Detection and Localization Algorithm Based on Sparse Reconstruction

Zhao Tang[1] and Xingcheng Liu[1,2,3(✉)]

[1] School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China 510006
isslxc@mail.sysu.edu.cn
[2] School of Information Science, Guangzhou Xinhua University, Guangzhou, China 510520
[3] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

Abstract. Wireless sensor networks (WSN) have received wide attention in many fields of applications. Secure localization is a critical issue in WSN. In the presence of malicious anchors, the traditional solution is to detect the malicious anchors, use the information collected from the normal anchors and then estimate the target location. The way of thinking and operating is reformed by modeling the behavior of the malicious anchors as perturbations. The secure localization is formulated as a sparse reconstruction problem. A gradient projection algorithm with variable step sizes is proposed to solve the sparse reconstruction. The proposed algorithm utilizes sparse reconstruction formulation for obtaining anchors information and identifying the malicious anchors by exploiting the sparsity of malicious anchors. The proposed algorithm is further modified to enhance the accuracy. The simulation results demonstrate that the proposed algorithm can effectively identify the cheating anchors and achieve great target anchors localization accuracy. The proposed algorithm performs better than any other algorithms of interest.

Keywords: Wireless Sensor Networks (WSN) · Malicious anchor detection · Sparse recovery · Gradient projection · Secure localization

1 Introduction

In the scenario of wireless sensor networks (WSN), a large quantities of wireless sensor nodes are anchored and deployed to collect information and process data [1]. In the coverage, WSN monitor the objects effectively and send considerable

This work was supported by the Key Project of NSFC-Guangdong Province Joint Program (Grant No. U2001204), the National Natural Science Foundation of China (Grant Nos. 61873290 and 61972431), the Science and Technology Program of Guangzhou, China (Grant No. 202002030470), and the Funding Project of Featured Major of Guangzhou Xinhua University (2021TZ002).

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 866–879, 2021.
https://doi.org/10.1007/978-3-030-89814-4_64

information to the observer [2]. WSN have been applied in various fields such as underwater exploration, environment monitoring, fire surveillance [3]. Due to the key roles of WSN and fragility of nodes anchored in the environment, node secure location is a significant issue [4].

Limited by the function of WSN and the vulnerability of the nodes anchored in the wild environment, the node secure localization significantly matters. The current node localization formulation in WSN is usually classified into two categories: range-based and range-free mechanisms [5]. The range-based algorithms adopt ranging technology to gather the distance information among nodes and unknown nodes, by means of Radio Signal Strength Indicator(RSSI) [6], Time-Difference of Arrival (TDoA), Angle of Arrival (AoA) [7]. The range-free algorithms make use of the connectivity of networks to gather information of the target anchors, by means of Distance Vector Hop(DV-Hop) [8], Approximate Point-in-Triangulation Test(APIT) [9] and so on.

Localization systems are vulnerable to attackers, who wish to invalidate the WSN' functionality. Therefore, it is of significance to focus on the accuracy and robustness of the localization. Secure localization algorithms proposed before are straightforward: the first step is to filter out cheating anchors based on the consistency of the anchors signal data set, the second one is to locate the target. An algorithm implementing Isolation Forest is proposed to filter out the malicious anchors [10]. MNDC and EMDC algorithms exploit cluster and evaluation of the consistency of RSSI and ToA measurement to detect the cheating anchors [11]. Gradient Descent (GD) method with a selective pruning stage for inconsistent measurements is used to achieve localization [12].

In this paper, we propose a detection algorithm by modeling malicious anchors' misbehavior into perturbations and reconstructing the sparse vector to detect the malicious anchors and then locating the target. The paper is organized as follows: Section 2 presents the network model and formulation. The proposed algorithm is developed in Sect. 3. The comparative experiments and simulation results are demonstrated in Sect. 4. The summary and conclusion are drawn in Sect. 5.

2 Network Model and Problem Formulations

2.1 Network Model

The network localization and the algorithm are considered in two-dimensional sensor networks where the measurement of distance is stable and available through ranging technology of TDoA. Each node provides location reference, including its location information and the measured distance. The notations used in this paper are listed in Table 1.

2.2 Problem Formulation

Especially, there is a WSN including anchor set $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\}$. To describe easily, we make one common node as the target anchor with the real location as $\boldsymbol{t}=[t_x, t_y]^T$. And thus the true distance between target anchors and others can

Table 1. Summary of notations

Notations	Meanings
n	Number of anchors
m	Number of malicious anchors
r	Number of reference anchors
k	Number of observations
$\boldsymbol{A_i}$	Location of the i-th anchor
$\boldsymbol{t}$	Location of the target node
d_i	Measured distance of the i-th anchor
n_i	Noise components of the i-th anchor
$\boldsymbol{p}$	Sparse vector
u_i	Attack components of the i-th anchor
ϕ_i	Positive part of u_i
ψ_i	Negative part of u_i
α^k	Step size of the k-th iteration
λ^k	Positive Scalars of the k-th iteration
P	Operation of projection
β, μ	Scalars for size election
σ_δ	Strength of the attacks

be represented $\|\boldsymbol{A_i} - \boldsymbol{t}\|, i = 1, 2, ..., n$. The cheating anchors report their fake measurement results, which can be simulated in Eq. 1.

$$d_i = \|\boldsymbol{A_i} - \boldsymbol{t}\| + n_i + u_i, i = 1, 2, ..., n. \tag{1}$$

where d_i is the distance in the presence of measurement errors and malicious anchors. n_i represents the random errors, which are given by $n_i \sim \mathcal{N}(0, \sigma^2)$. u_i simulates these misbehaviors attributed by malicious anchors, which are bounded by $\mathcal{N}(\mu_\delta, \sigma_\delta^2)$. μ_δ is the mean, σ_δ^2 is the variance.

The component of u_i is nonzero if i-th anchors is cheating, otherwise the value of u_i is zero or nearly zero. Since the set of u_i and the target location are unknown in advance, we arrange the set of u_i, t_x and t_y to sparse vector $\boldsymbol{p} = \{u_1, u_2, ...u_n, t_x, t_y\}$. The goal is converted to the recovery of $\boldsymbol{p}$.

3 Proposed Algorithm

Based on the above assumption, malicious anchors detection and target anchors localization are formulated into the sparse reconstruction problem. In this section, we proposed the algorithm using Basic Gradient Projection [13] for sparse reconstruction and sequential probability ratio testing to locate the target anchors and identify the cheating anchors. The algorithm includes the following steps:

1. Determine the initial target localization by recursive weighted least squares.
2. Perform the sparse reconstruction by Basic Gradient Projection.

The flow chart of the proposed algorithm is presented in Fig. 1. More specific description is presented as follows.

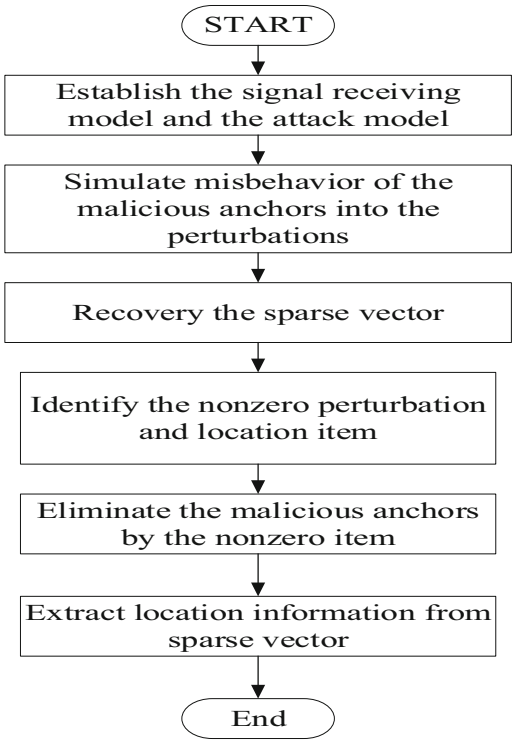

Fig. 1. Flow chart of the proposed algorithm

3.1 Determine the Initial Location

The following steps of the proposed algorithm benefit from starting points with an accurate value. Hence, we firstly proposed recursive weighted linear least squares for locating the initial location. The recursive weighted least squares are relatively independent.

In the received signal strength measurement, we can make use of the energy loss of signal to measure the distance between transmitter and receivers. The logarithmic distance path loss model [13] is Eq. 2.

$$P_R = P_{Ti} - 10a \log \frac{d_i}{d_0} + \varepsilon_i, i = 1, 2, ..., n. \tag{2}$$

where P_{Ti} presents the power of the i-th transmitter, and P_R represents the signal of power from the target anchor. a denotes path loss exponent. d_i denotes the

distance between the i-th anchor and the target node, while d_0 is the referenced distance. ε_i is the measurement noise, which is bounded with $N(0, \sigma_i^2)$. After rewriting the above equal by dividing $10a$, we can get Eq. 3.

$$z_i = 10^{y_i}, i = 1, 2, ..., n. \tag{3}$$

where $z_i = 10^{\frac{P_{Ti} - P_R}{10a}}$, $y_i = lgd_i + \frac{\varepsilon_i}{10a}$. After UT and mathematical transform, we can get the mean and variance of z_i, which is denoted as Eq. 4, Eq. 4.

$$\overline{z_i} \approx \alpha_i \|\boldsymbol{A_i} - \boldsymbol{t}\|^2, \tag{4}$$

$$\sigma_{z_i}^2 \approx \beta_i \|\boldsymbol{A_i} - \boldsymbol{t}\|^4. \tag{5}$$

where $\alpha_i = \frac{2}{3} + \frac{1}{6}10^{\frac{\sqrt{3}\sigma_i^2}{5a}} + \frac{1}{6}10^{-\frac{\sqrt{3}\sigma_i^2}{5a}}$ and $\beta_i = \frac{2}{3}(1 - \alpha_i)^2 + (10^{\frac{\sqrt{3}\sigma_i^2}{5a}} - \alpha_i)^2 + \frac{1}{6}(10^{-\frac{\sqrt{3}\sigma_i^2}{5a}} - \alpha_i)^2$ [13]. The formulate the linear system model is Eq. 6,

$$\mathbf{b} = \mathbf{A}\mathbf{t} + \mathbf{w}. \tag{6}$$

We select the r-th anchor node as the reference anchor node, $\mathbf{b}$ is the observed vector,

$$\mathbf{b} = \begin{bmatrix} \frac{z_1}{\alpha_1} - \frac{z_r}{\alpha_r} + (x_r^2 + x_r^2) - (x_1^2 + x_1^2) \\ \frac{z_2}{\alpha_2} - \frac{z_r}{\alpha_r} + (x_r^2 + x_r^2) - (x_2^2 + x_2^2) \\ \cdots \\ \frac{z_n}{\alpha_n} - \frac{z_r}{\alpha_r} + (x_r^2 + x_r^2) - (x_n^2 + x_n^2) \end{bmatrix}. \tag{7}$$

while $\mathbf{A}$ is coefficient matrix,

$$\mathbf{A} = 2 \begin{bmatrix} x_r - x_1 & y_r - y_1 \\ x_r - x_2 & y_r - y_2 \\ \vdots & \vdots \\ x_r - x_n & y_r - y_n \end{bmatrix} \tag{8}$$

We denoted the covariance matrix as $\mathbf{c}(\mathbf{t})$. The solution of the weighted linear least squares for Eq. 6 is reduced to the minimization objective function $f(\mathbf{t})$, with the solution as Eq. 10.

$$f(\mathbf{t}) = (\mathbf{b} - \mathbf{A}\mathbf{t})^T * \mathbf{c}(\mathbf{t})^{-1} * (\mathbf{b} - \mathbf{A}\mathbf{t}). \tag{9}$$

$$\hat{\mathbf{t}} = [\mathbf{A}^T \mathbf{c}(\hat{\mathbf{t}})^{-1} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{c}(\hat{\mathbf{t}})^{-1} \mathbf{b}. \tag{10}$$

In order to calculate accurately, the recursive formula can be obtained by putting the solution into the covariance matrix as Eq. 11.

$$\hat{\mathbf{t}}^{\mathbf{k}} = [\mathbf{A}^T \mathbf{c}(\hat{\mathbf{t}}^{\mathbf{k-1}})^{-1} \mathbf{A}]^{-1} \mathbf{A}^t \mathbf{c}(\hat{\mathbf{t}}^{\mathbf{k-1}})^{-1} \mathbf{b} \tag{11}$$

We terminate with the solution $\hat{\mathbf{t}}^{\mathbf{k}}$ if the stopping criterion, $\|\hat{\mathbf{t}}^{\mathbf{k}} - \hat{\mathbf{t}}^{\mathbf{k-1}}\| \leq \gamma$, is satisfied.

3.2 Gradient Projection for Sparse Reconstruction

In this section, we assemble the unknown perturbation u_i, i=1, 2, ...,n and the target location $\mathbf{t}=[t_x, t_y]^T$ into unknown vector $\boldsymbol{p}=[u_1, u_2, ..., u_n, t_x, t_y]$ with sparse features. The goal is converted to the recovery of $\boldsymbol{p}$. The problem can be transformed to the optimization problem:

$$\hat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}} \mathbf{G}(\boldsymbol{p}) = \arg\min_{\boldsymbol{p}} \sum (d_i - \|\mathbf{A} - \mathbf{t}\|_2 - u_i)^2 + \lambda \|\mathbf{u}\|_1 \tag{12}$$

To recover this vector, the Basic Gradient Projection algorithm is proposed to reconstruct the vector $\boldsymbol{p}$. Two vectors with positive value, $\boldsymbol{\phi}=[\phi_1, \phi_2, ..., \phi_n]^T$ and $\boldsymbol{\psi}=[\psi_1, \psi_2, ..., \psi_N]^T$, are introduced to split the $\boldsymbol{p}$ into positive and negative part. $\mathbf{G}(\boldsymbol{p})$ can be split into loss function part and regularization function part. Let the vector $\boldsymbol{\tau}=[\boldsymbol{\phi}^T, \boldsymbol{\psi}^T, \mathbf{t}^T]^T$ be the entire unknown vector in this process. Equation 12 can be rewritten as:

$$\hat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}} \mathbf{G}(\boldsymbol{p}) = \Sigma (d_i - \|\mathbf{A} - \mathbf{t}\|_2 - \phi_i + \omega_i)^2 + \lambda \cdot 1_N^T (\boldsymbol{\phi} + \boldsymbol{\psi}). \tag{13}$$

The process of iteration is:

$$\begin{cases} \boldsymbol{v}^k = P(\boldsymbol{\tau}^k - \alpha^k \nabla G(\boldsymbol{\tau}^k)) \\ \boldsymbol{\tau}^{k+1} = \boldsymbol{\tau}^k + \lambda^k (\boldsymbol{v}^k - \boldsymbol{\tau}^k) \end{cases} \tag{14}$$

where α^k is the variable step size, λ^k is a positive scalar. $P(\mathbf{z})$ denotes the operation of projecting $\mathbf{z}$, specially projecting onto the corresponding positive orthant along the negative gradient direction. Before the initial estimation, we can get the start point $\hat{\mathbf{t}}=[\hat{t_x}, \hat{t_y}]$.

There are several step selection schemes. In our case, the iteration points produced tend to locate the boundary of the set. We choose the Armijo rule [13] along the projection arc, in which the value of λ^k is 1 for all k and α^k is the first number in the sequence of $\{1, \beta, \beta^2, ...\}$ until the inequality. 15 meets,

$$\begin{aligned} &\mathbf{G}(P(\boldsymbol{\tau}^k - \alpha^k \nabla \mathbf{G}(\boldsymbol{\tau}^k))) \\ \leq\ &\mathbf{G}(\boldsymbol{\tau}^k) - \mu \nabla \mathbf{G}(\boldsymbol{\tau}^k)^T (\boldsymbol{\tau}^k - P(\boldsymbol{\tau}^k - \alpha^k \nabla \mathbf{G}(\tau^k))), \end{aligned} \tag{15}$$

where $\beta \in (1, 2)$ and $\mu \in (0, 0.5)$. After fixing the value of α^k, we set $\boldsymbol{\tau}^k = P(\boldsymbol{\tau}^k - \alpha^k \nabla \mathbf{G}(\boldsymbol{\tau}^k))$. The iteration is terminated with the solution $\boldsymbol{\tau}^{k+1}$ by the stopping criterion.

4 Simulation Results

To evaluate the proposed algorithm, the mean localization Error(MLE) and the metrics set are introduced [14], including True Positive Rate(TPR) referring the proportion of correctly identifying malicious anchors; False Negative Rate(FNR) referring the possibility falsely identifying the cheating node as an honest node;

False Positive Rate (FPR) referring the possibility falsely identifying an honest node as a malicious one.

The evaluation of the proposed algorithms also includes comparison with other algorithms, Malicious Anchor Node Detection based on Isolation Forest(MANDIF) [10], using isolation forest and sequential probability ratio testing to detecting the malicious nodes; Malicious Nodes Detection using Clustering and Consistency (MNDC) and Enhanced Malicious Nodes Detection using Clustering and Consistency (EMDC) [11]; GD algorithm with fixed steps and variable steps [12]. Two kinds of GD algorithms are proposed, one is the fixed step size algorithm GD, the other is the variable step size algorithm GD. For better performance, we took the variable step size algorithm into the experiment. The rule of the change of the step size is $\gamma(i) = 15 - \frac{15(i-1)}{M}$, in which $\gamma(i)$ represents the step size of the i-th iteration, and M is the maximum number of iterations.

Table 2. Setting of Experiment I

n	m	β	μ	α_0	λ^k	K
30	1	0.5	0.1	1	1	1000

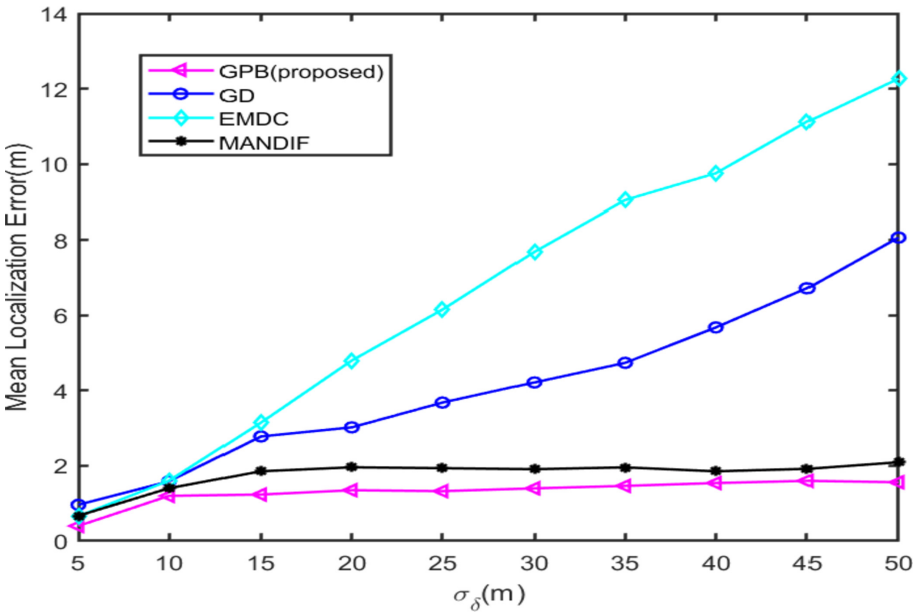

Fig. 2. Mean Localization Error (MLE) curves with σ_δ

In the simulation experiment, we deploy m anchors containing n malicious anchors randomly in the square field of $100m \times 100m$. The experiments below were repeated over 1000 times to obtain accurate and stable results. In Experiment I, we set the relatively simple environment where the malicious anchors are in a small scale. The corresponding parameters are summarized in Table 2. σ_δ is the stength of the attacks. The Mean Localization Error (MLE) with varying σ_δ from 5 to 50 is manifested as Fig. 2. It is can be depicted that while the EMDC and GD performance degrades as σ_δ increases, the MANDIF and GPB algorithms perform with stable and excellent features. Within the value range of σ_δ, the average value of MLE of EMDC, GD, MANDIF and GPB are 6.623, 3.948, 1.914 and 1.308 respectively. The proposed algorithm can decrease the localization error remarkably: by 80.3%, 66.9%, 31.7% compared to EMDC, GD and MANDIF respectively.

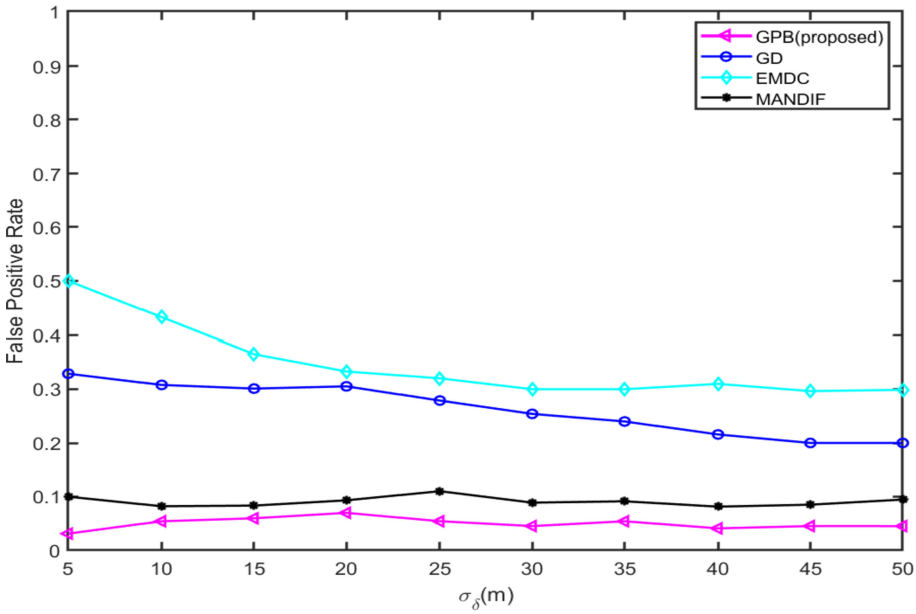

Fig. 3. FPR curves with σ_δ

The False Negative Ratio(FNR), True Positive Ratio(TPR), and False Negative Ratio(FNR) curves of comparison algorithms and the proposed algorithm with varying σ_δ are shown in Fig. 3, Fig. 4, Fig. 5 respectively. MNDC is proposed to achieve malicious anchors detection and secure localization. There is a very important premise condition to implement the method in EMDC: to guarantee

the measurements of RSSI are not attacked while the measurements of ToA are under attack. This precondition can be difficult to guarantee in the practical scenario. The more violent the attack is, the larger the detection interval will be. As a result there will be a lower false detection rate. However, the FPR of GPB remains below 0.0699, The TPR remains above 0.9200, performing well in a violent-attack environment and soft-attack environment. The stable and excellent capability of the proposed algorithm comes from the fact that sparse recovering is not affected by the size of the value of a non-zero item.

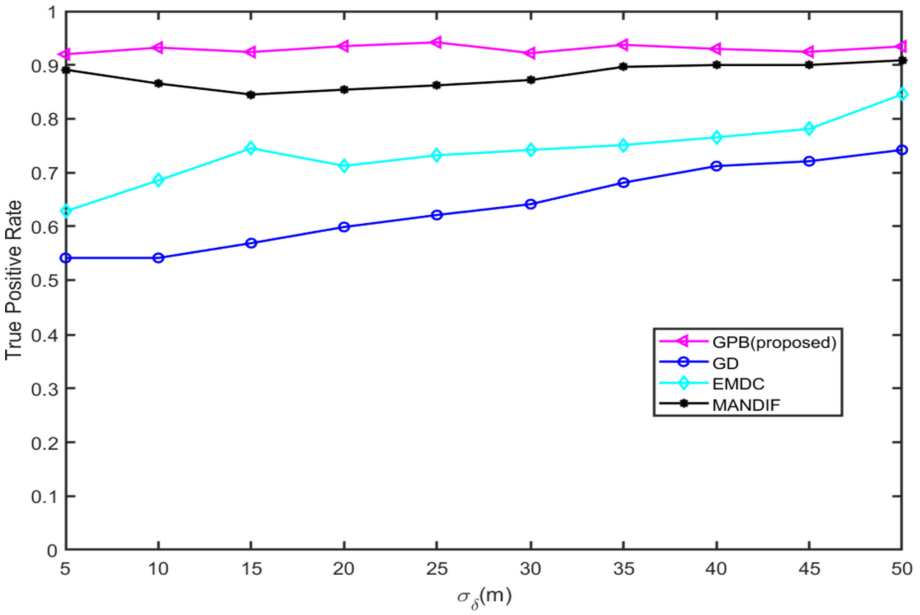

Fig. 4. TPR curves with σ_δ

Then, we consider a situation where the attack is more violent and the malicious anchor nodes occupy a larger proportion. Based on the theorem and the fact that: as long as the number of the malicious nodes $m \leq \frac{n-2}{2}$, the target node localization and all malicious anchors identification can be achieved at the same time. We set the number of the malicious nodes to take up 40% of the total nodes. Like Experiment I, m anchors including n malicious anchors were deployed randomly in the field of $100m \times 100m$. The corresponding parameters are summarized in Table 3.

Table 3. Setting of experiment II

n	m	β	μ	α_0	λ^k	K
30	9	0.5	0.1	1	1	1000

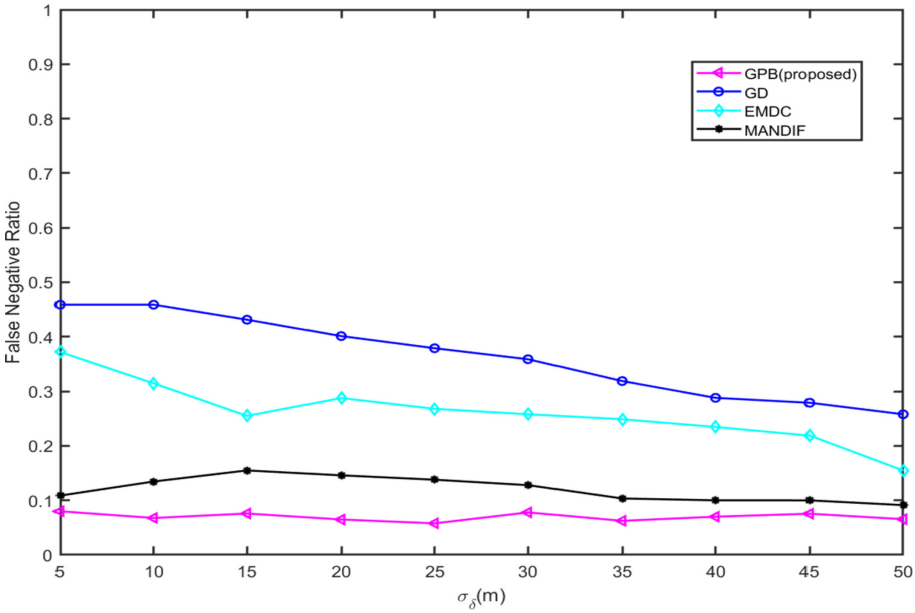

Fig. 5. FNR curves with σ_δ

The Mean Localization Error (MLE) manifested as Fig. 6. It is worth of denoting that, within the value range of σ_δ, the average value of localization error of EMDC, GD, MANDIF and GPB are 20.78, 18.84, 13.59 and 13.59. The positioning accuracy of other algorithms has declined, while the proposed one remains accurate, decreasing by 43.1%, 37.2%, 13.0% compared with EMDC, GD and MANDIF respectively.

The FPR, TPR and FNR curves of the proposed algorithm and other algorithms are shown in Fig. 7, Fig. 8, Fig. 9. The FPR of GPB keeps below 0.0812, The TPR remains above 0.9010, performing better than other algorithms.

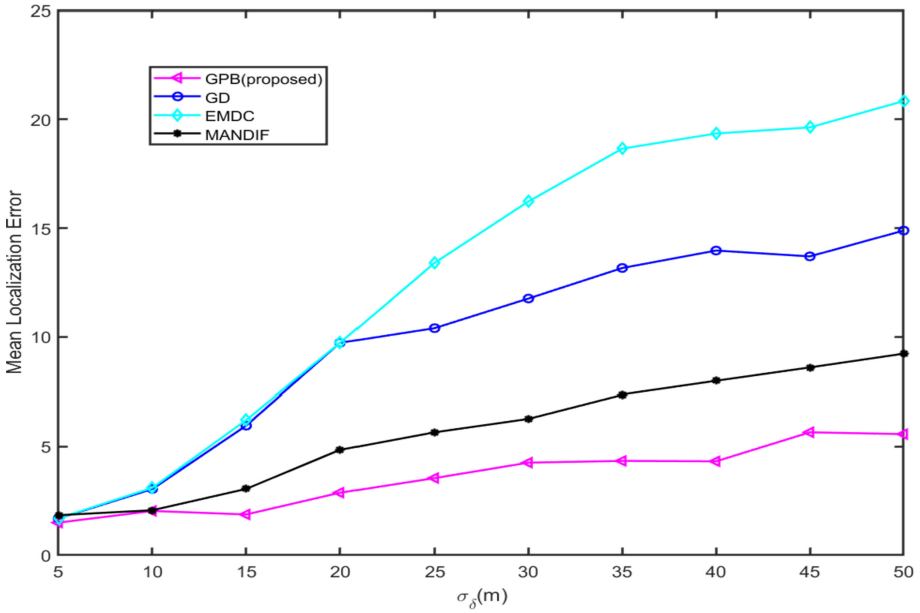

Fig. 6. Mean Localization Error (MLE) curves with σ_δ

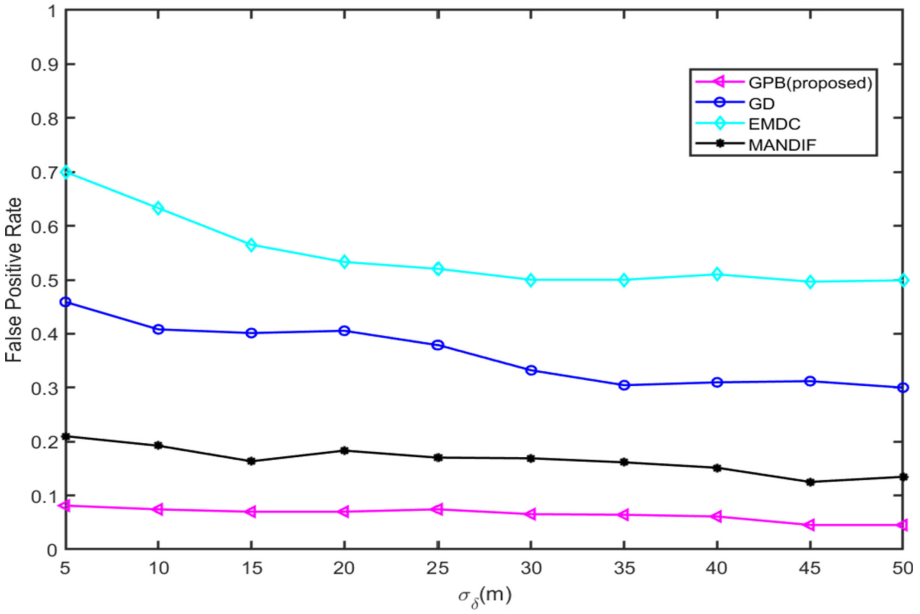

Fig. 7. FPR curves with σ_δ

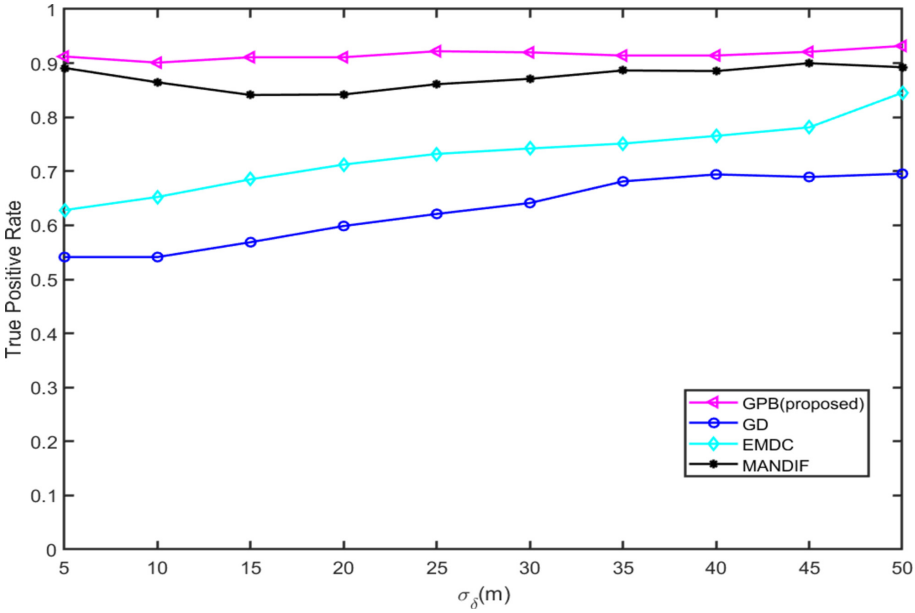

Fig. 8. TPR curves with σ_δ

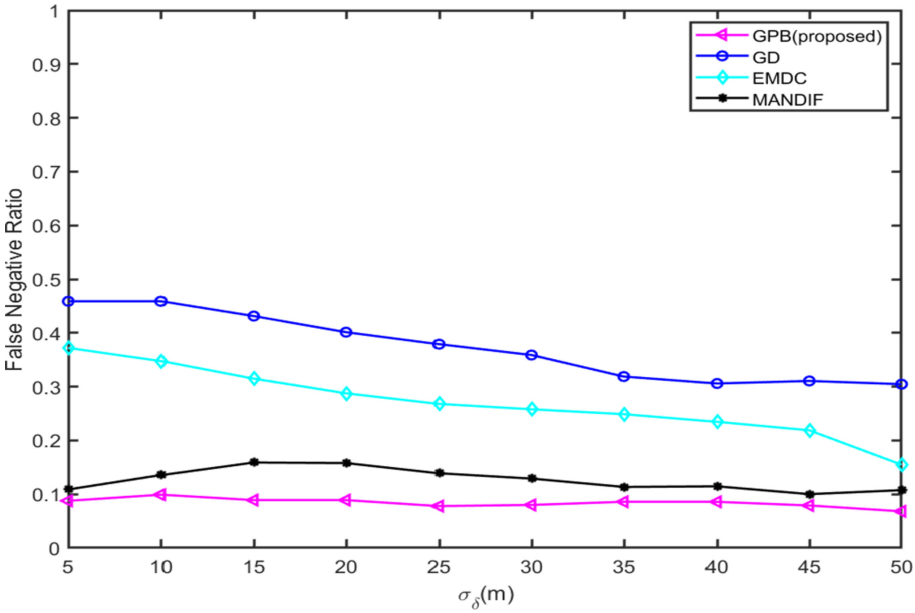

Fig. 9. FNR curves with σ_δ

5 Conclusions

In this paper, the localization problem is formulated as the sparse vector recovery problem. The Gradient Projection Basic algorithm(GPB) is proposed to identify the non-zero item and detect the malicious anchors. In the early stage of the proposed algorithm, the recursive weighted linear square is proposed to obtain the initially estimated position. The comparative experiments and simulation results demonstrate that the proposed algorithm can identify the malicious anchors and achieve successful localization with the probability above 0.9, which outperforms other algorithms of interest.

References

1. Xiong, J., Zhao, M., Bhuiyan, M.Z.A., Chen, L., Tian, Y.: An AI-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of IoT. IEEE Trans. Industr. Inf. **17**(2), 922–933 (2021)
2. Tian, Y., Wang, Z., Xiong, J., Ma, J.: A blockchain-based secure key management scheme with trustworthiness in DWSNs. IEEE Trans. Industr. Inf. **16**(9), 6193–6202 (2020)
3. Zhou, B., Chen, Q.: On the particle-assisted stochastic search mechanism in wireless cooperative localization. IEEE Trans. Wireless Commun. **15**(7), 4765–4777 (2016)
4. Jiang, W., Xu C., Pei, L., Yu, W. Multidimensional scaling-based TDOA localization scheme using an auxiliary line. IEEE Signal Process. Lett. **23**(4), 546–550 (2016)
5. Patwari, N., Ash, J.N., Kyperountas, S., Hero, A.O., Moses, R.L., Correal, N.S.: Locating the nodes: cooperative localization in wireless sensor networks. IEEE Signal Process. Mag. **22**(4), 54–69 (2005)
6. Luo, Q., Peng, Y., Li, J., Peng, X.: Rssi-based localization through uncertain data mapping for wireless sensor networks. IEEE Sens. J. **16**(9), 3155–3162 (2016)
7. Huang, B., Xie, L., Yang, Z.: Tdoa-based source localization with distance-dependent noises. IEEE Trans. Wireless Commun. **14**(1), 468–480 (2015)
8. Liu, X., Yin, J., Zhang, S., Ding, B., Guo, S., Wang, K.: Range-based localization for sparse 3-d sensor networks. IEEE Internet Things J. **6**(1), 753–764 (2019)
9. Liu, X., Xiong, N., Li, W., Xie, Y.: An optimization scheme of adaptive dynamic energy consumption based on joint network-channel coding in wireless sensor networks. IEEE Sens. J. **15**(9), 5158–5168 (2015)
10. Peng, J., Liu, X.: A malicious anchor detection algorithm based on isolation forest and sequential probability ratio testing (SPRT). In: Guo, S., Liu, K., Chen, C., Huang, H. (eds.) CWSN 2019. CCIS, vol. 1101, pp. 90–100. Springer, Singapore (2019). https://doi.org/10.1007/978-981-15-1785-3_7
11. Liu, X., Su, S., Han, F., Liu, Y., Pan, Z.: A range-based secure localization algorithm for wireless sensor networks. IEEE Sensors J. **19**(2), 785–796 (2019)
12. Garg, R., Avinash, L.V.: An efficient gradient descent approach to secure localization in resource constrained wireless sensor networks. IEEE Trans. Inf. Forensics Secur. (2012)

13. Hamidi, S., Shahbazpanahi, S.: Sparse signal recovery based imaging in the presence of mode conversion with application to non-destructive testing. IEEE Trans. Signal Process. 1 (2015)
14. Mukhopadhyay, B., Srirangarajan, S., Kar S.: Robust range-based secure localization in wireless sensor networks. In: 2018 IEEE Global Communications Conference (GLOBECOM) (2019)