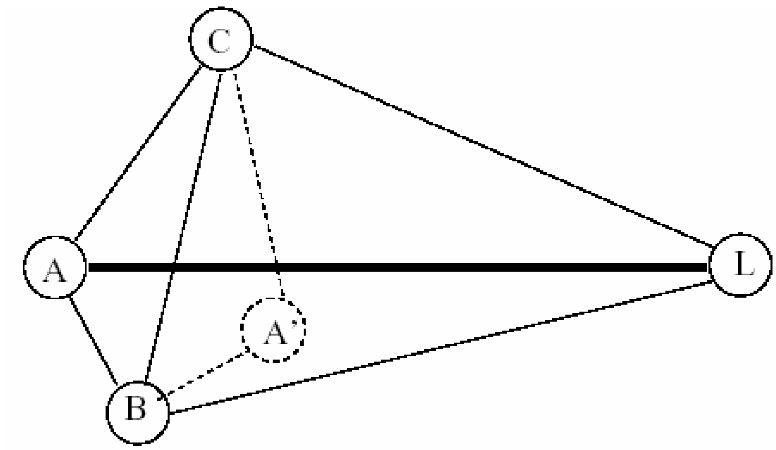


Fig 2.1 Euclidean method.

learns the distances *BC*, *BL* and *CL* by communicating with these two neighbors. Thus, a quadrilateral *ABCL* is formed. Since the length of all the sides and one of the diagonals, *BC*, are known, node *A* is able to compute the second diagonal *AL*, which is the distance between node *A* and the reference node, or node *L*.

However, according to the localization comparisons conducted by Langendoen and Reijers [24], the Euclidean method has a few issues to address. First, a node has uncertainty in choosing between two possible solutions in location (i.e., position A and A' in Fig. 2,1). Besides, two neighbors with estimated distance (i.e., node B and node C in Fig. 2.1) to a reference node are needed in computing a location, thus making many nodes unable to compute their locations in a network with low connectivity. Also, the Euclidean method is highly dependent on the accuracy of range estimation. Therefore, alternative algorithms should be devised for sparse and non-uniform networks.

2.3.2 Robust Positioning

A robust and fully distributed positioning algorithm, *Robust Positioning* [34], is proposed to estimate the locations of the sensor nodes in ad hoc wireless networks. The Robust Positioning algorithm is split into two phases: Hop-Terrain and Refinement phases. Hop-Terrain algorithm roughly estimates the positions of the nodes for further refinement in the second phase. The Hop-Terrain phase is similar to DV-Hop, where hop-counts and average hop-distance are propagated by two floodings throughout the network until, ideally, all the nodes in the network have the information from all the reference nodes in the network. The nodes compute their position using triangulation to obtain coarse positions.

In the Refinement phase, each node repeats the triangulation calculation, but this time they use their one-hop neighbors as the new reference nodes. In this phase, each node obtains the estimated positions and the ranges computed from the Hop-Terrain phase from each of its neighbors. Then, the nodes perform triangulation repeatedly to determine their new positions. This is an iterative process in which position broadcast and triangulation are repeated until certain stopping criterion is met.

However, the Refinement phase has a few drawbacks. Error propagates fast throughout the network. Firstly, an error introduced by a node would have been propagated to every node in the network by d iterations, where d is the network diameter in hop-counts. Secondly, it is a priori not unknown under what conditions the refinement will converge and how accurate the final solution is. Thirdly, according to a localization quantitative study [24], the accuracy of this refinement is highly dependent on the estimated range between neighbors. Robust Positioning also suggests that if a node has low confidence in its estimated location (for example, when it has low number of neighbors and it suspects that its estimation may not be accurate), it may be filtered out from the iterations. Since some neighbors are not involved in the iterative computation, it results in low percentage of nodes for which a position is determined [24].

2.3.3 Ad Hoc Localization System (AHLoS)

Three multilateration methods are proposed in AHLoS [35][36], i.e., atomic multilateration, iterative multilateration and collaborative multilateration. The selection of which method to be used by a node depends on the distance between the node and reference nodes and also the number of reference nodes in the network. If a node has three or more reference nodes as *immediate neighbors*, it uses simple triangulation, i.e., atomic multilateration, to determine its position. This can be done since the reference nodes are within one hop and thus the distances from the reference nodes can be measured directly by ranging techniques such as RSSI or ultrasound.

After atomic multilateration is carried out, iterative multilateration can be used to estimate the positions of nodes that do not have three or more reference nodes as immediate neighbors. In other words, iterative multilateration is a continuation of atomic multilateration. After atomic multilateration is applied, the nodes that have computed their positions are upgraded to reference nodes status. This allows the rest of the nodes to estimate their positions using these newly upgraded reference nodes. Despite the simplicity, iterative multilateration requires high reference node ratio in the network such that large fraction of nodes have at least three immediate reference node neighbors to enable every node in the network to compute its position.

Collaborative multilateration (also known as *N*-hop multilateration primitive [36]) is used if the reference node ratio in the network is low. Nodes collaborate with each other to propagate and accumulate range measurement over multiple hops. Then, the nodes estimate their positions using Min-Max technique (explained in Section 2.4.1). Two computation models, i.e., centralized and distributed, are proposed. The distributed computation model induces lower computation latency compared to the centralized model and thus it is more suitable for resource-constrained networks.

AHLoS has some setbacks. Iterative multilateration requires large number of reference nodes in the network. The multilateration computation cannot proceed if the number of reference nodes is low. Furthermore, error introduced by a node can be propagated easily throughout the network in iterative multilateration, and AHLoS is also sensitive to the accuracy of inter-node range estimation.

2.3.4 Gradient and Multilateration

Nagpal et al. [26] propose a similar hop-count localization technique by using a set of 'seed' sensors that are preprogrammed with position information. A *gradient* process, which is similar to flooding, is initiated so that each node can obtain minimum hop-counts from the seeds. The network density is assumed to be high and uniform. Multilateration is used to compute a node's position.

A refinement method, *local averaging,* is suggested; where each sensor collects its neighboring hop-count values and computes an average of itself and neighbors' values. However, this method is only suitable for evenly spaced sensors [26].

2.3.5 Mobility-enhanced Localization

Lim and Rao [25] improve the accuracy of hop-count localization by using mobile nodes to do averaging and correction. They show that by intentionally introducing a small group of mobile nodes to a network that initially comprises of only static nodes, the estimation accuracy is increased. Works from Sichitiu and Ramadurai [37], and Pathirana et al. [32] also utilize the mobility of reference nodes to compute node localization, and in comparisons, their works are based on Received Signal Strength Indicator (RSSI) range estimation instead of using hop-counts.

2.3.6 Other Works Affected by Density Issue

According to Cho and Chandrakasan [8], sensor density can range from a few to a few hundred in a region that is less than 10m in diameter. Cerpa et al. [7] point out that in habitat monitoring, the number of sensors can be 25 to 100 per region. This implies that node density is not uniform in the whole network. A region can have many times more or less sensors than the other regions in a sensor network. Therefore, the impact of non-uniform node density should be taken into consideration in hop-count localization.

Node density also affects power management, network connectivity management, and data aggregation. Intanagonwiwat et al. [20] propose a data-centric routing with in-network data aggregation mechanism so that information dissemination is energy-efficient. In a high density network, they state that the greedy-tree aggregation approach achieves more significant energy savings (up to 45%) than the opportunistic aggregation. Ganesan et al. [14] propose using multipath routing in wireless sensor networks to increase resilience to node failure. They discover that at high node density, the maintenance overhead of two-disjoint paths is nearly an order of magnitude higher than braid path. On the other hand, at low node density, they find that path construction sometimes fails to find an alternate path. The Geographical Adaptive Fidelity (GAF) algorithm [41] conserves energy by identifying nodes that are equivalent from a routing perspective and turning off unnecessary nodes. The results from GAF suggests that network lifetime increases proportionally with node density, where a four-fold increase in node density can lead to network lifetime increases by 3 to 6 times. Bulusu et al. [5] improves localization quality by placement of new reference nodes at low node density and rotating functionality among redundant reference nodes at high node density. Thus, node density is an interesting issue not only in localization, but also in other areas in sensor networks.

2.4 Position Computation Methodologies

To determine a node's specific location within a coordinate system, some position computation methodologies are needed. The complexity of localization computation with distance estimated is analyzed theoretically in [11]. Two techniques used to solve for unknown locations are explained below.

2.4.1 Triangulation

Triangulation [31] is a computation technique used to locate nodes within a coordinate system. A node's location is uniquely identified when at least three reference nodes are associated with it in a two-dimensional space, or at least four reference nodes in a three-dimensional space. Triangulation can be computed by a node when distances or angles from reference nodes are known. The algorithm in this thesis makes use of one form of triangulation, known as lateration, in which only the distances from reference nodes are considered. The computation is explained below.

After an arbitrary node with position (u,v) obtains estimated distances, $d_1, \ldots, d_K$, from K number of RNs which have corresponding positions of $(X_1,Y_1),\ldots,(X_K,Y_K)$, the following equations are derived:

$$
\begin{aligned}
&(X_1 - u)^2 + (Y_1 - v)^2 = d_1^2,\\
&\vdots\\
&(X_K - u)^2 + (Y_K - v)^2 = d_K^2,
\end{aligned}
$$

The list of equations can be linearized by subtracting the last row of equation from the previous K-1 equations.

$$
\begin{aligned}
&X_1^2 - X_K^2 - 2(X_1 - X_K)u + Y_1^2 - Y_K^2 - 2(Y_1 - Y_K)v = d_1^2 - d_K^2,\\
&\vdots\\
&X_{K-1}^2 - X_K^2 - 2(X_{K-1} - X_K)u + Y_{K-1}^2 - Y_K^2 - 2(Y_{K-1} - Y_K)v = d_{K-1}^2 - d_K^2
\end{aligned}
$$

Reshuffling the equations gives a proper system of linear equations in the form of $AU=b$, where

$$
A = \begin{bmatrix} 2(X_1 - X_K) & 2(Y_1 - Y_K) \\ \vdots & \vdots \\ 2(X_{K-1} - X_K) & 2(Y_{K-1} - Y_K) \end{bmatrix},
$$

$$
U = \begin{bmatrix} u \\ v \end{bmatrix},
$$

$$
b = \begin{bmatrix} X_1^2 - X_K^2 + Y_1^2 - Y_K^2 + d_K^2 - d_1^2 \\ \vdots \\ X_{K-1}^2 - X_K^2 + Y_{K-1}^2 - Y_K^2 + d_K^2 - d_{K-1}^2 \end{bmatrix}.
$$

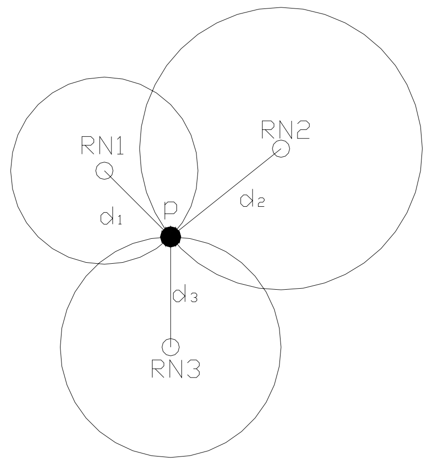


Fig 2.2 Position computation using Lateration.

The solution of the above matrix, $U = (A^TA)^{-1}A^Tb$, can be obtained by using a standard least-squares approach [16]. Using triangulation, an object is uniquely positioned when distances from at least three non-collinear reference nodes are known in a two-dimensional space.

To illustrate how a node *p*, computes its position using triangulation technique, consider a two-dimensional space with three reference nodes, RN_1, RN_2, and RN_3 (Fig. 2.2). After *p* obtains its first distance from RN_1, d_1, it can deduce that its possible location is a point on the circumference of the circle of radius d_1 centered at RN_1. The second distance from RN_2, d_2, reduces the possible locations of *p* to two, which are the two intersection points of the two circles, centered at RN_1 and RN_2 respectively. With the knowledge of third distance from RN_3, d_3, the position of *p* is confirmed, which is the point where the three circles intersect exactly.

This concept can be extended to a three-dimensional space if there is at least one more reference node. From the first known distance, *p* can conclude that it is a point on the surface of the sphere of radius d_1 centered at RN_1. The second distance reduces the possibilities to a circle, which is on a two-dimensional plane. Then, the third and forth distances would finally determine the position of *p*, as explained in the two-dimensional case above.

2.4.2 Min-Max

Another position computation technique which is simpler but provides coarser solution is Min-Max operation. After an arbitrary node, *p*, obtains estimated distances, $d_1, \ldots, d_m$, from the reference nodes *1, ..., m*, bounding boxes that enclose the circles originating from each reference node with radii of $d_1, \ldots, d_m$, are constructed (Fig. 2.3). The four edges of a bounding box from a reference node *i* can be created by adding and subtracting the estimated distance d_i from reference node position (X_i, Y_i), as shown below.

Top edge => $Y_i + d_i$;

Bottom edge => $Y_i - d_i$;

Left edge => $X_i - d_i$;

Right edge => $X_i + d_i$;

Then, the intersection of the bounding boxes is determined by taking the maximum of all coordinate minimums and the minimum of all coordinate maximums.

Top edge => $\min(Y_1 + d_1, \cdots, Y_m + d_m)$

Bottom edge => $\max(Y_1 - d_1, \cdots, Y_m - d_m)$

Left edge => $\max(X_1 - d_1, \cdots, X_m - d_m)$

Right edge => $\min(X_1 + d_1, \cdots, X_m + d_m)$

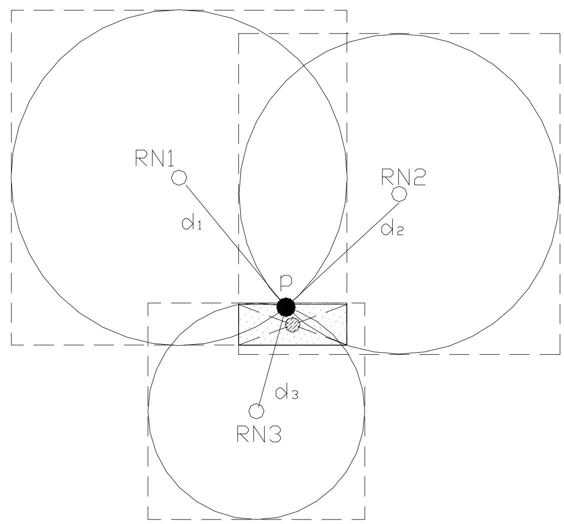


Fig. 2.3 Position computation using Min-Max operation.

The estimated position of the node is set to the intersection of this small bounding box. The estimated coordinates are the average values from the four corner coordinates.

2.5 Conclusion

In Chapter 2, introductory background of ad hoc wireless sensor networks, localization schemes and mathematical computation methodologies are reviewed. In the following chapter, the issues our algorithm addresses are discussed and our algorithm is presented.

Chapter 3 Density-aware Hop-count Localization (DHL) Algorithm

The problem of localization, i.e., determining where a node is physically located in a particular coordinate system, is crucial for many applications in wireless sensor networks. Yet, the inherent network constraints pose challenges to the design of robust localization algorithms. As discussed in Chapter 1, two potential issues in conventional hop-count localization algorithms are identified: (a) *density issue*; and (b) *path-length issue*. In this thesis, the main goals of our algorithm are to address these two issues and to provide a localization solution that is suitable for sparse and non-uniform ad hoc wireless sensor networks.

In the subsequent sections, we discuss how the abovementioned issues arise in wireless sensor networks. Section 3.1 presents an overview of the issues being addressed and the algorithm being proposed. Section 3.1.1 and Section 3.1.2 investigates *density issue* and *path-length* issue respectively. Section 3.1.3 discusses Density-aware Hop-count Location (DHL) algorithm in details. Subsequently, Section 3.2 describes the method to determine parameters in DHL whereas Section 3.3 presents the complexity of communication overheads in DHL. Lastly, Section 3.4 concludes Chapter 3.

3.1 Density-aware Hop-Count Localization (DHL) Algorithm

In the Density-aware Hop-count Localization (DHL) [39][40] algorithm, the sensor network is assumed to be fully connected and there is no node partition. The sensors have moderately low mobility. Due to broadcast nature of wireless channel, each node is assumed to know the number of its neighbors after a network is deployed. An omni-directional radio propagation model and a 2D network model that is extendable to 3D are assumed. The radio range of the sensors is denoted by R.

In our network model, there exists a total of N sensors, of which only K sensors (where $0<K<N$), known as reference nodes/sensors, are equipped with position information while the rest of the nodes seek to discover their positions through multi-hop communication. Two nodes can communicate if their distance is less than R, where R is the radio range (which varies with the transmission power and technology used). Local density is defined as the number of neighboring nodes per unit transmission area. For simplicity, the number of neighboring nodes or local connectivity, c, is used to estimate the density surrounding a node. We also define the incremented distance by traveling a hop as *hop-distance*.

We describe in detail the two issues of concern, i.e., *density issue* and *path length issue*, below before presenting the details of the algorithm.

3. 1.1 Density Issue

3.1.1.1 Factors of Density Variation

Most sensor networks are deployed outdoors. Thus, the sensor distribution can be affected by various factors, as elaborated below.

a) Method of deployment and terrain contour

The number of sensors to be deployed in a wireless sensor network can be substantial; a network may be composed of hundreds or thousands of nodes. Thus, manual deployment is not a simple task and sensors are more commonly deployed in ad hoc fashion, via means like air drop or artillery launch. However, using ad hoc deployment, sensor distribution tends to be affected by terrain contour. For instance, sensors tend to accumulate at the bottom of a slope or hilly terrain, thus, causing node density to be higher at the bottom than at the peak of a slope. Ad hoc deployment and terrain contour makes it difficult to decide accurately the location and orientation of each sensor node.

b) Hostile environment

Environmental obstacles can prevent nodes from being placed at certain intended locations in order to create a uniform network. Thus, it is not easy to ensure uniformity in node distribution. Even if nodes can be impeccably placed at the beginning of network deployment, hostile environment or unpredicted weather can alter a node's position or cause it to malfunction. For example, sensors can be swept away by strong current, corroded by harsh chemical solution, moved away by animals or damaged by the enemies. Thus, network density and sensor distribution can be easily altered.

c) Network dynamism

Over time, the battery power of a sensor may have dwindled to a level where the node can no longer be active at all times. In a worse case, the energy of a sensor is depleted and it is no longer functioning. Also, a node may move out of transmission range of the other nodes. Due to network dynamism, nodes may switch between active and sleep modes, enter functional or breakdown states, join or leave a network from time to time. Therefore, node density is not consistent throughout a sensor network lifetime.

3.1.1.2 Euclidean Distance and Range Ratio

Assume an arbitrary reference node, P_i, $i=1,\ldots,K$, is deployed at a point (X_i, Y_i). For an arbitrary sensor S_j at (u_j,v_j), $j=1,\ldots,N\text{-}K$, we denote the Euclidean distance between them as $d(P_i, S_j)=\sqrt{(X_i-u_j)^2+(Y_i-v_j)^2}$. We define the Euclidean path as a path consisting of the minimum number of hops, m, to propagate a packet from P_i to S_j, i.e., where $d(P_i, S_j) = mR$ (Fig. 3.1a). If sensor deployment is very dense and uniform (Fig. 3.1b), a path approximating the Euclidean path can be constructed, and S_j is able to approximate its distance from P_i by $d(P_i, S_j) \approx m(\lambda R)$, where m is the minimum hop-count and λ is the average range ratio; λR is also the average *hop-distance (*i.e., distance per hop).

However, in a non-uniform network (Fig. 3.1c), the variance of hop-distance is high, causing $d(P_i, S_j) \approx \mu_1 R+ \mu_2 R+\ldots+ \mu_m R$, where m is the minimum hops and μ is the range ratio ($0<\mu\leq 1$). μ is a function of an intermediate node's local density (i.e., connectivity/πR^2) since hop-distance depends on the availability of the next node close to

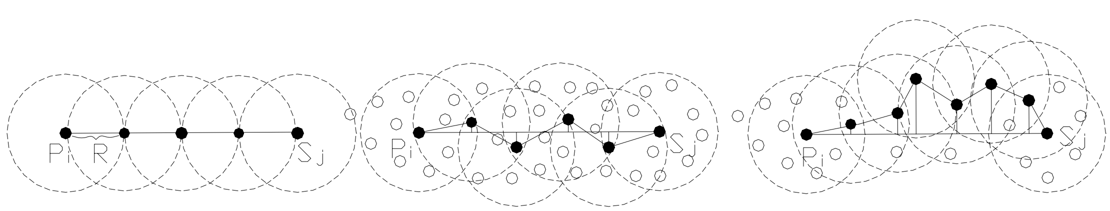


Figure 3. 1 (a) Euclidean distance, (b) uniform network, (c) non-uniform network

the transmission range and at the direction of propagation, i.e., $\mu=f(D)$, where D is local density.

In Fig. 3.2, when a network is dense and uniform, using hop-count parameter, the estimated distance between an arbitrary node can be approximated accurately. For example, in Case 0, the distance between the reference node and destination node is approximated by "$2 \times R$". The estimation gives close approximation to the actual distance. However, if the node density is sparse, three cases can arise that can cause over-estimation of the actual distance. The following three cases illustrate how distance overestimation happens.

a) Case 1

If the next forwarding node is not located sufficiently close to the transmission boundary, the distance traversed for each hop does not equate to the propagation range (Fig. 3.2a). Thus, more hops are taken in order to propagate the packet to the intended node.

b) Case 2

The next forwarding node is located on the boundary of the transmission range. However, the end-to-end path taken is not straight (Fig. 3.2b). The winding and twisted path taken accumulates more hop-counts.

c) Case 3

This is a hybrid case of the previous two cases (Fig. 3.2c). Some of the forwarding nodes are not close to the transmission boundary. In addition to that, end-to-end forwarding path is not straight. Case 3 usually causes greater distance over-estimation compared to Case 1 and Case 2.

Thus, it can be summarized that in a sparse network, the actual distance for each hop is less than R since ***the probability of finding a point close to the boundary in the***

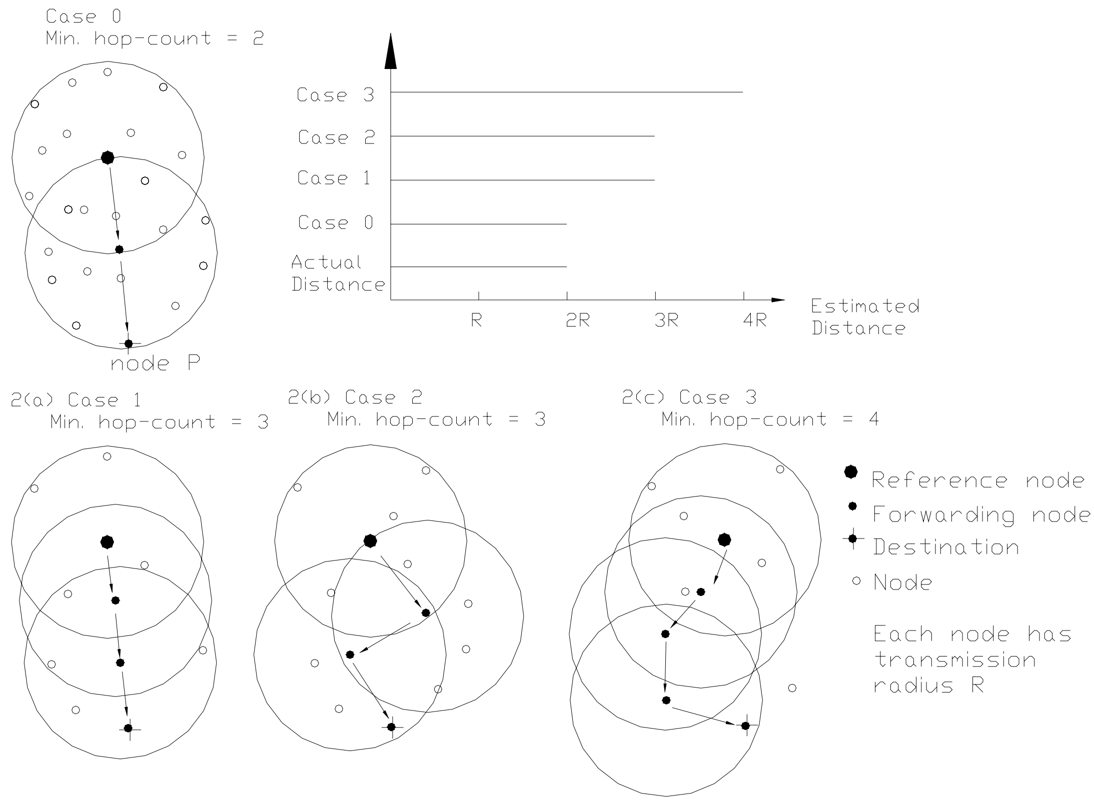


Fig. 3.2 Comparison of distance over-estimation due to (a) Case 1, (b) Case 2, and (c) Case 3.

direction of travel diminishes. In other words, the probability of having sufficient nodes constituting straight and short paths in hop-count propagation directions decreases. The variation of hop-distance is directly affected by the degree of uniformity of node distribution.

As discussed in Section 2.3.1, DV-Hop uses average distance per hop-count, D_{avg}, as a correction. The purpose of using D_{avg} is to reduce distance over-estimation when the network is sparse and uniform. When the node density is high, the probability of having a straight and short hop-count propagation path is high. In this case, DV-Hop shows highly accurate estimation (Fig. 3.3a). If the number of nodes in the network is smaller, but the overall node distribution is still uniform, D_{avg} is computed as a smaller value to account for the decreased distance per hop-count (Fig. 3.3b). This is because the propagation path tends to be a winding one, thus the distance traversed in each hop is shorter. In this case, using D_{avg} as a correction also shows good estimation (Fig. 3.3b). However, this is not the case when a network has a mix of dense and sparse regions, i.e., non-uniform node distribution. Using DV-Hop shows degraded performance (Fig. 3.3c). This is because the distance traversed for each hop is no longer consistent. The distance per hop is generally greater in dense regions and generally shorter in sparse regions.

DV-Hop points out that its drawback is that localization only gives good performance if the network is isotropic, that is, when the properties of the graph are the same in all directions [27][29]. In fact, Langendoen and Reijers[24] who conducted comparisons of distributed localization algorithms stated that "a drawback of DV-Hop is that it fails for highly irregular network topologies, where the variance in actual hop-distance is very large".

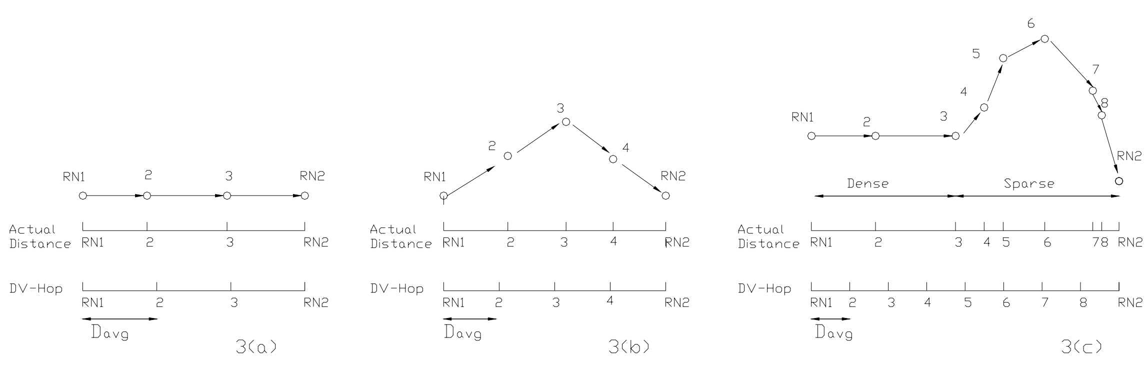


Fig.3.3 Estimated distance from RN1 by DV-Hop in a (a) uniform and high density network, (b) uniform and low density network, (c) non-uniform network.

In actual physical deployment, the node distribution in wireless sensor networks is unpredictable; thus, we face the challenge of devising hop-count localization that can accommodate networks with non-uniform distribution. In view of this, density-awareness is an issue worth exploring to extend the implementation of hop-count based localization to non-uniform networks.

Thus, we propose to incorporate density awareness and assign hop-distance dynamically based on a node's local density.

3. 1.2 Path Length Issue

A downside of distance summation using hop-count localization is that estimation error accumulates when hop-count is incremented over multiple hops. This cumulative error becomes increasingly significant with the increase of hop-counts. It happens especially for large networks with few reference nodes where long propagation paths tend to take place.

For each hop, the actual traversed distance is either less than or equal to transmission range, R. This difference between actual progressed distance and transmission radius is accumulated with the increase of hop-counts. Therefore, the distance estimation error tends to increase with hop-counts.

In reality, an estimated hop-distance, L, is imprecise and the uncertainty should be reflected in the expression, i.e., $L \pm \varepsilon$, where ε is the maximum error. If S_j is m hops from P_i, its estimated distance is $m(L \pm \varepsilon)$, i.e., $mR[\lambda \pm \varepsilon/R]$ (uniform networks) or $R\sum^{m}[\mu_i \pm \varepsilon_i / R]$ (non-uniform networks). From the two equations, when R is infinitely large and sensors are within hearing range from one another, the error is negligible, but this is infeasible since the transmission power of sensors is limited. However, error can be reduced if the distance is associated with fewer hops. To further improve the performance of our scheme, path-length is taken into account in DHL, where an estimated distance computed from a comparatively fewer number of hops *m* is given a higher confidence rating.

Realizing that there is room for improvement in dealing with the *density* and *long path* issues, our algorithm introduces density-awareness to adjust hop-count in non-uniform networks [39] and path-length awareness [40] to reduce errors in large networks with low reference node density. Hop-distance is adjusted based on a node's surrounding density whereas distances computed from large hop-count are identified as potentially having larger errors, and thus given a lower confidence level.

Thus, the distance between a reference node and an arbitrary node is not easily approximated accurately due to high variance of hop-distance and accumulated error. Therefore, we require a novel algorithm to handle this density issue in sensor localization. Next, we explain our DHL algorithm in the following section.

3. 1.3 Main Algorithm

Unlike conventional hop-count localization algorithms, DHL [39][40] does not require network-wide uniformity. Within a network, some regions may have higher or lower density. This type of non-uniform node distribution is more often encountered in actual network scenarios. The neighbors of a node are assumed to be distributed randomly around the node.

We define the incremented distance by traveling a hop as *hop-distance*. Depending on local connectivity, we classify the node density into a few categories and each category has a corresponding range ratio. Range ratio, μ, represents the ratio of expected hop-distance to the transmission range for a particular local density. The algorithm strives to integrate density-awareness when propagating hop-counts throughout the network. Range ratio is a function of local density, i.e., a sensor node's connectivity per unit transmission coverage.

Due to the broadcast nature of wireless channels, S_j is assumed to know its local density after a network is deployed. A network manager predefines a set of density categories, e.g., low, medium, high, etc, and each category covers a certain range of local density. A sensor, S_j, deduces the category it falls into based on its local density. Each category is mapped to a corresponding range ratio μ that reflects the ratio of transmission range a packet most probably advances if forwarded to the next hop. The selection of number of density categories is a tradeoff between accuracy and overhead. Increasing the number of categories can increase the accuracy of expected hop-distance, but at the expense of higher number of exchanged messages. The flow of the algorithm is described below and the methods to determine the range ratio and confidence level are described in the next section.

We perform a one-time computation, as follows:

Step A: P_i broadcasts a set of tuples, consisting of {ID(P_i), Position(P_i), Total Hops to P_i, Total Range Ratio to P_i }, i.e.,{ID, (X_i, Y_i), $\sum k_i=0$, $\sum\mu_i=0$}.

Step B: S_j stores {ID, (X_i, Y_i), $(\sum k_i)+1$, $(\sum\mu_i)+\mu$}and forwards the information.

Step C: S_j estimates distance to P_i by $L_i=(\sum\mu_i)\times R$.

Step D: If S_j subsequently receives packet with smaller $\sum k_i$ or $\sum \mu_i$, it repeats Step *B* to *C*.

Step E: S_j associates L_i with a low or high confidence rating, as described in the next section. When sufficient number of distances from the reference nodes is received, S_i will perform triangulation. Where possible, only L_i associated with high confidence

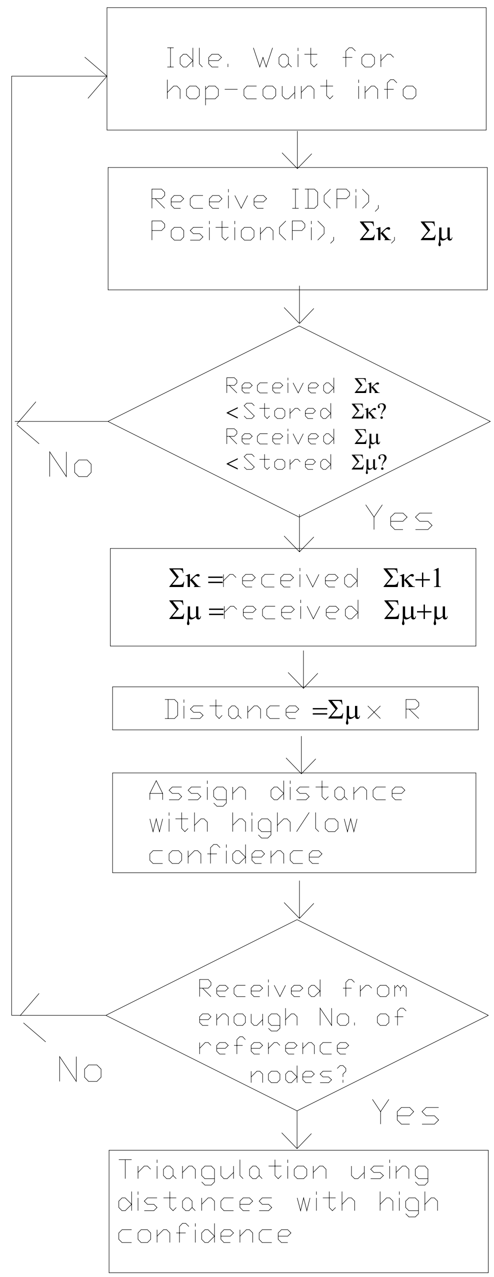


Fig.3.4 Flow chart showing the states a node enters in DHL

rating will be used.

Flow chart shown in Fig. 3.4 gives a detailed description the states a sensor enters and also the actions the sensor performs in DHL.

The algorithm is basically divided into two phases. The purpose of the first phase, *Density-aware Phase* (*Step A* to *D*), is to enable individual nodes to share hop-count information collaboratively in order to determine their distances from individual reference nodes. The hop-count information incorporates density information so that it provides more accurate distance estimation. In the second phase, *Path-Length aware Phase* (*Step E*), a node determines the confidence level for each estimated distance and decides if the distance should be used in position computation using triangulation. The first phase uses a node's local density information to address the *density issue*, whereas the second phase assigns confidence level to address the *path length issue*.

In *Step A*, a reference node broadcasts information that consists of its ID, its position, total number of hop-counts from itself and total range ratio to itself. Immediate neighbors that hear the broadcast discover that they are within one hop from the reference node. Thus, in *Step B*, the total number of hop-count from the reference node is incremented by one. The range ratio, μ, is estimated individually based on the receiving node's surrounding density. Subsequently the receiving node forwards the information that consists of the reference node ID, reference node position, the new total hop-count and the new total range ratio. In *Step C*, a receiving node estimates its distance from the reference node by computing the product of total range ratio and transmission range. The rest of the nodes repeat the same procedure, i.e., increment received hop-count and range ratio and then forward the information. If a node subsequently receives hop-count information that gives smaller total number of hop-count, it discards the old stored values and repeats *Step B* to *Step C*. The frequency of repeating *Step B* to *Step C* mainly depends on the uniformity of the network. In a non-uniform network, a node has higher tendency to receive different total range ratio from time to time, thus causing a new round of range ratio re-adjustment and re-broadcast.

In *Step E*, each estimated distance is associated with a 'Confidence level' whose value is in the range of [0,1]. The confidence level is inversely proportional to the number of hop-counts from a reference node. This is because comparing actual hop-distance and transmission range, the actual hop-distance can be equal to or less than the transmission range. If a localization algorithm assumes that hop-distance is equivalent to one transmission range, the shortfall from transmission range becomes estimation error. Thus, localization error accumulates with increasing hop-counts. Also, the chance that a propagation path is straight and direct decreases as path length becomes longer. A winding path tends to accumulate more unnecessary hop-counts than a direct path. Consequently, a sensor node that is positioned far from a reference point tends to accumulate more errors. After assigning the confidence level, a node can select only those estimated distances with high confidence and ignore those with low confidence in position computation by methods such as triangulation [31].

We illustrate the difference in computing hop-distance between DHL and a general hop-count localization algorithm that does not make use of local density information (DV-Hop [27][29]) in Fig 3.5. As shown in the figure, if hop-count is propagated from reference node RN1 to reference node RN2 through regions with different densities (a high density region, followed by a low density region and another high density region), for DV-Hop, increasing hop-distance by an average distance does not show good performance. In contrast, applying DHL, hop-distance is increased by greater extent in dense region and lesser extent in sparse region. This gives better distance estimation. In this example, the hop-distance traveled in a dense region is the distance between RN1 and node 2, $D_{RN1\text{-}2}$ whereas the hop-distance traveled in sparse

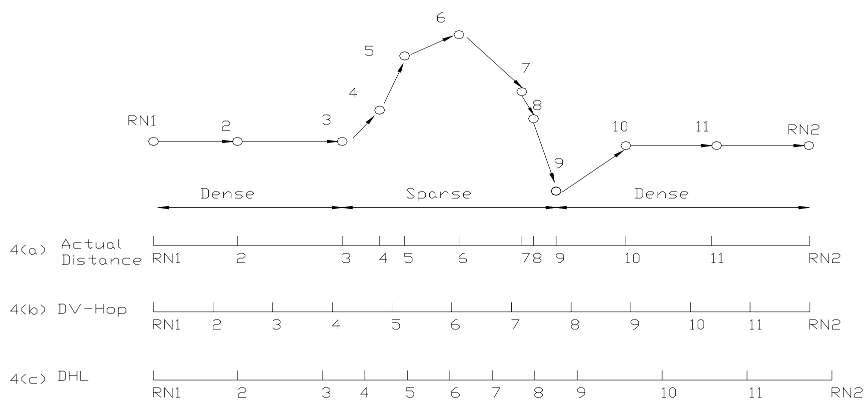


Fig. 3.5 Comparison of (a) actual distance from RN1, (b) estimated distance from RN1 by DV-Hop, (c) estimated distance from RN1 by DHL.

region is ½$D_{RN1\text{-}2}$. By using range ratio, when the density is low, each hop traversed is not necessarily equivalent to one hop-count. Thus, distance overestimation in sparse regions in the network is accounted for.

3. 2 Determination of Range Ratio and Confidence Level

Fig. 3.6 illustrates how hop-distance is affected by high and low local density. In the diagram, a node, N_a, propagates hop-count packets to an arbitrary node, N_c, that is multi-hop away via the shortest path. They are separated by a distance D_1+D_2. To propagate as close as possible in absolute distance to N_c at the next hop, D_1 should be

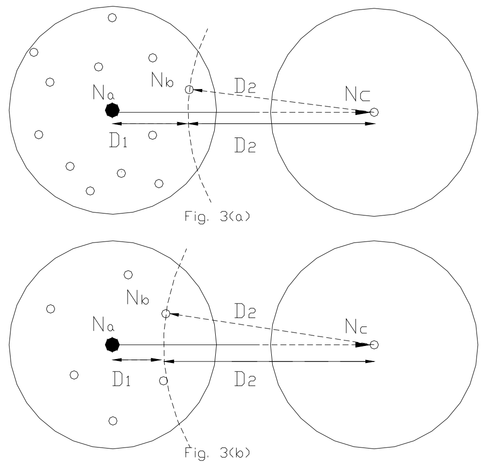


Fig.3.6 Hop-distance due to (a) high local density, (b) low local density

maximized. Thus, hop-distance can be represented by D_1, i.e., maximum distance traversed such that it is closer to node N_c in a hop. If the local density of N_a is higher (Fig. 3.6a), the hop-distance, D_1, tends to be larger compared to the hop distance, D_2, when the local density is lower (Fig. 3.6b). Thus, range ratios are used to reflect the ratio of hop-distance to the transmission radius and its value is larger if a node's local density is higher. In the ideal case where the local density is infinite, the range ratio has a value of one.

We now describe how we determine the range ratio to be used in our scheme. Range ratio as a function of local connectivity, c, has been derived in [23]. Using a continuous function to determine the range ratio can result in unlimited density categories and immense transmission overhead. If densities are divided into n categories, a node at m hops from a reference node can potentially receive n+(n-1)(m-2) different accumulated range ratios, triggering more packet forwarding. We decided to take a more heuristic

TABLE 3.1
RANGE RATIO FOR DIFFERENT DENSITY CATEGORIES

Categories	Local Density	Range Ratio
Low Density	1-6	μ_l=0.6
Medium Density	7-12	μ_m=0.7
High Density	>12	μ_h=0.8

approach by investigating the relationship between local connectivity and range ratio through simulations (Section 4.2). To create a network of connectivity *c*, a total of $cA/(\pi R^2)$ nodes are created randomly, where *A* is the network area. We define "accuracy" as the percentage of nodes with estimated locations that are within one transmission range from their actual locations. From the results, we decided to use three main categories with three corresponding optimum range ratio (Table 3.1).

We next describe the way we determine the confidence level to use. Assuming network diameter is *x*, a distance computed from more than $\frac{x}{R}$ hops is unlikely to approximate a Euclidean path and thus can be associated with low confidence level. Since a node requires at least three (four) reference nodes to perform triangulation for 2D (3D) networks, it assigns hop counts from the three (four) nearest reference nodes with high confidence. A confidence threshold can be determined within the range of *y* and $\frac{x}{R}$ to select hop counts with high reliability, where *y* is the largest hop counts from among the three (four) nearest reference nodes. For simplicity, a node can assign hop counts from other reference nodes with high confidence if they are less than $\frac{1}{2}(\frac{x}{R}+y)$. Only hop counts from reference nodes with high confidence levels will be used in the triangulation.

3.3 Communication Overheads

Generally, conventional hop-count localization requires *two* separate flooding stages, i.e., one for (a) Hop-Count Accumulation, and another one for (b) Correction. In the Hop-Count Accumulation flooding stage, hop-count information is disseminated from each reference node to, ideally, all the nodes in the network so that each node has a coarse estimation of its position. In the second Correction stage, the flooding can be used to spread information that enhances the estimation accuracy. For example, DV-Hop [27][29] broadcast average hop-distance to every node in the network through controlled flooding, Robust Positioning [34] disseminates each node's coarse position for subsequent iterative triangulation computations, and Gradient and Multilateration [26] broadcasts each node's coarse position for local averaging. However, network-wide flooding is an expensive process since it involves every node in the network. Since each node is involved in storing and forwarding the information, a lot of energy is consumed for computation and communication. Flooding also causes scaling problems. The overhead increases linearly with the number of nodes and reference nodes ratio in the network. While increasing reference nodes ratio in the network aids in increasing the localization accuracy, it also tends to increase communication overheads between nodes.

Some algorithms propose using Time-to-Live (TTL) to limit the number of hop-count propagation [27], so that transmission overhead is reduced. This method can be used only if the reference nodes ratio in the network is high. Otherwise, a large fraction of nodes in the network may not be able to receive sufficient information to compute their positions. *Flood limit* parameter is another proposal to reduce communication overhead. A node stops forwarding once it has received hop-count information from "*flood limit*" [24] number of reference nodes. However, if reference nodes initiate hop-count broadcasting at different times, a node may receive information from reference nodes that are further away and stop forwarding once the "*flood limit*" number of reference nodes has been reached. Thus, the node is unable to take advantage of hop-count information from nearer reference nodes that initiate the flooding later. As explained in Section 3.2, distances computed from smaller hop-counts tend to have better accuracy.

In comparison, DHL has less concern of reference nodes ratio and flooding initiation time. DHL combines the correction process in the hop-count accumulation stage to account for the localization errors caused by density variation. When hop-count is accumulated in the flooding process, the correction by range ratio is applied simultaneously to all the nodes in the network. Therefore, we do not require a separate flooding stage to forward the correction.

This combination approach effectively helps to reduce the number of transmitted messages, conserve network energy, and reduce the time consumed in computing a node's position. Comparing the first flooding stage, DHL has slightly more packet transmissions due to more hop-count adjustment. However, since DHL eliminates the second flooding stage, the total number of packets transmitted by DHL is less than that required by conventional hop-count localization.

3.4 Conclusion

In this chapter, the two issues in conventional hop-count localization algorithms, i.e., *density issue* and *path-length issue*, are discussed. Our algorithm, Density-aware Hop-count Localization (DHL), which addresses these two issues, is presented. In the following chapter, verification and experimentation results are given, where the performance of DHL is compared against DV-Hop by simulations.

Chapter 4 Simulation Results

4.1 Simulator Program

To evaluate and analyze the performance of DHL, we conducted simulations using a discrete event-driven simulator written in C language. The discrete-event simulator initializes the entire simulation by reading the network parameters and creating the appropriate network size, number of nodes and number of reference nodes. The simulator consists of a single event-list managed by a scheduler function. A broadcast from a reference node is designated as an event. A broadcast from a reference node triggers hop-count packet forwarding process in the network. The hop-count packet is incremented and forwarded by each node in the network. Each re-broadcast is an event, and thus, a sequence of events is generated. Each event is associated with a processing time. This time designates when the event should take place. These events are queued into a list to be processed when the virtual simulator time reaches the specific processing time.

To manage the list of events (sending and receiving of hop-count packets), the discrete-event scheduler maintains a data structure. This data structure is essentially a time-ordered queue of events. Any event occurring is queued into the list. Some events may trigger additional events that will subsequently be added to the queue according to the time it is supposed to occur. The discrete-event scheduler basically inserts each event into the queue, and then processes the queue in temporal order. When it processes an event, it also updates the simulation clock accordingly.

After every node has obtained the hop-count information (Density hop-counts, and Normal hop-counts), a node, N_k, computes its estimated distance from each reference node, RN_j, by

$$d_{k-RN_j} = DHC_{k-RN_j} \times R, \tag{4.1}$$

where $j=1, \ldots, m,$ (m is the total number of reference nodes in the network), d_{k-RN_j} is the distance between node N_k and RN_j, and DHC_{k-RN_j} is the accumulated Density Hop-Counts between node N_k and RN_j.

Using the estimated distance, a simple triangulation is used to obtain the estimated position of node N_k, i.e., $(\tilde{u}_k, \tilde{v}_k)$ so that the solution is as close as possible to the actual position (u_k, v_k). The basic idea is to solve the following set of equations between a node and each reference node, i.e.,

$$\left(X_{RN_j} - \tilde{u}_k\right)^2 + \left(Y_{RN_j} - \tilde{v}_k\right)^2 = d^2_{RN_j-k}, \tag{4.2}$$

where (X_{RN_j}, Y_{RN_j}) is the coordinate of RN_j , where $j=1, \ldots, m$ and (u_k, v_k) is the coordinate of node N_k.

Then, a least mean square method is used. Equations for $j = 2$ to m are subtracted by the equation for j =1, thus, the following set of equation is obtained,

$$(X_{RN_j} - X_{RN_1})\tilde{u}_k + (Y_{RN_j} - Y_{RN_1})\tilde{v}_k = \frac{1}{2}(X^2_{RN_j} + Y^2_{RN_j} - X^2_{RN_1} - Y^2_{RN_1} + d^2_{RN_1-k} - d^2_{RN_j-k}), \tag{4.3}$$

where j = 2 to m. Subsequently, Eqn. 4.3 is represented by A $\begin{bmatrix} \tilde{u}_k \\ \tilde{v}_k \end{bmatrix}$ =B, where

$$A = \begin{bmatrix} \left(X_{RN_2} - X_{RN_1}\right) & (Y_{RN_2} - Y_{RN_1}) \\ \vdots & \vdots \\ \left(X_{RN_m} - X_{RN_1}\right) & (Y_{RN_m} - Y_{RN_1}) \end{bmatrix}, \tag{4.4}$$

$$B = \begin{bmatrix} 1/2(X^2_{RN_2} + Y^2_{RN_2} - X^2_{RN_1} - Y^2_{RN_1} + d^2_{k-RN_1} - d^2_{k-RN_2} \\ \vdots \\ 1/2(X^2_{RN_m} + Y^2_{RN_m} - X^2_{RN_1} - Y^2_{RN_1} + d^2_{k-RN_1} - d^2_{k-RN_m} \end{bmatrix}. \tag{4.5}$$

Then, the following equation is solved by using pseudo-inverse of matrix.

$$\begin{bmatrix} \tilde{u}_k \\ \tilde{v}_k \end{bmatrix} = (A^T A)^{-1} A^T B \tag{4.6}$$

Position accuracy is then computed by comparing the obtained position ($\tilde{u}_k, \tilde{v}_k$) and the actual position (u_k, v_k). Distance accuracy is also computed by comparing the estimated distance and the actual distance from RN_j where $j=1, ..., m$.

4.2 Range Ratio Determination

First, simulations are conducted to investigate how localization accuracy is affected by range ratio for a uniform network of α local density. If on average each node has α neighbors and by incrementing received hop-count by a constant range ratio, we determined the percentage of nodes with estimated locations within the accuracy of less than one transmission range, *R*, from their actual locations. We evaluated each simulation scenario over 50 trials for a network of $50 \times 50\text{m}^2$ area and 5m transmission range; range ratio is increased by 0.1 at each step from 0.1 to 0.9. Simulations are conducted for local densities 6 and above since networks start showing severe partitioning for local density less than 6.

Fig. 4.1 shows the accuracy results of using different range ratio for average local densities from 6 to 20. Simulation results show that for a network with average local density of 6, if each node increments its received hop-count by a range ratio of 0.6, the localization accuracy is the highest compared to the use of any other values of range ratio. Similarly, for local density 7 and 12, the optimum range ratio found from the simulations is 0.7. Simulation results for local density between 7 and 12 also show similar trend, i.e., the highest accuracy is achieved when range ratio is 0.7. For local

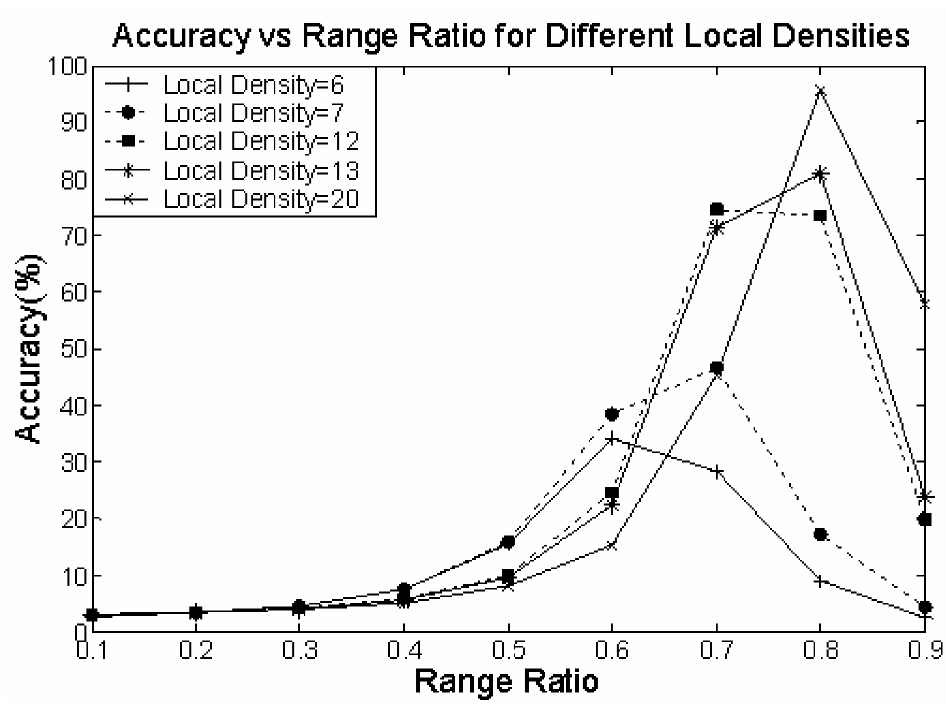


Fig. 4.1 Localization accuracy vs. range ratio for variable local densities.

density 13 and 20, the optimum range ratio is 0.8 and similarly for the cases when local density is between 13 and 20.

The results from the simulations serve as a guide for our selection of local densities and the corresponding range ratio for low, medium and high density categories (Table 4.1).

TABLE 4.1
RANGE RATIO FOR DIFFERENT DENSITY CATEGORIES

Categories	Local Density	Range ratio
Low Density	1-6	$\mu_l = 0.6$
Medium Density	7-12	$\mu_m = 0.7$
High Density	>12	$\mu_h = 0.8$

If a node has local density of 6 or below, we regard it as having low density. This is because from the simulations, networks are not fully connected if local density is less than 6. The optimum range ratio for local density 6 is assigned to a value of 0.6 from the results shown in Fig. 4.1.

For sensor networks, a local density of 10 is generally perceived as common. Thus, local density close to 10 is regarded as medium density. From the simulations, local density of 7 to 12 shares the same optimum range ratio, i.e., 0.7, in the simulations. Thus, they are assigned to the same density category, i.e., medium density. Local density higher than 12 is assigned to the high density category. The assigned range ratio for this category, 0.8, is chosen based on the optimum weight in Fig. 4.1.

4.3 Non-Uniform Network Simulations

Subsequently, simulations are conducted to compare localization accuracy between DV-Hop and DHL in a non-uniform network. The objective is to examine whether the introduction of density-awareness can improve the accuracy of hop-count localization in non-uniform networks.

We observe that the degree of non-uniformity of a network can be affected by three factors, i.e., (a) the number of regions with different local density from their surrounding regions, (b) the local density of each of these regions, and (c) the area of each of these regions. For example, suppose that a network has k number of regions which have different average local density from their surrounding regions, where the corresponding local density and the areas are $L=\{L_1, L_2, \ldots, L_k\}$, and $A=\{A_1,A_2, \ldots, A_k\}$, respectively. The degree of the network non-uniformity increases if k increases, standard deviation of L increases, or standard deviation of A decreases. In other words, degree of non-uniformity increases if the number of regions with different connectivity increases, the difference in connectivity becomes wider, or the area among the regions becomes more equal. If two areas with different connectivity have very dissimilar size, i.e., one is

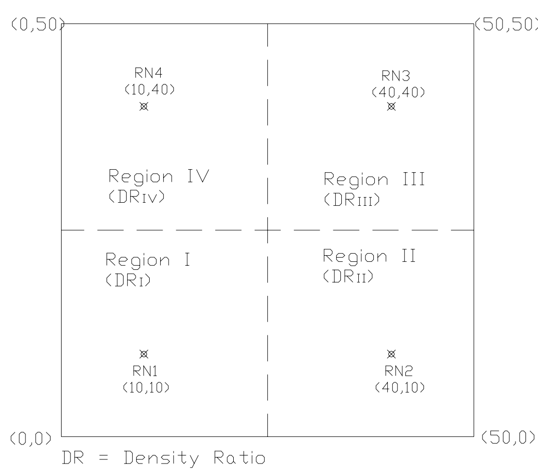


Fig. 4.2 Simulation setting for overheads comparison.

infinitely larger than the other; the non-uniformity caused by the small area becomes insignificant.

In our simulations, the number of regions with different local density from their surrounding regions is selected as 4, i.e., Region I to Region IV, and the area of each the region is equal, i.e., A_1=A_2=A_3=A_4 (Fig. 4.2). The network size is $50 \times 50\text{m}^2$ and the transmission range is 5m. A total of 10 reference nodes is placed randomly in the network.

To create non-uniformity in local density, a total of 500 nodes are deployed randomly with Density Ratio (DR) for the four regions, DR_I:DR_{II}:DR_{III}:DR_{IV}, 3:1:3:1. In such a deployment, Region I and Region III have three times more nodes than Region II and Region IV. For DHL, range ratios are assigned according to Table 4.1, i.e., $(W_l, W_m, W_h) = (0.6, 0.7, 0.8)$.

4.3.1 Distance Accuracy with *Density-awareness*

We evaluated Phase 1 of DHL, i.e., updating hop-counts with range ratio based on a node's local density. The simulation scenario is tested with 50 trials. Distance error, δ_d, is computed as a ratio of the "difference between a node's estimated distance, L_e, and actual distance, L_a" to the "transmission range, R". In other words, the computed error represents the deviation of the estimated distance relative to a node's transmission range,

$$\delta_d = \frac{L_e - L_a}{R} \times 100\% . \tag{4.1}$$

Fig. 4.3 compares the accuracy in distance estimation between DV-Hop and DHL. The figure illustrates the percentage of estimated distances with errors from "shorter than actual distance by 2R" (-200%R) to "longer than actual distance by 2R" (200%R). Each bar represents 20% width, i.e., the distance error is shown from "-200% to -180%", "-180% to -160%", and so on.

The results show that more distances are estimated with less error using DHL. From Fig. 4.3, using DHL, almost 82% of estimated distances have less than 60%R error (-60% to 60%R from Fig. 4.3) whereas around 71% distances estimated from DV-Hop achieves the same. Besides, DV-Hop has more distances estimated with greater than 100%R error than DHL, i.e., 12.6% as compared to 6%.

DV-Hop can cause both distance underestimation and overestimation with almost equal probability (Fig. 4.3). This is because the hop-distance is computed as an average value from the hop-counts accumulated along paths between reference nodes. A path can pass through a few regions with different densities, thus the distance progressed with

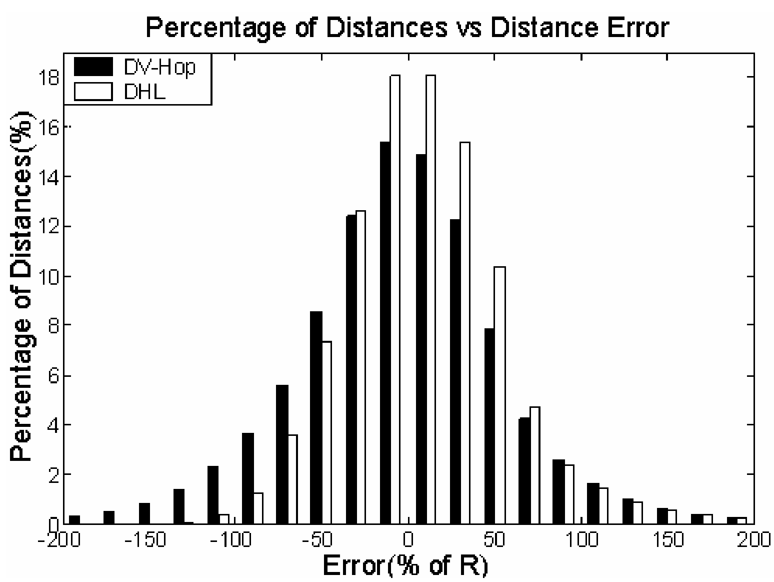


Fig. 4.3 Distance error distribution.

each hop tends to be different from one another. It results in high variance in the actual hop-distances. Using an average hop-distance value, an estimated distance could easily be shorter or longer than the actual distance. For example, in Fig. 3.3c, if the shown hop-count propagation path is the shortest from RN1 to RN2 and vice versa, using DV-Hop, Node 2 would underestimate its distance from RN1, but overestimate its distance from RN2. The same happens to Node 3 to Node 8, but with different degree of deviation. In

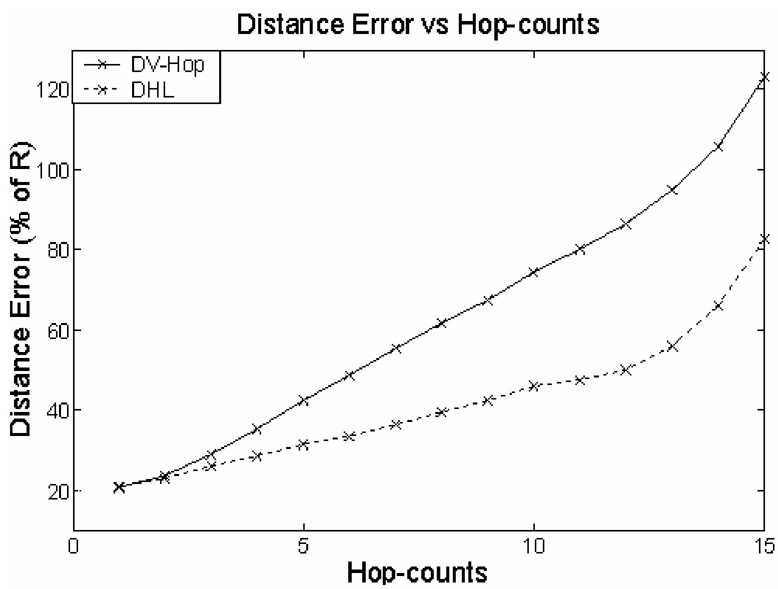


Fig. 4.4 Distance Error vs. Hop-counts.

comparison, DHL has lesser tendency to underestimate the estimated distance. Fig. 4.3 shows that DHL causes almost no estimated distance with less than -100%R errors.

Besides distance error distribution, a comparison between distance error and hop-counts is also plotted in Fig. 4.4. In this figure, absolute distance error is used. From the figure, it can be deduced that distances with larger errors are mostly associated with larger hop-counts. DHL manages to reduce distance errors when the hop-counts increase.

4.3.2 Position Accuracy with Density-awareness

Better distance estimation may not be sufficient to indicate better estimated positions. This is because in triangulation computation [31], calculating a node's position in two dimensional network requires estimated distances from at least three non-collinear RNs. One badly estimated distance can adversely affect the final estimated position. Therefore, estimated positions are also computed using lateration, which is a form of triangulation, and the position accuracy comparison between DV-Hop and DHL is shown in Fig. 4.5.

Position error, δ_p, is computed as a ratio of the "difference between a node's actual position (u, v) and estimated position ($\tilde{u}, \tilde{v}$) " to the "transmission range, R", as shown in Eqn. 4.2. In other words, the computed error represents the deviation of the estimated positions relative to a node's transmission range.

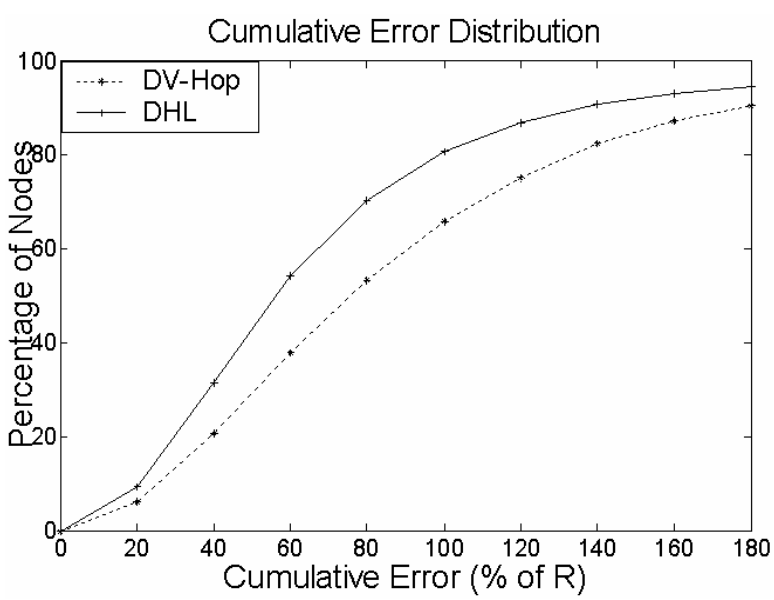


Fig. 4.5 Cumulative Error Distribution – Effect of Density awareness.

$$\delta_p = \frac{\sqrt{(u-\tilde{u})^2 + (v-\tilde{v})^2}}{R} \times 100\% \tag{4.2}$$

The results show that using DHL, 80% of the nodes managed to estimate their locations within one transmission range from the actual locations. Comparatively, using DV-Hop, only around 60% of the nodes managed to accomplish the same results. In the non-uniform network (Fig. 4.2), the average local density in Region I and Region III is approximately 23 whereas the average local density in Region II and Region IV is about 7. Thus, the difference in local density is about 3.3 ($C_I/C_{II} = C_{III}/C_{IV} = 23/7$). In other words, nodes in Region I and Region III have higher local density, i.e., around three times more neighbors, than those in Region II and Region IV. According to Table 4.1, nodes in Region I and Region III have high local density whereas nodes in Region II and IV have medium local density. As they are in different density categories, DHL treats them differently in the hop-count computation.

Nodes in Region I and Region III are expected to advance each hop with larger hop-distance, and thus, are assigned higher range ratio by DHL. Conversely, nodes in Region II and Region IV which have sparser density are assigned lower range ratio. In contrast, DV-Hop increases each hop by one and uses average hop-distance in estimating the distances of a node from RNs. By taking into account the impacts of network non-uniformity, DHL's accuracy in the distance estimation of each node from the reference nodes is higher compared to DV-Hop. Therefore, after triangulation, the estimated positions of most nodes are also closer to the actual positions.

4.3.3 Position Accuracy with Confidence Level (CL)

Fig. 4.6 shows the cumulative error distribution when *confidence level* (CL) is associated with estimated distances. In DV-Hop-CL, after a node computes its distance by computing "$HC \times D_{avg}$", it only selects distances with high confidence level for triangulation. Similarly, in DHL-CL, after a node computes its distance by computing "$\sum W \times R$", only distances with high confidence level are used in triangulation. In this simulation, if the accumulated hop-count is less than ten, it is associated with high confidence.

When estimated distances with low CL are ignored in position computation using lateration, the results (Fig. 4.6) show that DHL-CL performs better than DHL while DV-Hop-CL performs better than DV-Hop. Among these four schemes, DHL-CL achieves the highest accuracy, with the most number of nodes (83%) managing to estimate their

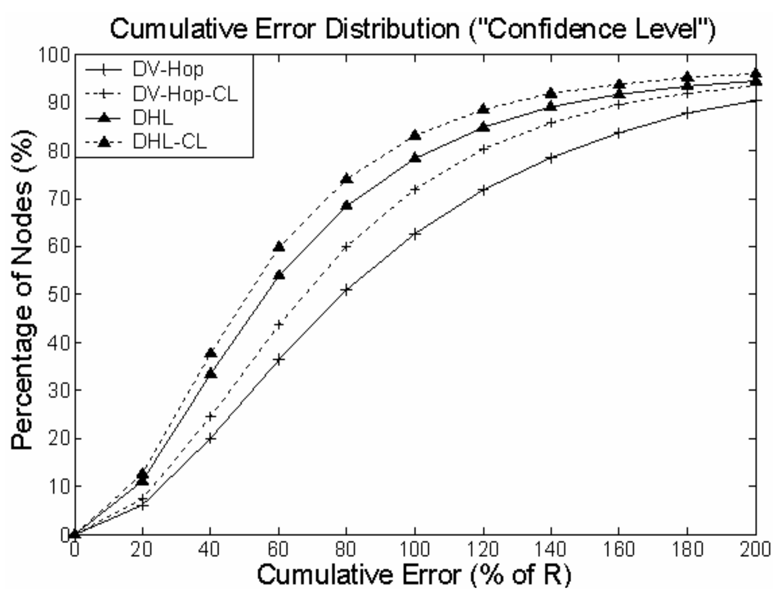


Fig. 4.6 Cumulative Error Distribution -Effect of Confidence Level (CL).

positions to the accuracy of less than one transmission range from their actual locations, compared to DHL (78%), DV-Hop-CL (72%) and DV-Hop (63%). It shows that density-awareness coupled with “Confidence Level” achieves the best results among these four schemes.

When a propagation path is long, more errors tend to be accumulated; this is more evident especially in the case when the propagation path passes through sparse regions in the network. The propagation path is less likely to be direct, straight and the shortest, thus accumulating more extra hop-counts. Besides, the actual distance per hop is either less than or equal to one transmission range. This difference is negligible if the propagation path is short; however, error accumulates and becomes significant when the propagation path is long. Thus, for a node that is further in hop-counts from a particular reference node, the corresponding estimated distances tend to have higher errors. Consequently, putting higher confidence in distance acquired from smaller hop-counts in the position computation process can help to improve localization accuracy.

4.3.4 Geographic Error Distribution

Another useful way to investigate error distribution is to take into account individual node's geographical location. Fig. 4.7 and Fig. 4.8 give detailed looks at the distribution of position error as a function of individual nodes' physical locations in the square network area of 50m $\times$ 50m.

Comparing Fig. 4.7 and Fig. 4.8, DV-Hop localization error is higher than DHL localization. The range of DV-Hop error distribution is approximately 100%*R* for most of the interior nodes whereas a small portion of nodes at edges have localization error up to approximately 300%*R*. In contrast, localization error for most of the interior nodes of DHL hovers around 50%*R* while a small percentage of nodes at edges has up to around 250%*R* error. This shows that DHL has better performance than DV-Hop for nodes

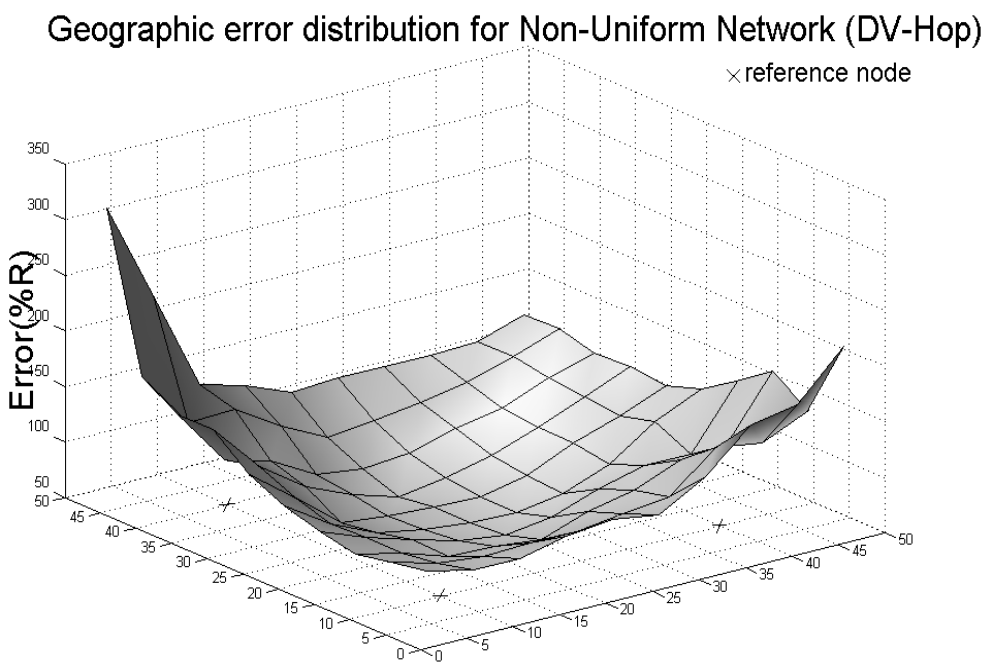


Fig. 4.7 Geographic Error Distribution - DV-Hop

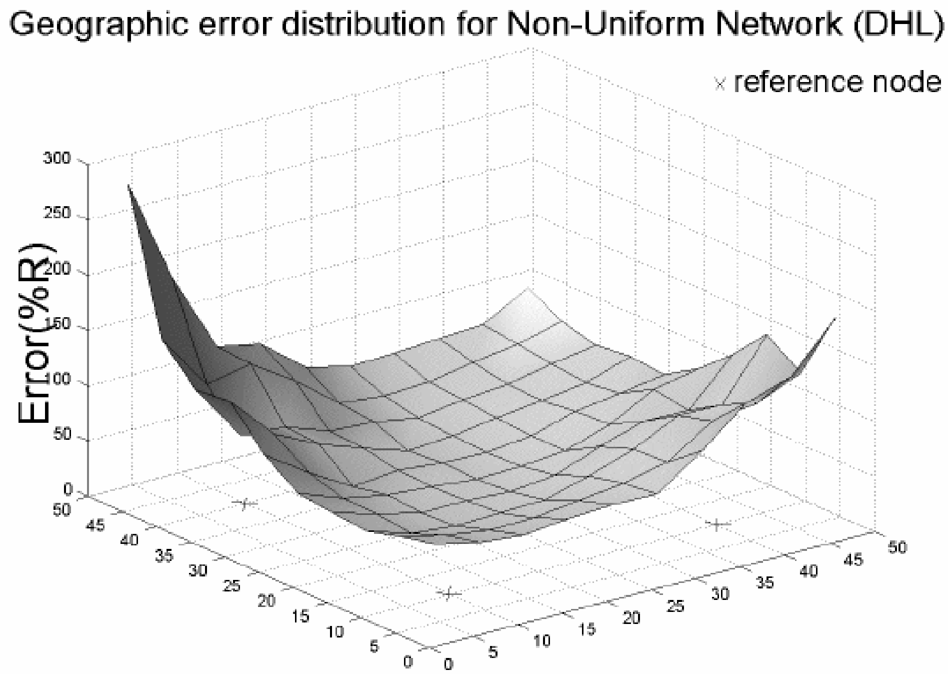


Fig. 4.8 Geographic Error Distribution - DHL.

scattered around anywhere in the network. The observation also shows that both DV-Hop and DHL shares a common phenomenon, whereby the nodes near the network edges and corners are susceptible to higher localization error compared to those located near the center of the network. We illustrate how this phenomenon can arise in Fig. 4.9 and Fig. 4.10.

Fig. 4.9 shows two propagation paths from a reference node located at a corner of the network to another node located at the network edge (Path 1) and at the network center (Path 2) respectively. The figure shows that the first path tends to follow a longer route compared to the second path. This is mainly because areas along the network boundary tend to have lower concentration of intermediate nodes such that the probability of locating a next propagating node that is close to the transmission range and in the direction of propagation is much lower compared to areas at the network center. We

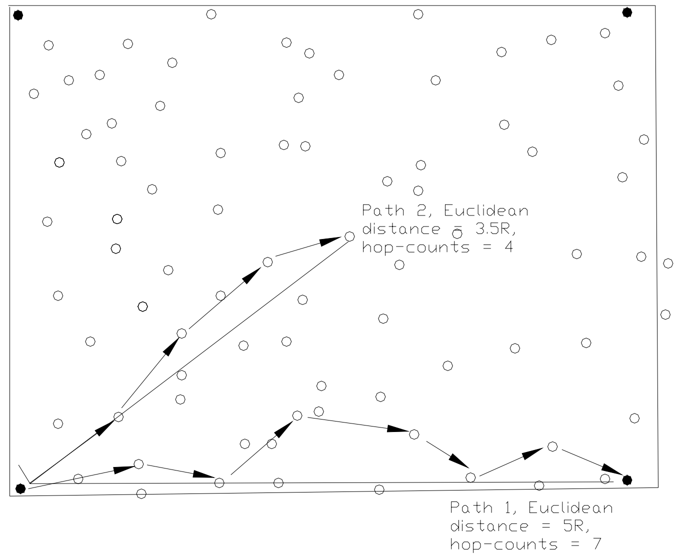


Fig. 4.9 Propagation paths along a network edge (Path 1), and towards network center (Path 2)

illustrate the possible forwarding transmission area in Fig. 4.10. In order to propagate a packet in the forward direction, a node at the network center can forward to any node located in the shaded area (Fig. 4.10a), preferably to those near the transmission range. However, for a node located along the network edge (Fig. 4.10b), the shaded area is reduced by half since no intermediate node is available outside the network region. Strategically placing reference nodes near the network edges so that most nodes at edges can have direct communication with reference nodes could be a good future study topic to reduce the impact of such phenomenon.

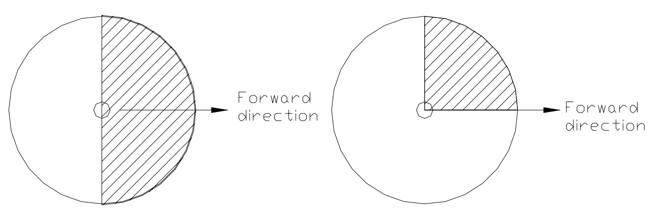


Fig. 4.10 Forward propagation area for (a) a node at network center, (b) a node at network edge

4.4 Random Network Simulations

The performance of DV-Hop and DHL are also compared in random networks. In random network scenario, nodes are positioned randomly throughout the network. In this case, the nodes are scattered quite uniformly where each node has approximately the same number of neighbors. The network does not have any particular regions with higher or lower node density. The total number of nodes being scattered in the network is increased from 500 to 700. The network size is $50 \times 50\text{m}^2$ and the transmission range is 5m. A total of 10 reference nodes are placed randomly in the network. From the simulation results (Fig. 4.11), we found that both schemes manage to locate large percentage of nodes to high accuracy and the accuracy achieved by both schemes is quite comparable. This is because in random networks where nodes are distributed uniformly,

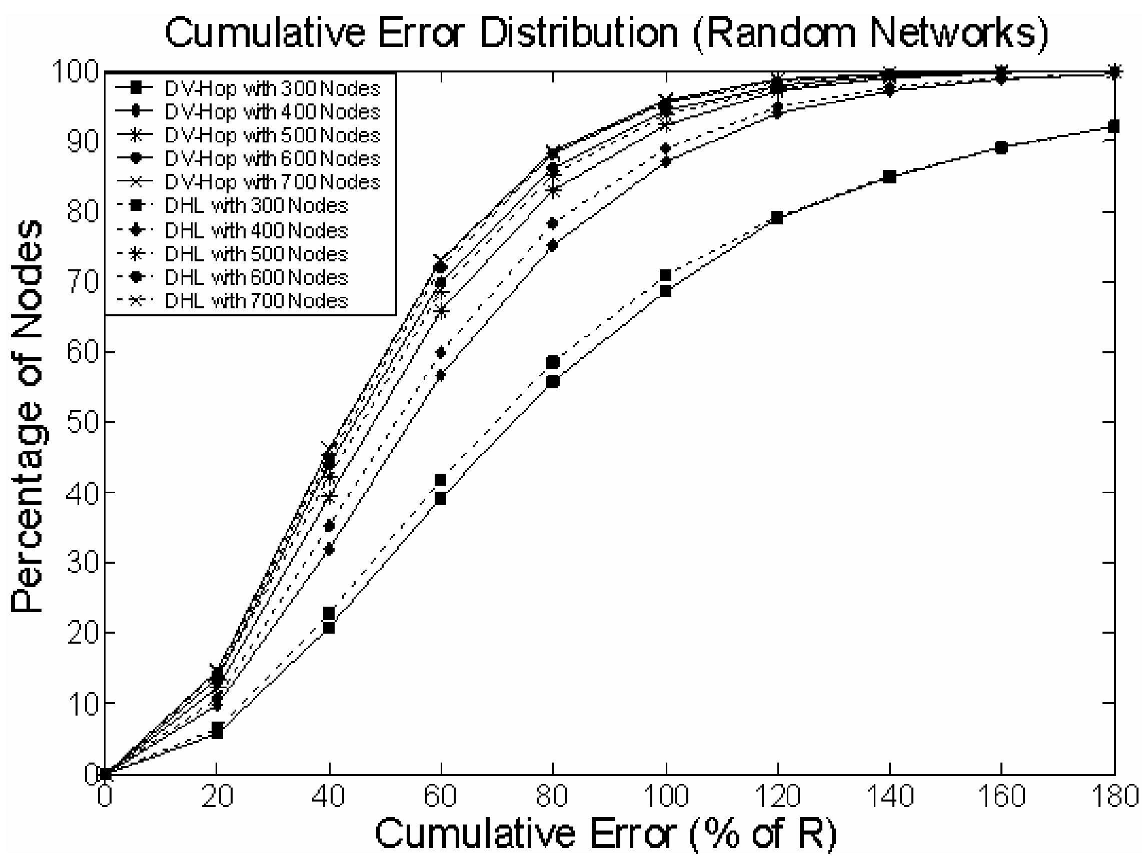


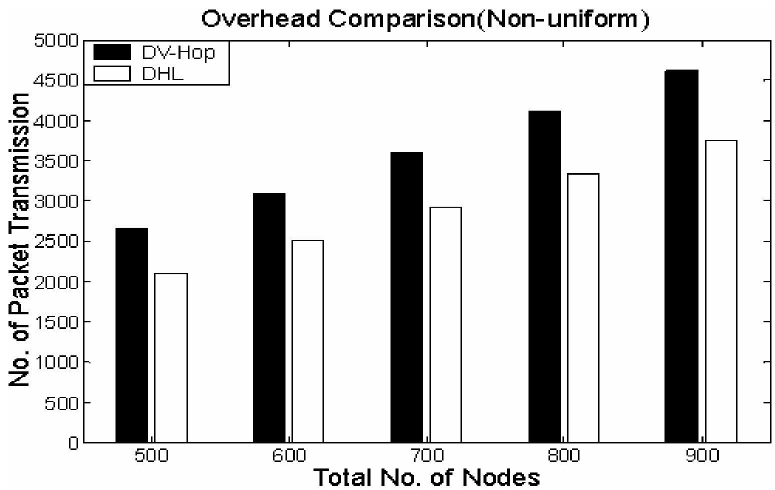
Fig. 4.11 Cumulative Error Distribution – Random Networks.

average hop-distance computed by DV-Hop shows good approximation to the actual hop-distance. Besides, DHL is also capable of achieving comparable results with the use of range ratio.

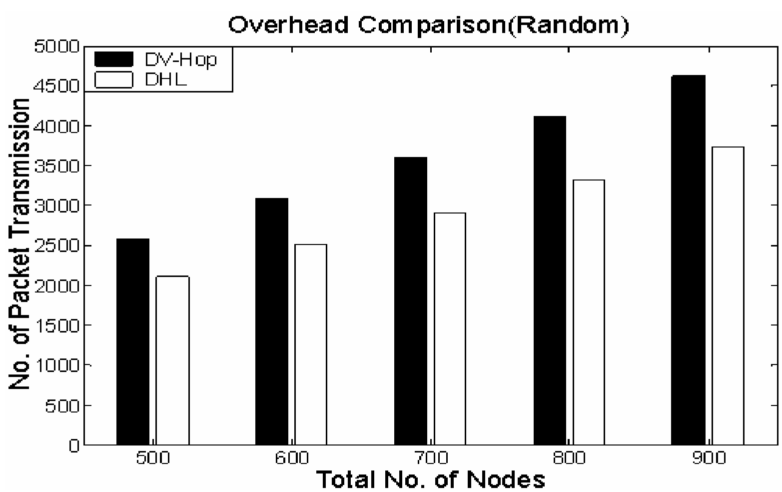
4.5 Overhead Comparisons

Packet transmission overheads for both DV-Hop and DHL are compared in non-uniform and random networks. The total number of nodes in the network is increased from 500 to 900 to investigate how packet transmission overheads change with the increase of total nodes. In non-uniform network setting, four reference nodes are placed as shown in Fig. 4.2. The reference nodes are close to the network boundary and surrounded by randomly placed nodes in all directions. Thus, the area of transmission is circular and the density surrounding a reference node is affected mainly by its connectivity. In the random network setting, nodes are randomly scattered throughout the network. The overhead comparison results for this non-uniform networks are shown in Fig. 4.12(a) while the results for random networks are shown in Fig. 4.12(b).

The reason DV-Hop incurs higher number of packet transmissions is due to an additional Correction flooding stage. The scheme floods the network twice. The first flooding involves accumulating hop-counts and the second flooding involves spreading computed D_{avg}, *average distance per hop-count*. In comparison, DHL integrates the correction with the hop-count accumulation stage. Thus, it eliminates any additional flooding stage. This effectively reduces the time needed for a node to compute its locations, and thus reduces the response time for location-related queries. Although DHL involves more frequent hop-count adjustment in the hop-count accumulation stage, the total number of transmission is still less than DV-Hop as DHL uses only one flooding stage. Since most sensors have limited power supply, energy efficiency is an important factor in algorithm design. By maintaining lower packet transmission overheads, DHL helps to reduce power consumption, and thus achieves better energy efficiency.



(a)



(b)

Fig. 4.12 Overhead comparison for (a) non-uniform network, and (b) random networks.

4.6 Discussion of DHL Issues

4.6.1 Local Density Representation

The current representation of local density is based on a node's connectivity, or the number of neighboring nodes. However, this representation may not be appropriate for nodes that do not have circular transmission coverage, e.g., nodes that are located near the network boundary or nodes that use directional antenna. For these nodes, their neighboring nodes are not randomly placed in all directions surrounding them, but located at particular angles. Thus, even though a node has a large number of neighboring nodes, these neighbors are not helpful in forwarding a packet to particular directions. Therefore, the proportional relationship between local density and hop-distance is no longer true. Some alternative theoretical methods in defining local density are needed for nodes without circular coverage. The definition of local density should take into consideration the area and the angle of transmission coverage.

4.6.2 Range Ratio Assignment

The current values of range ratio are selected based on experimentation results. Kleinrock and Silvester [23] have independently conducted theoretical analysis on optimum connectivity for wireless networks. Part of their analysis is related to finding the effective distance traversed per hop for multi-hop wireless networks. Using their analysis, a node can compute its hop-distance on-the-fly based on its local density. However, their analysis is based on the assumptions of Poisson node distribution and short distance between source and destination nodes. This may not be true in all network scenarios. The analysis from Xue and Kumar [42] contrasts with the studies by Kleinrock and Silverster which recommended some "*magic numbers*" of nearest neighbors to maintain network-wide connectivity. Instead, Xue and Kumar show that in a network with n randomly placed nodes, each node should be connected to $\Theta(\log n)$ nearest neighbors in order to avoid network partitioning. In this scenario, the number of neighbors a node maintains could vary with time depending on how frequently the total number of nodes in the network changes. Therefore, further studies can be conducted to determine the number of links a node is connected to at a particular time. Besides, methods to assign range ratio when the connectivity of a node varies with time should also be studied.

If nodes are not classified into density categories but they are allowed to compute its own hop-distance based on individual local density (i.e., unlimited density categories), the total transmission overhead will be substantial. The frequency of hop-counts re-adjustment will be high since a node tends to receive a new minimum hop-count from time to time and subsequently triggers another round of broadcasting.

As the number of density categories increases, the range ratio that a node computes has high chances to be different from that computed by its neighbors. For example, in the case when there are only two density categories, a node has fifty percent chances that its range ratio is different from its neighbors. When the number of density categories increase to ten, the probability increases to ninety percent. Thus, the accumulated hop-counts between nodes tend to be different from each other. In any case when hop-counts are different for two neighboring nodes, the node that has higher hop-counts may need to re-compute its hop-counts and retransmit. Thus, the frequency of hop-count adjustment and message exchanges is high.

4.6.3 Node Mobility

The current experimentations and simulations are conducted for static nodes. This is because the nodes in the target network, i.e., wireless sensor network, are commonly associated with low mobility. In mobile networks, modifications or enhancement can be added into the algorithm. A mobile node can obtain hop-counts from its new neighbors to compute triangulation. Alternatively, a node can obtain the estimated positions from its new neighbors and compute an average value. In this way, a mobile node is able to compute new positions with minimum communication signaling.

If the reference nodes are mobile, they can assist in localization refinement. This is because their positions can act as new reference points to the nodes in close proximity. Thus, after triangulation, a node usually is able to estimate its position with better approximation.

4.7 Conclusion

In this chapter, experimental results are presented and discussed. Firstly, the impacts of the two issues, i.e., *sparse nodes issue* and *long path issue*, are investigated. Then, range ratios for DHL are determined, followed by accuracy comparison between DHL and DV-Hop in non-uniform networks. Communication overheads are also evaluated. The results show that DHL achieves better distance and position estimation in non-uniform networks, with less transmission overheads. In the next chapter, a brief summary of our work is described and conclusion is given.

Chapter 5 Conclusion and Future Works

5.1 Conclusion

In this thesis, we described a self-configuring localization algorithm, Density-aware Hop-count Localization (DHL). The design motivation is to address two issues: (a) *sparse nodes issue*, where localization accuracy drops at low local density; and (b) *long path issue*, where distance error accumulates with hop-counts. To address the non-uniform node distribution issue, a novel concept of density-based hop-count update is developed. We identify density as an important parameter in characterizing hop-distance, thus, we proposed an algorithm for self-localization based on node density. We also evaluated and demonstrated the effectiveness of our solutions.

Our design is driven by a major goal, i.e., to improve localization accuracy in sparse and non-uniform networks. DHL makes use of the multi-hop feature of ad hoc sensor networks to estimate distances with respect to some known location nodes. Propagated hop-count is incremented with range ratio, which is the ratio of progressed distance with respect to transmission range. A node that obtains distances from more than three reference nodes only select distances computed from small hop-counts in triangulation. These distances are associated with high confidence level since error tends to increase with hop-counts.

Simulations showed that when a network has non-uniform node distribution, the introduction of density-awareness is able to improve DV-Hop localization accuracy while incurring lower packet transmission overheads. The confidence associated with estimated distances improved the accuracy further in non-uniform networks. In random networks that have rather uniform distribution, DHL managed to achieve comparable accuracy as DV-Hop while maintaining lower packet transmission overheads.

5.2 Future Works

Based on the assignment of three density categories and the corresponding range ratios, we are able to achieve better location estimation accuracy while maintaining lower overheads compared to conventional schemes in non-uniform networks. As the achieved improvement may not be optimum in all cases, a possible extension to DHL is to analyze the impact of range ratios on other network settings, for example by varying the degree of network non-uniformity. Analysis can be conducted to explore the effect of the number of density categories on localization accuracy and transmission overheads. Besides, other than local density, factors such as propagation direction, which can affect hop-distance, can also be explored to enhance localization accuracy.

DHL issues that have been discussed in the previous chapter, i.e., local density representation, range ratio representation and node mobility can be explored further to improve the algorithm. Analysis can be performed to define local density for nodes that do not have circular coverage, for example for nodes that are located near the network edges or nodes that use directional antenna. Further theoretical and experimental studies can be conducted to map the relationship between range ratio and local density. If the local density for a node varies with time, the range ratio should also be adjusted when local density changes. The current algorithm is suitable for sensor networks that have static or low mobility nodes, but further studies need to be done for networks that comprise of highly mobile nodes in which the network density changes rapidly.

We can also look into placement strategies of reference nodes. From the simulations showing geographical error distribution in the previous chapter, nodes close to network edges and corners tend to have higher location estimation error due to low concentration of reference nodes. This issue can be tackled by strategically placing the reference nodes such that the nodes are able to have unhindered communication paths with the reference nodes.

In conclusion, a novel density-aware and path length-aware localization algorithm, i.e., DHL, has been presented for unevenly distributed sensor networks that potentially have long propagation paths. Further studies can be performed to enhance the algorithm so that it may be applied in different network scenarios.

Bibliography

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey", *Elsevier Computer Networks*, Vol. 38, No. 4, pp. 393-422, March, 2002.

[2] P. Bahl, and V.N. Padmanabhan, "RADAR: An In-building RF-based User Localization for Very Small devices", *Infocom 2000*, Tel Aviv, Israel, pp. 775-784, March, 2000.

[3] W. Brett, M. Last, B. Leibowitz, and K. Pister, "Smart Dust: Communicating with a Cubic-Millimeter Computer", *IEEE Computer*, Vol. 34, No. 1, pp. 45-51, January, 2001.

[4] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less Low-cost Outdoor Localization for Very Small Devices", *IEEE Personal Communications*, Vol. 7, No. 5, pp. 28-34, October, 2000.

[5] N. Bulusu, J. Heidemann, D. Estrin, and T. Tan, "Self-Configuring Localization Systems: Design and Experimental Evaluation", *ACM Transactions on Embedded Computing Systems*, Vol. 3, No. 1, pp. 24-60, February, 2004.

[6] S. Capkun, M. Hamdi, and J.-P. Hubaux, "GPS-free Positioning in Mobile Ad-hoc Networks", *Hawaii International Conference on System Sciences*, Hawaii, USA, pp. 3481-3490, January, 2001.

[7] A. Cerpa, J. Elson, M. Hamilton and J. Zhao, "Habitat Monitoring: Application Driver for Wireless Communications Technology", *SIGCOMM 2001*, San Jose, Costa Rica, pp. 26-41, April, 2001.

[8] S. Cho and A. Chandrakasan, "Energy-efficient Protocols for Low Duty Cycle Wireless Microsensor", *ICASSP 2001*, Utah, USA, pp. 2041-2044, May, 2001.

[9] C.-Y. Chong, and S.P. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges", *Proceedings of the IEEE*, Vol. 91, No. 8, pp. 1247-1256, August, 2003.

[10] L. Doherty, K.Pister, and Ghaoui, "Convex Position Estimation in Wireless Sensor Networks", *Infocom 2001*, Alaska, USA, pp.1655-1663, April, 2001.

[11] T. Eren, D. Goldenberg, W. Whiteley, Y. R. Yang, A. S. Morse, B. D. O. Anderson, and P. N. Belhumeur, "Rigidity, Computation, and Randomization in Network Localization", *Infocom 2004*, Hong Kong, pp.2673-2684, March, 2004.

[12] D. Estrin, D. Culler, and Kris Pister, "Connecting the Physical World with Pervasive Networks", *IEEE Pervasive Computing*, Vol. 1, No. 1, pp. 59-69, January, 2002.

[13] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the World with Wireless Sensor Networks", *ICASSP 2001,* Utah, USA, pp. 2033-2036, May, 2001.

[14] D. Ganesan, R. Govindan, S. Shenker and D. Estrin, "Highly Resilient, Energy Efficient Multipath Routing in Wireless Sensor Networks", *Mobile Computing and Communications Review*, Vol. 1, No. 2, pp. 1356-1361, April, 2002.

[15] L. Girod, and D. Estrin, "Robust Range Estimation using Acoustic and Multimodal Sensing", *IEEE/RSJ International Conference on Intelligent Robots and Systems,* Hawaii, USA, pp. 1312-1320, October, 2001.

[16] G. Golub, and C. F. Van Loan, "Matrix Computation", Third Edition, *The Johns Hopkins University Press*, Baltimore, USA, 1996.

[17] T. He, C. Huang, B.M. Blum, J.A. stankovic, and T. Abdelzaher, "Range-free Localization Schemes for Large Scale Sensor Networks", *Mobicom 2003*, San Diego, USA, pp. 348-352, September, 2003.

[18] J. Hightower, R. Want, and G. Borriello, "Location Systems for Ubiquitous Computing", *IEEE Computer*, Vol. 34, No. 8, pp. 57-66, August, 2001.

[19] C. Intanagonwiwat, R. Govindan and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks", *MobiCom 2000*, Massachusetts, USA, pp. 56–67, August, 2000.

[20] C. Intanagonwiwat, D. Estrin, R. Govindan and J Heidemann, "Impact of Network Density on Data Aggregation in Wireless Sensor Networks", *ICDCS 2002,* Vienna, Austria, pp. 67-69, July, 2002.

[21] B. Karp, and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks", *Mobicom 2000*, Massachusetts, USA, pp. 243-254, August, 2000.

[22] Y.-B. Ko, and N. H. Vaidya, "Location-aided Routing (LAR) in Mobile Ad Hoc Networks", *Mobicom 1998*, Texas, USA, pp. 87-96, October, 1998.

[23] L. Kleinrock and J.A. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number*", IEEE National Telecommunication Conference*, Alabama, USA, pp. 4.3.1-4.3.5, November, 1978.

[24] K. Langendoen and N. Reijers, "Distributed Localization in Wireless Sensor Networks: A Quantitative Comparison", *Elsevier Computer Networks*, Vol. 43, No. 3, pp. 499-518, November, 2003.

[25] J. G. Lim and SV Rao, “Mobility-enhanced Positioning in Ad Hoc Networks”, *WCNC 2003*, Louisiana, USA, pp. 1832-1837, March, 2003.

[26] R. Nagpal, H. Shrobe, and J. Bachrach, “Organizing a Global Coordinate System from Local Information on an Ad Hoc Sensor Network”, *IPSN 2003*, California, USA, pp. 333-348, April, 2003.

[27] D. Niculescu, and B. Nath, "Ad Hoc Positioning System (APS)", *Globecom 2001*, Texas, USA, pp.2826-2931, November, 2001.

[28] D. Niculscu, and B. Nath, “Ad Hoc Positioning System (APS) using AOA”, *Infocom 2003*, California, USA, pp. 1734-1743, March, 2003.

[29] D. Niculescu, and B. Nath, “DV-based Positioning in Ad Hoc Networks”, *Kluwer Journal of Telecommunication Systems*, Vol. 22, No. 1, pp. 267-280, January, 2003.

[30] D. Niculescu, and B. Nath, “Localized Positioning in Ad Hoc Networks”, *SNPA 2003, Anchorage*, Alaska, USA, pp.42-50, April, 2003.

[31] B. Parkinson and J. Spilker, “Global Positioning System: Theory and Application”, Washington, D.C., American Institute of Aeronautics and Astronautics, 1996.

[32] P. Pathirana, N. Bulusu, S. Jha, and A. Savkin, “Node Localization using Mobile Robots in Delay-tolerant Sensor Networks”, *IEEE Transaction on Mobile Computing*, Vol. 4, No. 3, pp. 1312-1328, July, 2005.

[33] G. Pottie, and W. Kaiser, “Wireless Integrated Network Sensors”, *Communications of the ACM*, Vol. 43, No. 5, pp. 51-58, May, 2000.

[34] C. Savarese, K. Langendoen, and J. Rabaey, “Robust Positioning Algorithms for Distributed Ad-Hoc Wireless Sensor Networks”, *USENIX Technical Annual Conference*, California, USA, pp. 317-327, June, 2002.

[35] A. Savvides, C.-C. Han, and M. Srivastava, “Dynamic Fine-Grained Localization in Ad-Hoc Networks of Sensors”, *Mobicom 2001*, Rome, Italy, pp. 166-179, July, 2001.

[36] A. Savvides, H. Park and M. B. Srivastava, “The *N*-Hop Multilateration Primitive for Node Localization Problems”, *ACM/Kluwer Mobile Networks and Applications (MONET)*, Vol. 8, No. 4, pp.443-451, August, 2003.

[37] M. L. Sichitiu, and V. Ramadurai, “Localization of Wireless Sensor Networks with a Mobile Beacon”, *the 1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, Florida, USA, pp. 174-183, October, 2004.

[38] Y.-C. Tseng, S.-L Wu, W.-H. Liao, and C.-M. Chao, "Location Awareness in Ad Hoc Wireless Mobile Networks", *IEEE Computer*, Vol. 34, No. 6, pp. 46-52, June, 2001.

[39] S. Y. Wong, J. G. Lim, SV Rao, and Winston KG Seah, "Density-Aware Hop-Count Localization (DHL) in Wireless Sensor Networks with Variable Density", *WCNC 2005*, Louisiana, USA, pp. 1848-1853, March, 2005.

[40] S. Y. Wong, J. G. Lim, SV Rao, and Winston KG Seah, "Multihop Localization with Density and Path Length Awareness in Non-Uniform Wireless Sensor Networks", *VTC 2005 spring*, Stockholm, Sweden, May, 2005.

[41] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed Energy Conservation for Ad Hoc Routing", *MobiCom 2001*, Rome, Italy, pp. 74-84, July, 2001.

[42] F. Xue, and P. R. Kumar, "The Number of Neighbors Needed for Connectivity of Wireless Networks", *Kluwer Wireless Networks*, Vol. 10, No. 2, pp. 169-181, March, 2004.