

## A Novel Multi-Hop Algorithm for Wireless Network with Unevenly Distributed Nodes

Yu Liu<sup>1</sup>, Zhong Yang<sup>2</sup>, Xiaoyong Yan<sup>3</sup>, Guangchi Liu<sup>4</sup> and Bo Hu<sup>5,\*</sup>

**Abstract:** Node location estimation is not only the promise of the wireless network for target recognition, monitoring, tracking and many other applications, but also one of the hot topics in wireless network research. In this paper, the localization algorithm for wireless network with unevenly distributed nodes is discussed, and a novel multi-hop localization algorithm based on Elastic Net is proposed. The proposed approach is formulated as a regression problem, which is solved by Elastic Net. Unlike other previous localization approaches, the proposed approach overcomes the shortcomings of traditional approaches assume that nodes are distributed in regular areas without holes or obstacles, therefore has a strong adaptability to the complex deployment environment. The proposed approach consists of three steps: the data collection step, mapping model building step, and location estimation step. In the data collection step, training information among anchor nodes of the given network is collected. In mapping model building step, the mapping model among the hop-counts and the Euclidean distances between anchor nodes is constructed using Elastic Net. In location estimation step, each normal node finds its exact location in a distributed manner. Realistic scenario experiments and simulation experiments do exhibit the excellent and robust location estimation performance.

**Keywords:** Multi-hop localization, Elastic Net, regularization, sparse.

### 1 Introduction

As a new media, wireless network can realize communication between people and object, and it has been broadly used. However, the location of people or object in the network is the basic information for communication. Current wireless localization technologies include GPS-based [Wang, Lu, Wang et al. (2017)], infrared-based [Park, Cho, Kim et al. (2014)], Bluetooth-based [Castillo-Cara, Lovón-Melgarejo, Bravo-Rocca et al. (2017)],

---

<sup>1</sup> School of Computer Engineering, Jinling Institute of Technology, Nanjing, 211169, China.

<sup>2</sup> School of Intelligence Science and Control Engineering, Jinling Institute of Technology, Nanjing, 211169, China.

<sup>3</sup> School of Modern Post & Institute of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China.

<sup>4</sup> Stratifyd Inc., Charlotte, NC 28208, USA.

<sup>5</sup> Center of Information Construction and Management, Nanjing Normal University of Special Education, Nanjing, 210038, China.

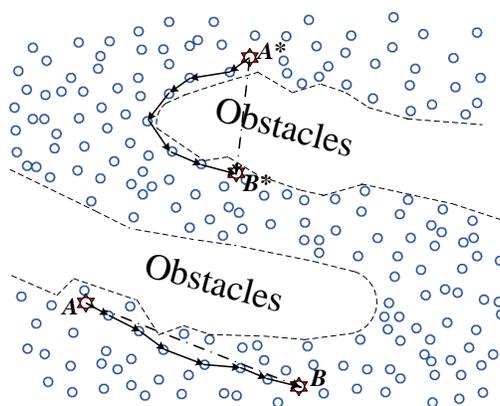
\* Corresponding Author: Bo Hu. Email: nj802006@163.com.

RFID-based [Zhou and Shi (2009)], WIFI-based and UWB-based [Mazhar, Khan and Sällberg (2017)], etc. Among them, GPS-based application is the most mature localization application with broadest application, but when there are obstacles, it is very difficult to satisfy the application requirement on the aspect of localization precision; the infrared ray has short propagation distance, which also tends to be interfered with by fluorescent lamp or indoor lighting, and there is certain limitation during localization; both the Bluetooth technology and RFID technology apply to indoor localization, but the Bluetooth technology has poor stability in complicated environment and system, while the RFID technology has short transmission distance, and it does not have the communication capacity; during the WIFI localization, the transmitter and receiver can only cover an area within 90 m, but it tends to be interfered with by other signals, and its locator involves high energy consumption; the UWB localization method can provide centimeter-level localization precision, but it is difficult to capture the UWB signal, and it also tends to cause significant interference with other signals. With the introduction of new wireless networks such as IoT (Internet of Things) and Mesh, wireless nodes can be randomly deployed in various types of monitoring environment, and the location information can be exchanged through the methods of multi-hop and self-organization.

The node localization method based on multi-hop and self-organization is called the “multi-hop localization” method. The multi-hop localization method can be roughly divided into two groups of range-based and range-free methods according to whether it requires range finding during the localization process [Yang, Wu and Liu (2014)]. The localization precision of range-based method depends on the measuring equipment, so it requires high equipment cost, and it is restricted to large-scale application. The range-free localization method is a localization technique of cost-effective and less limited for a wider range of applications, and it does not require direct measuring between nodes during the localization process, so this method has attracted more attention. The range-free technique generally employs hops to describe the Euclidean distance between nodes, and the locations of unknown nodes are estimated with the assistance of trilateration. By replacing the multi-hop localization of directly measuring with hops, it tends to be affected by node distribution. Under high density and even distribution of nodes, the range-free method can provide excellent localization performance; however, when the nodes have sparse and uneven distribution, it has very poor localization performance. Unfortunately, in a real environment, due to various reasons such as random distribution of nodes, obstacles and communication failure of nodes, after the network topology is determined, the whole network would present uneven distribution under the influence of its own or external factors. Due to uneven distribution of nodes in the network, it would generate severe deviation by using hops to describe the Euclidean distance with range-free localization method [Liu, Zhang and Bu (2016)].

Take Fig. 1 as an example, we use dotted line  $d_{AB}$  ( $d_{A^*B^*}$ ) to represent the Euclidean distance between nodes  $A(A^*)$  and  $B(B^*)$  in the network, and use arrowed straight line  $h_{AB}$  ( $h_{A^*B^*}$ ) to indicate the approximate length of the shortest path between nodes  $A(A^*)$  and  $B(B^*)$ . When the nodes are distributed evenly and densely in the network, we can see that  $d_{AB} \propto h_{AB}$ , however, when nodes are unevenly distributed or obstacles affect the

signal transmission, the nodes are distributed unevenly, which will result in  $d_{A^*B^*} \ll h_{A^*B^*}$ . Therefore, we can intuitively understand that if the nodes in the network are evenly distributed, the Euclidean distance between pairs of nodes is proportional to the hop count; conversely, the proportional relationship between the Euclidean distance and the hop count no longer holds.



**Figure 1:** Nodes uneven distribution network

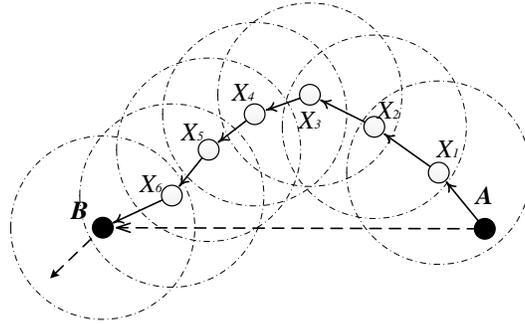
According to the above analysis, it is not difficult to see that when there are phenomena such as lack of distribution and uneven distribution of nodes, the hops between nodes will change, which will cause severe error during the estimation process of the shortest path. This kind of estimation error of the shortest path is directly related to the orientation of the shortest path, and the shortest path which does not pass through the area with uneven node distribution will not be affected. The distance estimation of shortest path between nodes is the basis of multi-hop localization problem, and only when the distance estimation error of shortest path between nodes is small can the localization of unknown node be conducted in a relatively accurate way; when the distance estimation error of shortest path between nodes is big, the localization precision of unknown node will be severely affected. In order to increase the adaptive capacity of range-free multi-hop localization method for various types of environment, we propose a range-free localization algorithm with high localization precision, which can adapt to various types of anisotropic network environment, i.e. Multi-hop Localization through Elastic Net (ML-EN). The Elastic Net method [Zou and Hastie (2015)] can be used to effectively handle strongly correlated variables, effectively choose variables and estimate related parameters, so the ML-EN method can provide higher localization precision and adapt to different network environment.

## 2 Relation work

DV-Hop is the best-known range-free multi-hop localization algorithm, which was proposed by Niculescu et al. [Niculescu and Nath (2003)] from Rutgers University, and this algorithm has been broadly discussed and quoted. The DV-Hop algorithm is built on some simple idea and easy to realize. In this algorithm, the distance from the unknown

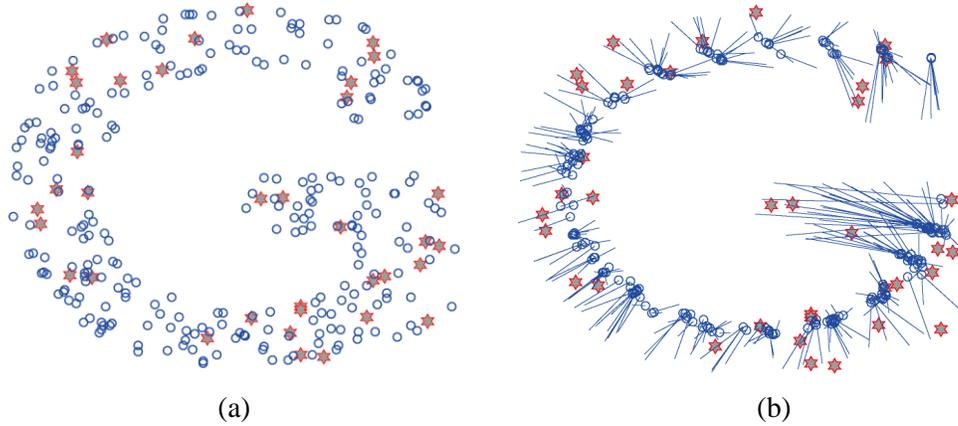
nodes to the anchor nodes are represented by the product of the per hop distance and the hops, and finally, the trilateration is used to estimate location of unknown nodes. The distance estimation method of DV-Hop algorithm has strict requirements on the distribution of nodes, which results in poor adaptability. According to the problem of uneven distribution of nodes, Shang et al. [Shang, Shi and Ahmed (2004)] proposed a method to use only four nearest anchor nodes from the unknown nodes to locate the unknown node (for convenience of discussion, we call this algorithm the “Nearest-4 algorithm”). The Nearest-4 algorithm believes that the fewer hops between nodes, the less affected by the uneven distribution of nodes. However, some unknown nodes may have few anchor nodes at close range and have to choose anchor nodes at a longer distance. In addition, the mechanism of selecting the most nearest four anchor nodes as reference nodes will also exclude some anchor nodes that are far away but have better distance estimation, thereby reducing the localization accuracy of unknown nodes. Shahzad et al. [Shahzad, Sheltami and Shakshuki (2017)] proposed the DV-maxHop method based on the Multi-objective Optimization method. The algorithm found that the value of MaxHop in the range of 5 to 15, gives satisfactory results. Recently, Zhao et al. [Zhao, Su and Shao (2018)] proposed the LWLR-DV-hop method based on the Nearest-4 algorithm. In this algorithm, three anchor nodes nearest to the unknown node are chosen as the reference nodes, and the local weighted least square method is used to further weaken the deviation between hops and Euclidean distance conversion.

In recent years, various improved DV-hop algorithms have been continuously proposed, but these improved methods still use the average hop distance from the unknown node to the anchor node to estimate the distance, and this kind of distance estimation method has the defects of poor model interpretability and poor adaptation to the nodes distribution. According to the defects of the DV-hop algorithm and its improved methods, the Hou research team at UIUC proposed the PDM algorithm [Lim and Hou (2009)]. The PDM algorithm directly adopts the skeleton between reference nodes to build the mapping model between hops and Euclidean distance, and the least square method is used to obtain the optimal linear conversion between hops and Euclidean distance. In essence, PDM is used to estimate the distance between nodes based on the hop-by-hop transmission characteristic of wireless network data, so there tends to be correlation between the hop matrix vectors during the estimation process. Take Fig. 2 as an example assume  $A$  is the source node and  $B$  is the destination node, and because it needs to pass through nodes  $X_1, X_2, X_3, X_4, X_5, X_6$  from  $A$  to  $B$ , there is also close correlation between the hop matrix vectors between nodes. When there is correlation between the hop matrix vectors, it needs to be penalized during the estimation process (i.e. regularization). Based on this problem, in the PDM algorithm, the TSVD regularization method [Hansen (1987)] is adopted during construction of the mapping model between hops and Euclidean distance.



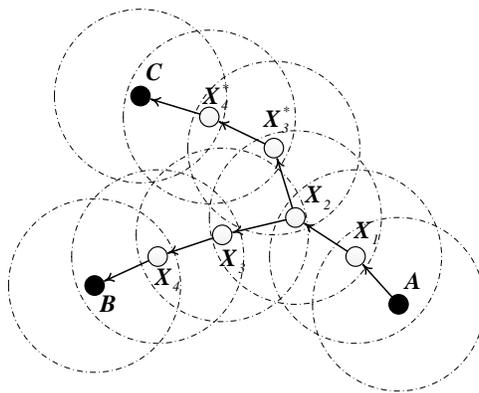
**Figure 2:** Node communication model in multi-hop networks

Compared with many previous DV-hop and its improved algorithm, the PDM method can more accurately capture the relation between hops and Euclidean distance, and effectively dig the node distribution and correlation information hidden behind data, so it can effectively improve the multi-hop localization performance. However, the PDM method has ignored the magnitude order conversion problem between hops and Euclidean distance, so it tends to cause fluctuation in localization performance of algorithm under different node densities and node communication radiuses. In addition, the regularization method adopted by the PDM method is essentially a method with empirical risk minimization, so the final localization performance of the PDM depends on the distribution of anchor nodes in the network. When there are a small number of anchor nodes, this method can only guarantee the localization precision of nodes near the anchor nodes. Inspired by the PDM method, Lee et al. [Lee, Chung and Kim (2013)] proposed two range-free multi-hop localization algorithms based on support vector regression, which are LSVR and MSVR respectively. By introducing the kernel function, these two algorithms adopt nonlinear mapping method to transform the localization problem to kernel regression problem. The support vector regression is a method with minimum structural risk, so the LSVR and MSVR methods are still able to provide great localization performance with a small amount of reference nodes. However, according to the bias–variance tradeoff theory of statistics and machine learning, it tends to cause the over-fitting problem when there a large number of model parameters. During actual application, the LSVR and MSVR algorithms are significantly affected by the over-fitting problem, and it might have poorer localization performance than some early PDM methods under certain scenarios. Take Fig. 3 as an example, Fig. 3(a) shows a classic network with uneven node distribution, there are 300 nodes in the network, and 36 of them are anchor nodes. Fig. 3(b) shows the results obtained with the LSVR localization method, where the circle represents the estimated location of the unknown node, the straight line connects the real coordinates of the unknown node and its estimated coordinates, and the longer the straight line, the greater the localization error. Obviously, LSVR is affected by the over-fitting problem, and the unknown node’s estimated location tends to be a curve.



**Figure 3:** (a) Node distribution, (b) localization result

Later, in 2016, Yan et al. [Yan, Song and Liu et al. (2015)] used the ridge regression (RR) to replace the multi-parameter SVR method and proposed the LRR localization algorithm, and the localization precision of algorithm has been significantly improved. The model built with RR is stable, it improves the prediction precision by reducing the estimation variance, but it has increased the estimation bias. In other words, the RR is a continuous method, which has reduced the regression coefficient without simply abandoning any variable. Because ridge regression fails to reduce any regression coefficient to 0, there are too many variables in the built model, and the model has poor interpretability. Again, take Fig. 4 as an example, assume  $A$  is the source node and  $B, C$  are the destination nodes. According to the neighborhood in Fig. 4, we can see that  $A \rightarrow B$  and  $A \rightarrow C$  only have common path on the two nodes of  $X_1, X_2$ . Therefore,  $A \rightarrow B$  and  $A \rightarrow C$  are only correlated at  $A \rightarrow X_1 \rightarrow X_2$ , and it is obviously unreasonable to impose penalty on the two paths of  $X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow B$  and  $X_2 \rightarrow X_3^* \rightarrow X_4^* \rightarrow C$  with no correlation.



**Figure 4:** Multiple communication path model in multi-hop networks

In 1996, Tibshirani [Tibshirani (1996)] proposed a landmark new variable selection method-LASSO. Different from the idea of traditional variable selection method, this method uses the absolute value function of model coefficient as penalty to compress the regression coefficient of model, so that some coefficients with smaller absolute value won't be directly compressed to 0, which can not only realize selection of significant variable, but also provide corresponding coefficient estimate. Although the LASSO method has some great properties, and it has been broadly used, some of its defects are also reflected in actual application. For example, the LASSO method does not have group's effect, which results in poor accuracy and effects. In addition, the LASSO method does not distinguish between various decomposition coefficients, it provides compression with the same degree, and this will cause over compression of some coefficients, which will further affect the estimation accuracy. In order to address these problems, in 2005, Zou et al. [Zou and Hastie (2015)] proposed the Elastic Net method. This method has combined the merits of both RR and LASSO, which has not only kept the stability of RR, but also possessed the characteristic of variable selection, and as a result, the estimation accuracy can be effectively improved.

### 3 Network model and localization problem description

Assume  $N$  nodes  $\{X_1, X_2, \dots, X_N\}$  are distributed in a network. Without loss of generality, let the first  $M$  ( $M \ll N$ ) nodes of set  $N$  to represent the anchor nodes whose locations are known, and the rest  $N-M$  nodes are unknown nodes. Each of these  $N$  nodes correspond to a definite location and a neighbor node set in the network. Therefore, this network can be abstracted to an undirected graph  $G=(V, E)$  in the network, in which, the vertex set is  $V = \{1, 2, \dots, N\}$ , and the edge set is  $E$ . When and only when node  $i$  and node  $j$  are neighbors in undirected graph  $G$  will  $(i, j) \in E$ . Assume the true location of each anchor node is represented by  $x_i, i \in \{1, 2, \dots, M\}$ ; the true location of node to be estimated is represented by  $x_j$ , and  $j \in \{M+1, M+2, \dots, N\}$ . If any node  $X_i$  is connected to node  $X_j$ , their measurement information can be measured, i.e. the minimum hops, which can be represented by  $p(X_i, X_j) = p_{ij} \triangleq \{0, 1, 2, \dots\}$ . The minimum hop vector of node  $X_j$  and other connected nodes in the network is  $p_j = [p_{j1}, p_{j2}, \dots, p_{jN}]^T$ , and the minimum hop matrix between nodes in the whole network is  $\tilde{P} = [p_1, p_2, \dots, p_N]$  where the hop count matrix between anchor nodes is  $P = [p_1, p_2, \dots, p_M]$ . The localization problem can also be abstracted to the process to estimate the location  $x_j$  of unknown node  $X_j$  under the constraint of  $x_i$  and  $p_{ij}$ . The constraint condition can be abstracted to a mapping function  $f: p(X_i, X_j) \rightarrow d(X_i, X_j), f(p_{ij}) = d_{ij}$  exists for the unknown node, and we expect the estimated location of node  $\hat{x}_j$  to be as close to  $x_j$  as possible. In it,  $d(X_i, X_j) = d_{ij}$  represents the Euclidean distance between nodes, and it has the

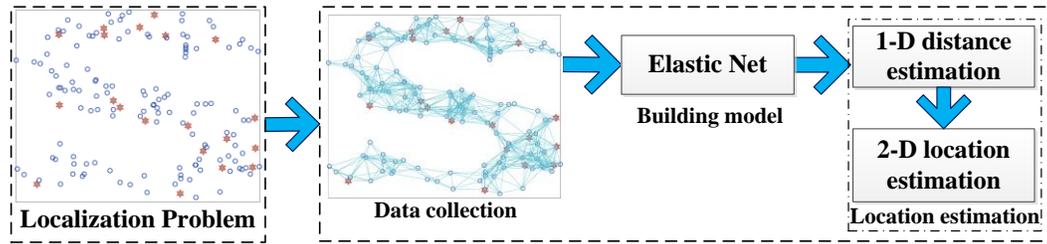
following definition:

$$d(X_i, X_j) = d_{ij} = \left( \sum_{k=1}^p (\mathbf{x}_{ik} - \mathbf{x}_{jk})^2 \right)^{\frac{1}{2}} \quad (1)$$

If node  $X_j$  is not connected to node  $X_i$ , then  $d_{ij} = \infty$ . Because the location of anchor node is known, the distance vector between anchor node  $X_i$  and the rest anchor nodes in the network is  $\mathbf{d}_i = [d_{i1}, d_{i2}, \dots, d_{iM}]^T$ , in which,  $d_{ii} = 0$ . The distance matrix between anchor nodes in the whole network is  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_M]$ .

#### 4 Node location estimation

The basic idea of multi-hop localization based on the Elastic Net is to transform localization to regression prediction problem by building the mapping relation between hops and Euclidean distance. The ML-EN algorithm consists of three steps (as shown in Fig. 5): First of all, measure the hops and Euclidean distance information between anchor nodes; then, build the mapping model between hops and Euclidean distance through EN; finally, estimate the coordinate of unknown node by the multilateration method.



**Figure 5:** The framework of proposed algorithm

The details of above three steps are discussed in the following subsections.

##### 4.1 Measurement step

First of all, after the program starts, any anchor node  $X_i, i \in \{1, 2, \dots, M\}$  will broadcast a HELLO packet to other nodes in the network. After the program has run for a while, the anchor node can obtain the shortest hops from it to other nodes. Then, the anchor node  $X_i$  will send an INFO packet to other anchor nodes in the network, and the packet contains the information of its own location  $\mathbf{x}_i$  and the shortest hops  $p_i$  to other nodes. In this way, any anchor node in the network can build a global shortest hop matrix  $\mathbf{P}$  between anchor nodes and a distance matrix  $\mathbf{D}$  between anchor nodes.

##### 4.2 Model building step

Through data collection during the measurement stage, any anchor node  $X_i, i \in \{1, 2, \dots, M\}$  can collect  $M$  hops-Euclidean distance data pairs  $(\mathbf{p}_k, d_{i,k})$ , which

can connect the anchor node  $X_i$  with the rest anchor nodes in the network. Assume the constraint condition for localization is a mapping function  $f$ ,

$$f_i(\mathbf{p}_k) = \boldsymbol{\theta}_i^T \mathbf{p}_k + e_i = d_{i,k} \quad (2)$$

In which,  $\boldsymbol{\theta}_i$  is the distance estimation model. In order to prevent inconsistent order of magnitudes during the conversion between hops and Euclidean distance with the change of network scale and node communication radius, we have conducted standardized processing to data pair  $(\mathbf{h}_k, d_{i,k})$ . For the whole network, Formula 2 can be transformed to:

$$\mathbf{d}_i = \mathbf{P}\boldsymbol{\theta}_i + \mathbf{e}_i \quad (3)$$

In order to obtain optimal distance estimation model, Elastic Net can be used for estimation, and the objective function of EN is:

$$L(\lambda_1, \lambda_2, \boldsymbol{\theta}_i) = \|\mathbf{d}_i - \mathbf{P}\boldsymbol{\theta}_i\|^2 + \lambda_2 |\boldsymbol{\theta}_i|^2 + \lambda_1 |\boldsymbol{\theta}_i|_1 \quad (4)$$

In which,  $|\boldsymbol{\theta}_i|^2 = \sum_{j=1}^p \theta_{ij}^2$ ,  $|\boldsymbol{\theta}_i|_1 = \sum_{j=1}^p |\theta_{ij}|$ , and  $\lambda_1$  and  $\lambda_2$  are two non-negative penalty parameters. Make  $\alpha = \lambda_1 / (\lambda_1 + \lambda_2)$ ,  $\lambda = \lambda_1 + \lambda_2$ , and Formula 4 can be rewritten as:

$$\hat{\boldsymbol{\theta}}_i = \arg \min \left\{ \|\mathbf{d}_i - \mathbf{P}\boldsymbol{\theta}_i\|^2 + \lambda \left[ (1-\alpha) |\boldsymbol{\theta}_i|_1 + \alpha |\boldsymbol{\theta}_i|^2 \right] \right\} \quad (5)$$

In which,  $\lambda \in [0, \infty)$  is the penalty coefficient;  $\alpha \sum_{j=1}^p |\theta_{ij}| + (1-\alpha) \sum_{j=1}^p \theta_{ij}^2$  can be called the

Elastic Net penalty;  $\alpha \sum_{j=1}^p |\theta_{ij}| + (1-\alpha) \sum_{j=1}^p \theta_{ij}^2$  is the convex combination of LASSO penalty and Elastic Net penalty. When  $\alpha = 0$ , Formula 5 turns to simple ridge regression; when  $\alpha = 1$ , Formula 5 turns to LASSO regression.

In order to obtain the optimal estimation for Formula 5, i.e. the optimal hops-Euclidean distance conversion model, we can use data pair  $(\mathbf{h}_k, d_{i,k})$  and the penalty coefficients  $\lambda_1$  and  $\lambda_2$  to build a new training dataset  $(\mathbf{P}, \mathbf{d}_i)$ :

$$(\mathbf{P}, \mathbf{d}_i) = (1 + \lambda_1)^{-1/2} \begin{bmatrix} \mathbf{P} \\ \sqrt{\lambda_1} \mathbf{I} \end{bmatrix}, \begin{bmatrix} \mathbf{d}_i \\ 0 \end{bmatrix} \quad (6)$$

The optimal conversion model  $\boldsymbol{\theta}_i$  between hops and Euclidean distance from anchor node  $X_i, i \in \{1, 2, \dots, M\}$  to other nodes in the area can be expressed as:

$$\begin{cases} \hat{\boldsymbol{\theta}}_i^* = \arg \min \left\{ \|\mathbf{d}_i - \mathbf{P}\boldsymbol{\theta}_i\|^2 + \left( \lambda_2 / \sqrt{1 + \lambda_1} \right) |\boldsymbol{\theta}_i^*|_1 \right\} \\ \hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i^* / \sqrt{1 + \lambda_1} \end{cases} \quad (7)$$

At this moment, the Least Angel Regression [Efron, Hastie, Johnstone et al. (2004)] and

AIC criterion [Akaike (2011)] can be used to obtain the optimal conversion model  $\hat{\theta}_i$ . According to the Elastic Net, in the optimal solution  $\hat{\theta}_i$  obtained from Formula (7), some penalty coefficients for nodes with weak correlation will be compressed to 0, so that the hops-Euclidean distance conversion model will have sparsity.

### **4.3 Location estimation step**

Through model building step, any unknown node  $X_k, k \in \{M+1, \dots, N\}$  receives  $M$  hops-Euclidean distance conversion models and shortest hops vectors  $\{\hat{\theta}_1, \mathbf{p}_1, \dots, \hat{\theta}_M, \mathbf{p}_M\}$  from the anchor nodes. Let

$$\hat{\mathbf{d}}_k = [d_{k,1}, \dots, d_{k,M}]^T \quad (8)$$

be the estimated distance vector representing the estimated distance between the unknown node  $X_k$  and  $M$  anchor nodes. The estimated distance vector  $\hat{\mathbf{d}}_k$  can be determined by

$$\hat{\mathbf{d}}_k = [\hat{d}_{k,1}, \dots, \hat{d}_{k,M}]^T = \hat{\theta}_1^T \mathbf{p}_k \cdots \hat{\theta}_M^T \mathbf{p}_k. \quad (9)$$

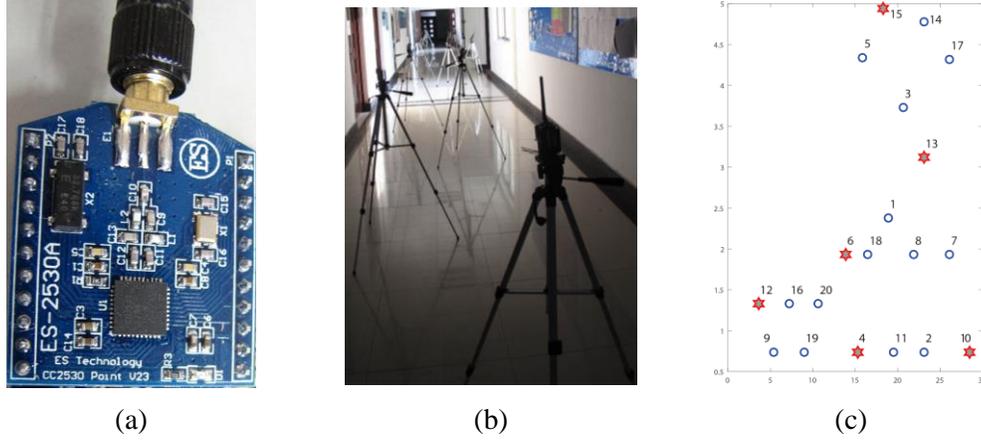
After the estimates of the distances  $\hat{d}_k$  to the anchor nodes are evaluated, multilateration is performed to localize the location of the node  $X_k$ .

## **5 Performance analysis**

In this section, through experiments in various environments, the performance of multi-hop localization algorithm based on the EN algorithm will be analyzed and evaluated. First of all, a group of experiments were conducted in the actual environment (as shown in Fig. 6). In the actual environment, the experiments were conducted in an indoor corridor, and its scope was around  $30 \times 5$  m. We deployed 20 wireless nodes in the corridor, and the node chip is CC2530 produced by TI Company.

Because this paper is dedicated to addressing the influence of uneven node distribution on the localization performance, we will not focus on such factors as the node communication cost, node lifecycle and information delivery method. In the experiment, the influence of the indices of node distribution, proportion of anchor nodes and node density on the localization precision was mainly investigated.

Sometimes, it is difficult to set these parameters. Therefore, in this section, we will still evaluate the performance of localization algorithm through software simulation.



**Figure 6:** (a) Wireless node; (b) Node distribution scenario; (c) Network topology with 20 nodes

We use the root mean square (RMS) as the criterion for localization precision, and RMS can be expressed as:

$$RMS = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} ((\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2)} \quad (10)$$

where  $(\hat{x}_i, \hat{y}_i)$  is the estimated coordinate of unknown node  $X_i, i \in \{M+1, \dots, N\}$ , and  $(x_i, y_i)$  is the real coordinate of unknown node  $X_i$ .

We also compare our method with seven previous methods: (1) the classic DV-hop algorithm proposed in Niculescu et al. [Niculescu and Nath (2003)]; (2) Nearest-4 proposed in Shang et al. [Shang, Shi and Ahmed (2004)]; (3) DV-Maxhop proposed in Shahzad et al. [Shahzad, Sheltami and Shakshuki (2017)]; (4) LWLR-DV-hop proposed in Zhao et al. [Zhao, Su and Shao (2018)]; (5) PDM proposed in Lim et al. [Lim and Hou (2009)]; (6) LSVR proposed in Lee et al. [Lee, Chung and Kim (2013)]; and (7) LRR proposed in Yan et al. [Yan, Song, Liu et al. (2015)].

**5.1 Localization experiment in real scene**

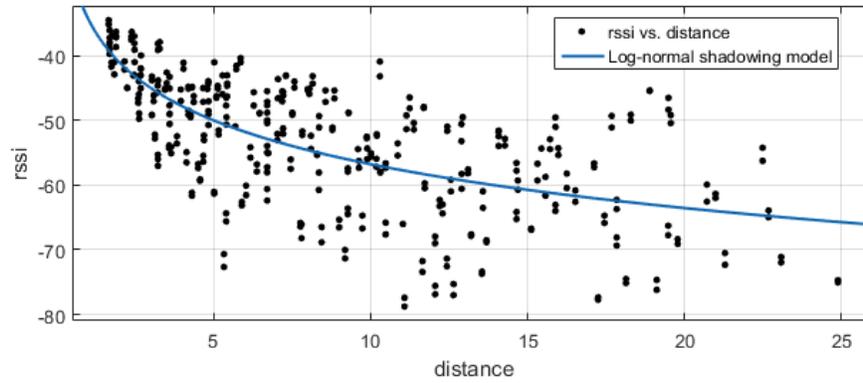
During the multi-hop localization, the hops between two nodes can be determined by setting the threshold value of signal attenuation. Therefore, it requires obtaining the relation between signal attenuation value RSSI and the distance. In order to obtain the relation between RSSI and the distance, for the scenario as shown in Fig. 6, we repeatedly collected communication among 20 nodes for 120 times. Considering that in the actual environment, the signal reflection, scattering or shielding will generate interference with the signal value collected by receiving node, we adopt the Log-normal Shadowing model [Yang, Wu and Liu (2014)] to depict how the signal intensity will change with the change of distance.

$$P_r(d) = P_0(d_0) - 10\eta \log\left(\frac{d}{d_0}\right) + X_\sigma \quad (11)$$

In which,  $P_r(d)$  represents the received signal strength at distance  $d$  and  $P_0(d_0)$  represents the received signal strength at some reference distance  $d_0$  (usually set  $d_0 = 1m$ ),  $\eta$  represents the path-loss exponent, and  $X_\sigma$  represents a log-normal random variable with standard deviation  $\sigma$  that accounts for fading effects. Therefore, after knowing the path loss exponent, the receiving node can measure its distance from the signal source according to the strength of received signal.

After collecting 120 times of data, we adopt the method in Literature [Yan and Qian (2013)] to clean the signal data, and after eliminating outlier signal values caused by the environmental factors, the mean value of signals collected in multiple times is used. In the environment as shown in Fig. 6, at a distance of 1 m, the mean value of RSSI obtained through multiple measurements is -31.55 dBm.

Up until now, we have obtained the signal and distance distribution diagram in the area of Fig. 6 (as shown in Fig. 7).



**Figure 7:** RSSI vs Distance plots for the 20 node deployment indoor

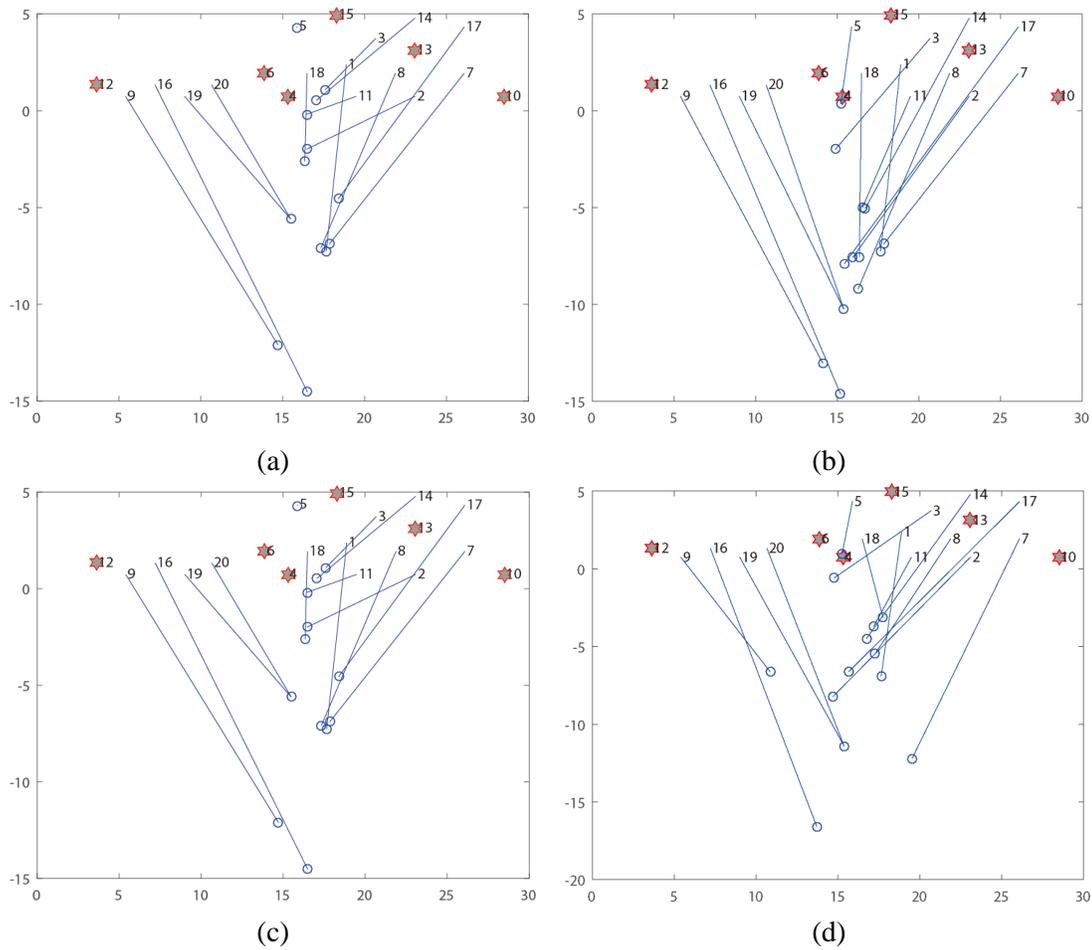
According to Formula 11, we can obtain the distance prediction formula,

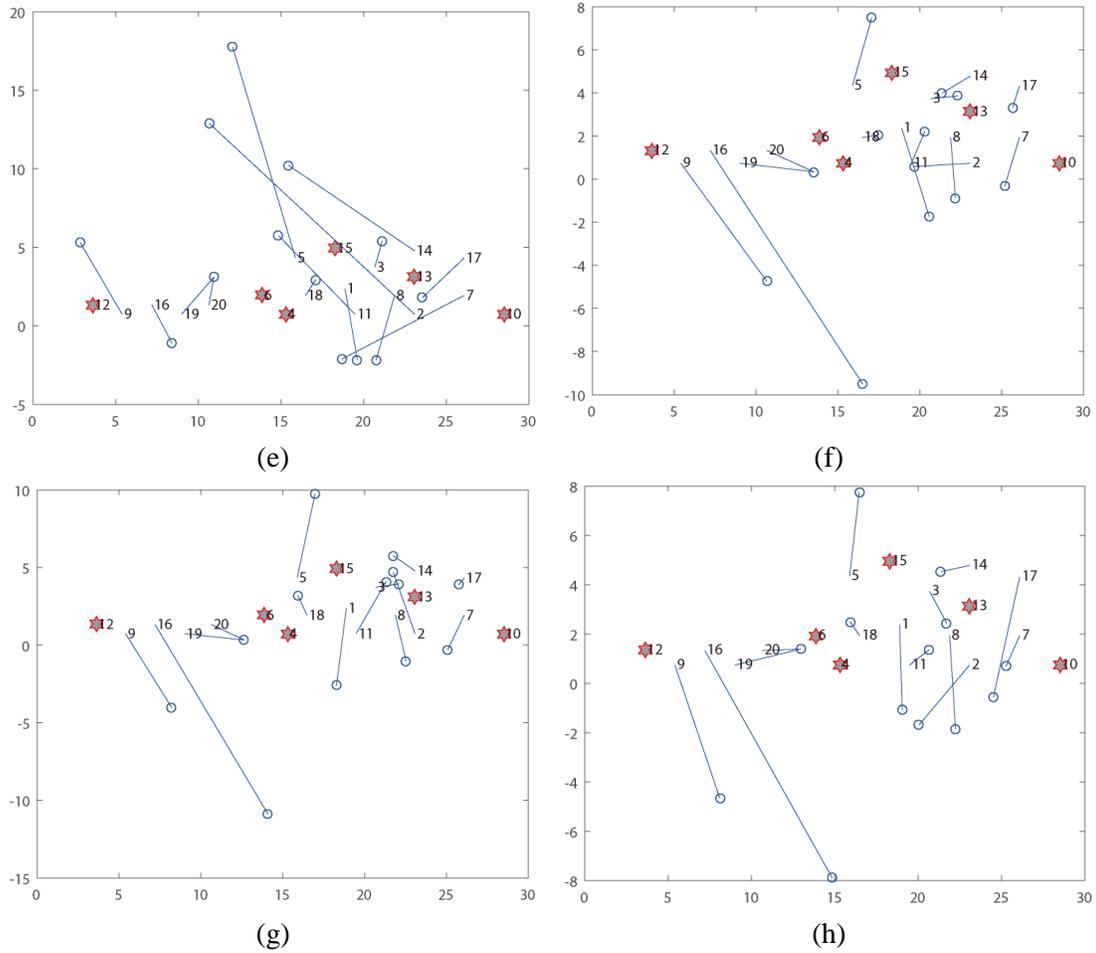
$$\hat{d} = \left( \frac{\text{Pr}}{P_0(d_0)} \right)^{-1/\eta} = \left( \frac{\text{Pr}}{-31.55} \right)^{-1/2.254} \quad (12)$$

In this group of experiments, the node communication radius is set at 4.5 m; nodes 4, 6, 10, 12, 13 and 15 in the network are chosen as the anchor nodes, and the rest 14 nodes are nodes to be localized. The Matlab software was used for computation of the ML-EN algorithm proposed in this paper and other 7 similar algorithms, and the final localization result can be obtained (as shown in Fig. 7). According to Fig. 8, we can see that the DV-Maxhop algorithm has close estimation precision as the DV-hop algorithm. This is because of the restriction of deployment space, which has resulted in failure of hops selection in the DV-Maxhop algorithm, and the DV-Maxhop algorithm will degrade to

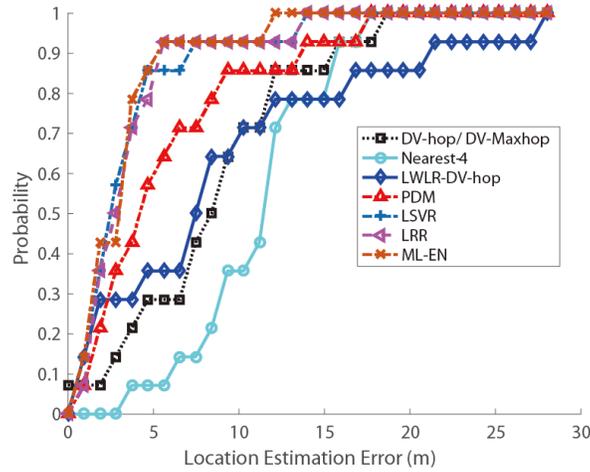
the DV-hop algorithm. The Nearest-4 algorithm and LWLR-DV-hop algorithm are two algorithms with similar theoretical basis, but due to restriction of the number of selectable anchor nodes and the topologic quality of anchor nodes, their localization precisions are lower than that of the DV-hop method. Fig. 8(h) shows the ML-EN method proposed in this paper, it has high localization precision, and the RMS value is 4.5595. The average localization precision (RMS) of ML-EN method is 58.1%, 68.2%, 61.1%, 39.3%, 4.2% and 6.4% higher than the localization precisions of Nearest-4, DV-Maxhop, LWLR-DV-hop, PDM, LSVR and LRR methods respectively.

For localization in the actual environment, we also employ the Cumulative Distribution Function (CDF) to analyze the accumulative distribution of errors of nodes to be localized. We can find when using ML-EN, the localization performance is effectively improved, e.g. more than 80% nodes are localized with RMS is less than 4 m in the real-world environment.





**Figure 8:** Results of location estimation in realistic scenarios. (a) DV-hop, RMS=9.9152; (b) Nearest-4, RMS=12.0545; (c) DV-Maxhop, RMS=9.9152; (d) LWLR-DV-hop, RMS=11.4967; (e) PDM, RMS=7.5795; (f) LSVR, RMS=5.0697; (g) LRR, RMS=5.0385 (h) ML-EN, RMS=4.5595



**Figure 9:** The CDF of the estimation errors in realistic scenarios

**5.2 Simulation experiment**

In this section, the performance of localization algorithm based on Elastic Net will be analyzed and evaluated through simulation experiment. The experiment analyzes the distribution of nodes in the network, the proportion of anchor nodes, and the density of nodes. The specific parameters of the network are shown in Tab. 1.

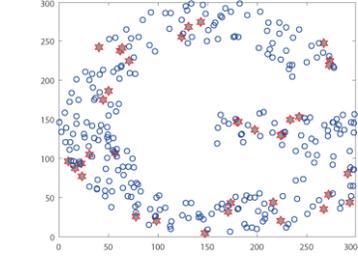
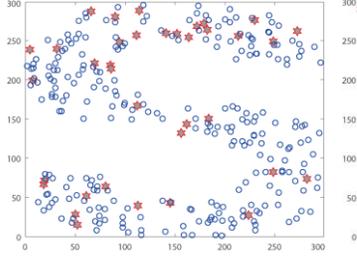
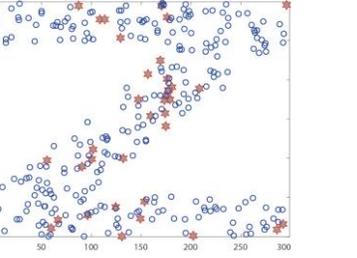
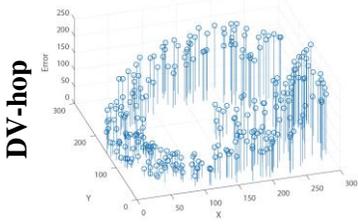
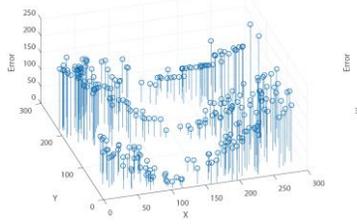
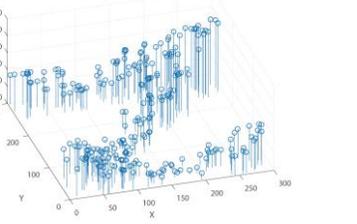
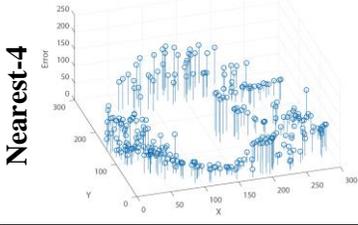
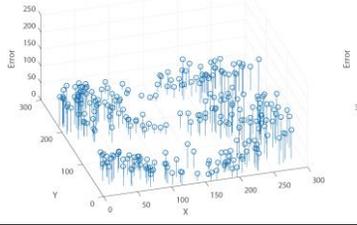
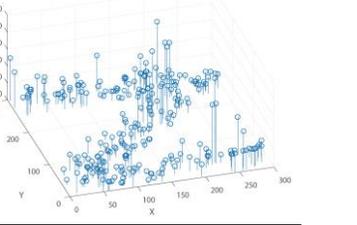
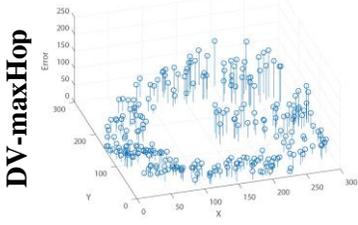
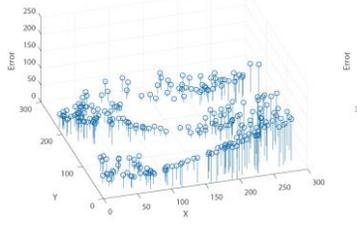
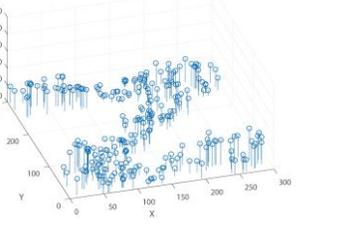
**Table 1:** Network parameters

Total number of nodes	Anchor node ratio	Node distribution	Communication radius
300	80%~16%	In the area of 300*300, the nodes are distributed into G-shaped, S-shaped and Z-shape	35~55

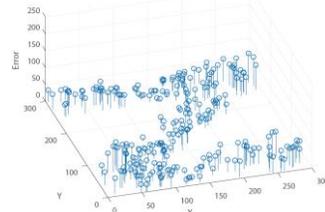
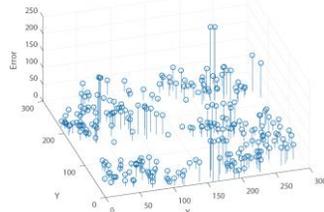
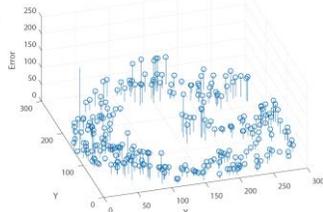
**5.2.1 The uneven distribution of nodes**

Experiments adopt the distribution of letter-shaped nodes to simulate the uneven distribution of nodes. Tab. 2 describes the node distribution and localization results of ML-EN and similar algorithms when the node communication radius is 40 and the anchor nodes account for 12% of total nodes. The first line in Tab. 2 shows the node distribution. The localization results are described with 3-D stem plot, the stem height represents the absolute error of estimation, and the higher the height is, the bigger the localization error.

**Table 2:** Localization results with different nodes distribution

			
<b>DV-hop</b>			
<b>RMS</b>	113.2706	110.3841	86.4903
<b>Nearest-4</b>			
<b>RMS</b>	56.6793	59.334	51.7108
<b>DV-maxHop</b>			
<b>RMS</b>	47.5032	52.9291	43.7317

**LWLR-DV-hop**



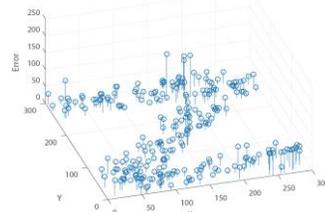
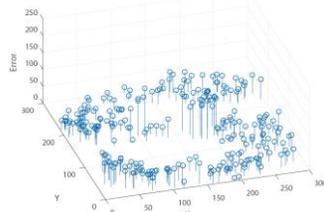
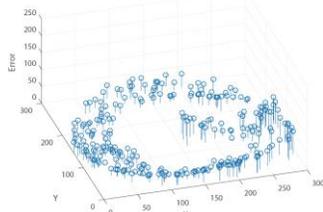
**RMS**

42.1533

48.3389

38.6941

**PDM**



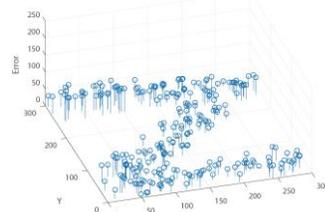
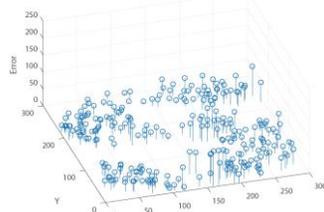
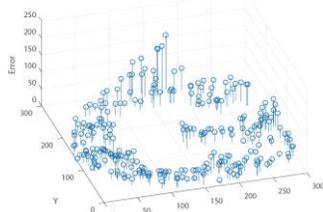
**RMS**

29.379

39.283

28.4368

**LSVR**



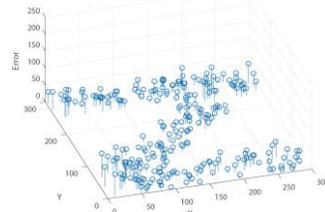
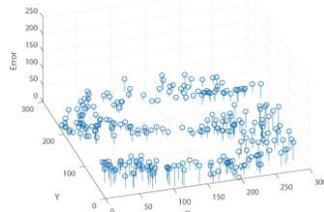
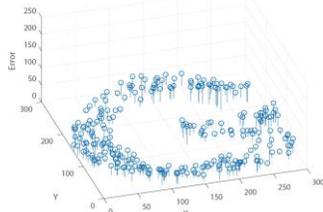
**RMS**

37.0462

31.2208

32.2912

**LRR**

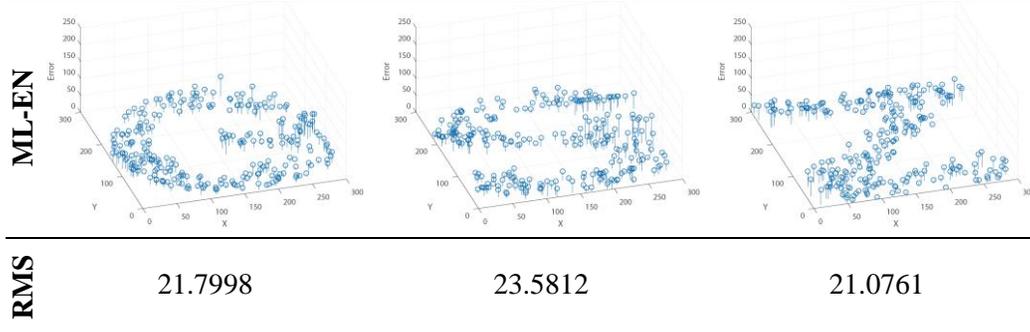


**RMS**

25.8327

28.8076

24.6351



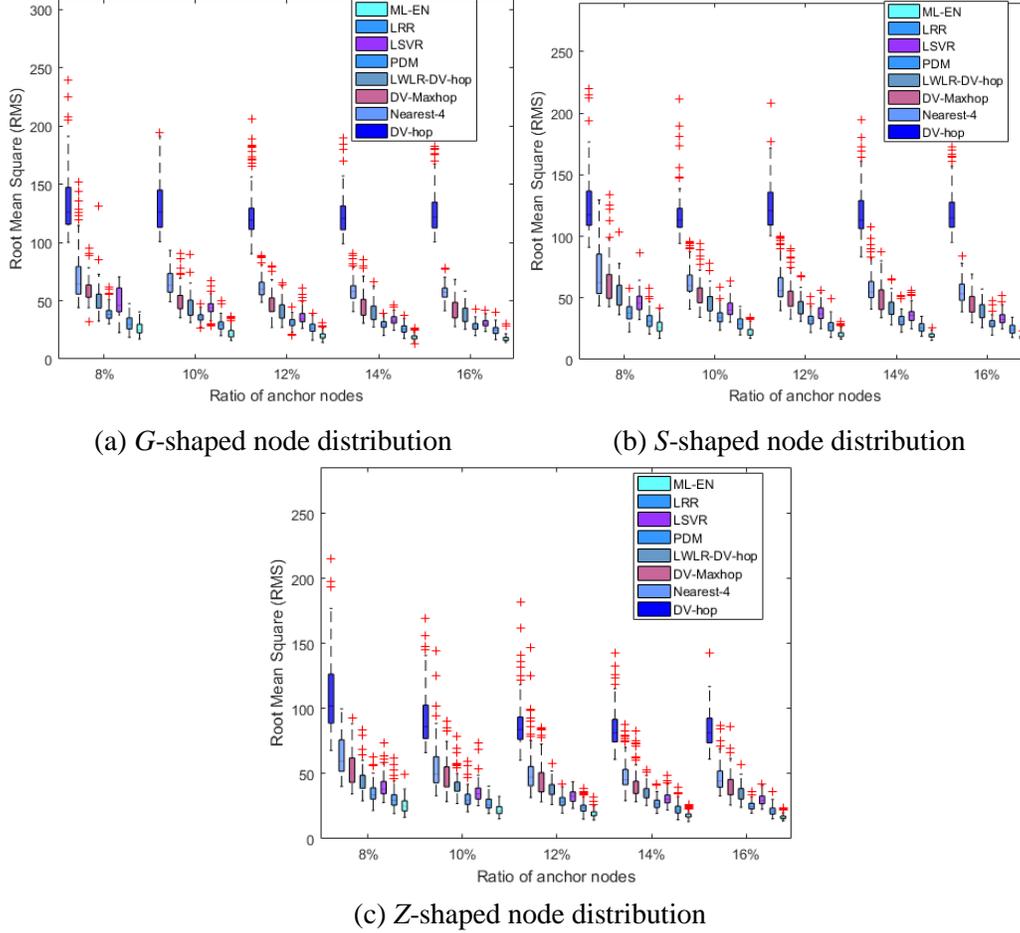
According to the localization result in Tab. 2, we can see that in a network with uneven node distribution, because DV-hop and its improved methods use fixed hops-Euclidean distance conversion coefficient, it results in error during the conversion between hops and Euclidean distance; although the PDM method can solve the error generated during conversion from hops to physical distance, however, due to restriction by problems such as conversion of magnitude order and minimum empirical risk, the improvement of localization performance is very limited; the LSVR method is restricted by the overfitting problem, which causes its localization precision to be higher than that of the PDM method; although the LRR method has higher localization precision than previous methods, the improvement in precision is limited due to fixed ridge parameter, The ML-EN method proposed in this paper has considered various factors constraining the localization precision, so it can provide higher localization precision.

### 5.2.2 Anchor node ratio

Another factor that affects the estimation accuracy of the nodes is the ratio of anchor nodes. In order to reduce the one-sidedness of the experimental results, we conducted 100 simulations for each node distribution network. Considering that the boxplot will not be affected by outliers, and it can describe the discrete distribution of data in a relatively stable way, in this group of experiments, we use boxplot to describe the localization results of multiple simulation experiments under different proportions of anchor nodes (8%~16%).

In Fig. 10, the experimental results show that the DV-hop method does not apply to the localization in anisotropic network. In the G-shape network and S-shape network, its RMS values are both higher than 100; in the Z-shape network, its RMS value is between 70 and 120, which also presents fluctuations with the increase of anchor nodes. The Nearest-4, DV-Maxhop, LRR methods and the ML-EN method proposed by us all have higher localization precision than the DV-hop method, and the precision presents gradual decline with the increase of anchor nodes. The ML-EN method proposed in this paper has the lowest RMS values in all three networks with uneven distributed nodes: in the G-shape network, the average estimation accuracy of ML-EN method is 83.85%, 67.33%, 57.26%, 51.9%, 35.53%, 43.98% and 24.95% higher than that of the DV-hop, Nearest-4, DV-Maxhop, LWLR-DV-hop, PDM, LSVR and LRR methods respectively; in the S-shape network, the average estimation accuracy of ML-EN method is 83.26%, 65.51%, 59.7%, 51.05%, 34.64%, 44.51% and 22.12% higher than that of the LRR method; in the

Z-shape network, the average estimation accuracy of ML-EN method is 77.86%, 62.42%, 57%, 48.01%, 32.1%, 40.04% and 18.97% higher than that of the Nearest-4 and LRR methods respectively.



**Figure 10:** Simulation results of the different algorithms with different number of anchors

### 5.2.3 Node density

In this section, we also evaluate the effect of node density on positioning errors. The number of nodes' average neighbors  $\rho = \frac{N_l \times R^2}{A}$  is usually used to represent the node

density, where  $A$  is the size of the node deployment area,  $R$  is the node communication radius, and  $N_l$  is the number of localizable nodes. In order to change the density of nodes in the network, we adopt the way to adjust the communication radius of the node in the experiment. When the node communication radius changes in the [35, 55] range, in the *G*-shaped network, the variation range of  $\rho$  is 2.16 to 4.19; in the *S*-shape area, the

variation range of  $\rho$  is 2.25 to 3.81; in the Z-shaped area, the variation range of  $\rho$  is 2.38 to 4.53. Fig. 11 shows the effects with different  $R$  on the localization performance. In theory, with the increase of node communication radius, the number of neighbors would increase accordingly, and the per-hop-distance of nodes is closer to the actual distance. However, this is not the case in reality. As described in the literature [Lee and Kim (2011)], in the multi-hop network, the node distribution conforms to the Poisson distribution, and the increase of communication radius will not increase the estimation accuracy, but results in fluctuation. According to Fig. 11, we can find that although the estimation accuracy presents fluctuation with the change of communication radius  $R$ , no matter how the communication radius  $R$  changes, the proposed method in this paper is always superior to the DV-hop, Nearest-4, DV-Maxhop, LWLR-DV-hop, PDM, LSVR and LRR methods.

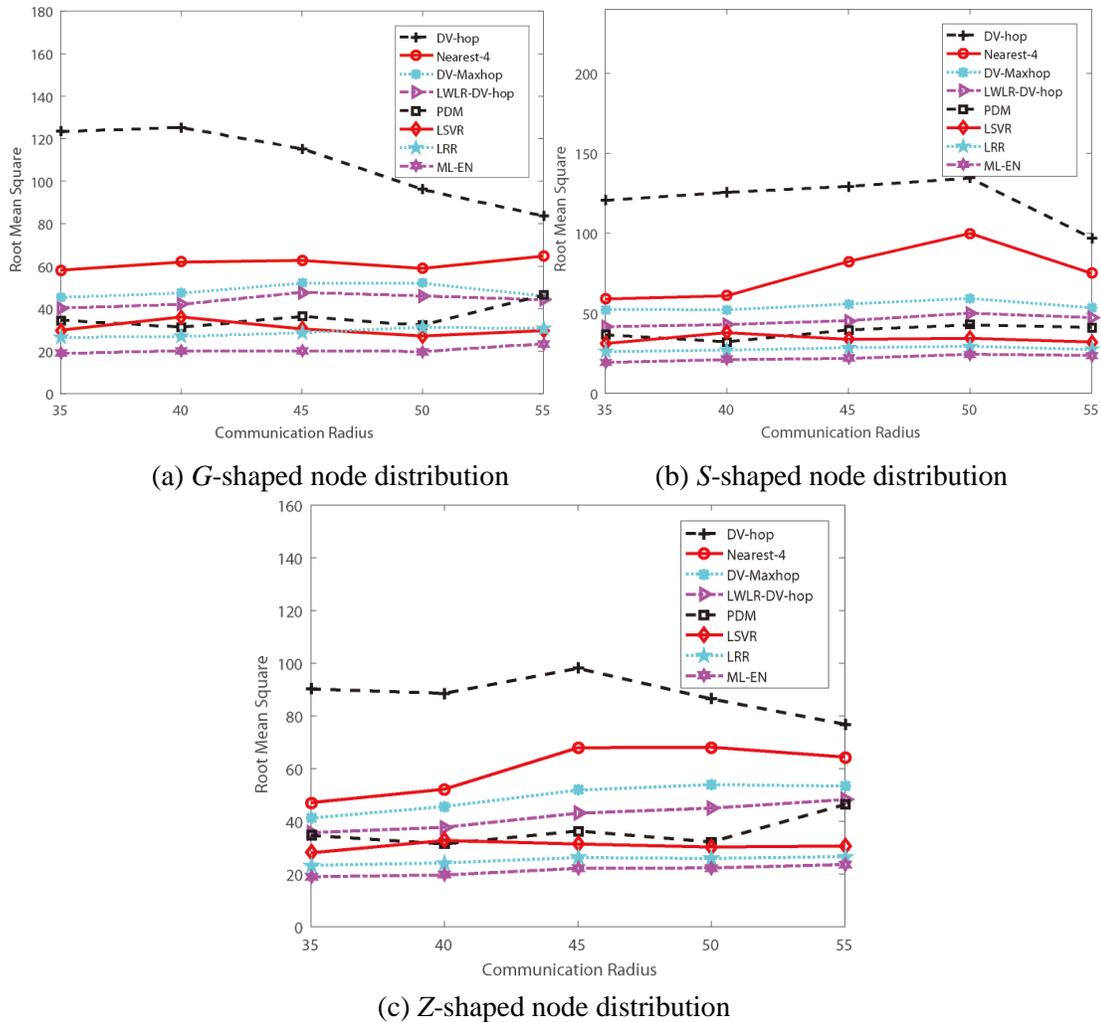


Figure 11: Simulation results of uneven node distribution with different radio range

## 6 Conclusions

A novel multi-hop algorithm for wireless network with unevenly distributed nodes is proposed in this paper. The proposed approach is based on the elastic net, which is a regularization method with good interpretability and stability. In the proposed method, each unknown node estimates its location using the mapping model by elastic net constructed with anchor nodes. In this mapping model, the location of the unknown node is estimated by the information related to the minimum hops of the unknown node to all of the anchor nodes. The proposed approach is applicable to the scenarios where nodes are unevenly distributed. The experiment demonstrated the superior performance of the proposed approach compared to previous multi-hop localization methods in a variety of different distribution environments.

**Acknowledgement:** The paper is sponsored by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (15KJB520009, 16KJD520004), China Postdoctoral Science Foundation (2016M601861), Jiangsu Postdoctoral Science Foundation (1701049A), and the Open Project Program of Jiangsu Key Laboratory of Remote Measurement and Control (YCCCK201603).

## References

- Akaike, H.** (2011): Akaike's information criterion. *International Encyclopedia of Statistical Science*, pp. 25-29.
- Castillo-Cara, M.; Lovón-Melgarejo, J.; Bravo-Rocca, G.; Orozco-Barbosa, L.; García-Varea, I.** (2017): An empirical study of the transmission power setting for bluetooth-based indoor localization mechanisms. *Sensors*, vol. 17, pp. 1-22.
- Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R.** (2004): Least angle regression. *Annals of Statistics*, vol. 32, no. 2, pp. 407-499.
- Hansen, P. C.** (1987): The truncated SVD as a method for regularization. *BIT Numerical Mathematics*, vol. 27, no. 4, pp. 534-553.
- Lee, J.; Chung, W.; Kim, E.** (2013): A new kernelized approach to wireless sensor network localization. *Information Sciences*, vol. 243, pp. 20-38.
- Lee, S.; Kim, K.** (2011): Determination of communication range for range-free multi-hop localization in wireless sensor networks. *Computer Communications and Networks*, pp. 1-4.
- Lim, H.; Hou, J. C.** (2009): Distributed localization for anisotropic sensor networks. *ACM Transactions on Sensor Networks*, vol. 5, no. 2, pp. 1-26.
- Liu, X.; Zhang, S. G.; Bu, K.** (2016): A locality-based range-free localization algorithm for anisotropic wireless sensor networks. *Telecommunication Systems*, vol. 62, no. 1, pp. 3-13.
- Mazhar, F.; Khan, M. G.; Sällberg, B.** (2017): Precise indoor positioning using UWB: a review of methods, algorithms and implementations. *Wireless Personal Communications*, vol. 97, no. 3, pp. 4467-4491.
- Niculescu, D.; Nath, B.** (2003): DV based positioning in Ad Hoc networks.

*Telecommunication Systems*, vol. 22, no. 1-4, pp. 267-280.

**Park, S.; Cho, I. M.; Kim, S. S.; Kim, S. M.** (2014): A portable mid-range localization system using infrared LEDs for visually impaired people. *Infrared Physics and Technology*, vol. 67, pp. 583-589.

**Shahzad, F.; Sheltami, T. R.; Shakshuki, E. M.** (2017): DV-maxHop: a fast and accurate range-free localization algorithm for anisotropic wireless networks. *IEEE Transactions on Mobile Computing*, vol. 16, no. 9, pp. 2494-2505.

**Shang, Y.; Shi, H. C.; Ahmed, A. A.** (2004): Performance study of localization methods for ad-hoc sensor networks. *IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, pp. 184-193.

**Tibshirani, R.** (1996): Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267-288.

**Wang, W. Y.; Lu, D.; Wang, L.; Jia, Q. Q.** (2017): Adaptive interference mitigation in GNSS. <https://www.springer.com/gb/book/9789811055706>.

**Yan, X. Y.; Qian, H. Y.** (2013): A selective beacon node 3D location estimation based on RSSI for wireless sensor network. *Information Technology Journal*, vol. 12, no. 2, pp. 50-60.

**Yan, X. Y.; Song, A. G.; Liu, Y.; Gu, X. Y.** (2015): An efficient tolerant-anisotropic localization for large-scale wireless sensor network. *Seventh International Symposium on Parallel Architectures, Algorithms and Programming*, pp. 41-44.

**Yang, Z.; Wu, C. S.; Liu, Y. H.** (2014): *Location-Based Computing: Localization and Localizability of Wireless Networks*. Tsinghua University Press, China.

**Zhao, W.; Su, S. B.; Shao, F.** (2018): Improved DV-Hop algorithm using locally weighted Lin-ear regression in anisotropic wireless sensor network. *Wireless Personal Communications*, vol. 98, no. 4, pp. 3335-3353.

**Zhou, J. Y.; Shi, J.** (2009): RFID localization algorithms and applications-a review. *Journal of Intelligent Manufacturing*, vol. 20, pp. 695-707.

**Zou, H.; Hastie, T.** (2015): Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B*, vol. 67, no. 2, pp. 301-320.