sensors

MDPI

Article

An MEF-Based Localization Algorithm against Outliers in Wireless Sensor Networks

Dandan Wang, Jiangwen Wan *, Meimei Wang and Qiang Zhang

School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Xueyuan Road No.37, Haidian District, Beijing 100191, China; wangdd@buaa.edu.cn (D.W.); may@aspe.buaa.edu.cn (M.W.); youleyuanzq@buaa.edu.cn (Q.Z.)

* Correspondence: jwwan@buaa.edu.cn; Tel.: +86-10-8233-9889

Academic Editor: Leonhard M. Reindl
Received: 31 March 2016; Accepted: 26 June 2016; Published: 7 July 2016

Abstract: Precise localization has attracted considerable interest in Wireless Sensor Networks (WSNs) localization systems. Due to the internal or external disturbance, the existence of the outliers, including both the distance outliers and the anchor outliers, severely decreases the localization accuracy. In order to eliminate both kinds of outliers simultaneously, an outlier detection method is proposed based on the maximum entropy principle and fuzzy set theory. Since not all the outliers can be detected in the detection process, the Maximum Entropy Function (MEF) method is utilized to tolerate the errors and calculate the optimal estimated locations of unknown nodes. Simulation results demonstrate that the proposed localization method remains stable while the outliers vary. Moreover, the localization accuracy is highly improved by wisely rejecting outliers.

Keywords: wireless sensor networks; localization; outliers; maximum entropy principle; fuzzy set theory

1. Introduction

Wireless Sensor Networks (WSNs), the networks of sensor nodes, have been widely used in many promising applications such as condition monitoring, target tracking, and home security. Precise localization plays an important role in WSNs localization systems. From the viewpoint of localization systems, there are two types of sensor nodes in WSNs. Anchor nodes, also known as beacon nodes, can obtain their location directly by using manual placement or Global Positioning System (GPS); unknown nodes, also known as regular nodes, derive their locations through localization methods. Up to now, most existing localization algorithms of WSNs could be classified as either range-based localization [1,2] or range-free localization [3,4]. Range-based localization algorithms use absolute point-to-point range measurements (distance or angle) to estimate unknown nodes' locations, while range-free localization algorithms depend on the contents of received messages. In this paper, the range-based localization methods are taken into consideration, since they are normally of high localization accuracy [5].

In localization methods, the calculation of unknown node's positions heavily relies on primary data, which are the distances between neighboring nodes and the position knowledge of anchors. In many applications of WSNs, sensor nodes are vulnerable to the internal or external disturbance. As a result of the inference, the measured distances and anchor positions can deviate from their true values. These inaccurate values are called outliers, including both the distance outliers and the anchor outliers. Due to the existence of outliers, the usage of such corrupted data can severely degrade the localization accuracy. Hence, the outlier detection process [6] is a necessary step to assure data quality in localization process. Up to the present, most existing outlier detection methods [7–9] simply assume that either distance or anchor position is the outlier, thus they are not comprehensive detecting methods. Furthermore, when the difference between the outlier value and the normal value is small enough, the outlier will not be detected, thus these methods will clearly be invalid. Therefore, an error-tolerant localization method is greatly needed to calculate the estimated locations of unknown nodes in the presence of undetected outliers. The error-tolerant localization method is a positioning refinement process which allows the existence of undetected outliers instead of discarding them. Through the error-tolerant localization process, the accuracy of localization will be improved efficiently.

In this paper, a novel secure localization method is developed to reject both the distance outliers and anchor outliers. Firstly, the uncertain value of the measured distances is obtained based on the maximum entropy theory in the lack of ranging error distribution. The uncertain value is served as the threshold in the membership function, which is compared with the difference between the Euclidean distance and the measured distance between every two neighboring anchor nodes. The Euclidean distance is calculated by the coordinates of the two anchors while the measured distance is obtained by the range-based methods. Secondly, a trust evaluation model is constructed based on the fuzzy set theory. In the trust evaluation model, a membership function is used to calculate the mutual trust values of anchor nodes. Through the data fusion, the trust value of each anchor node is obtained, and the lower trust value nodes are discarded. Finally, the Maximum Entropy Function (MEF) method is used to calculate the optimal estimated locations of unknown nodes by using the trustable data. Simulations demonstrate that the outliers can be detected effectively and the localization method can achieve high accuracy.

The rest of this paper is organized as follows. Section 2 reviews the related works. Section 3 shows a preliminary structure of sensor nodes localization system. Section 4 describes the outlier detection method. Section 5 presents the MEF method for calculating the optimal estimated locations of unknown nodes. Section 6 shows the simulation results. Finally, Section 7 concludes this paper.

2. Related Works

Generally speaking, the outliers have three anomalous causes: (1) hardware malfunctions; (2) environment interferences; or (3) malicious attacks [10]. For instance, in Time of Arrival (ToA) and Time Difference of Arrival (TDoA) systems, the transmission time or reception time of a packet can be delayed, thus resulting in distance enlargement or distance reduction [11]. In Received Signal Strength Indicator (RSSI)-based localization systems, the signal strength may be unstable or shadowed in the presence of natural or artificial interferences [12]. In malicious attacks, an attacker can increase or decrease the transmission power to make the measured distances deviate from their true values. In addition, the attacker can also capture anchor nodes to declare fake anchor positions to generate anchor outliers [13].

In practice, the existence of outliers is a fact that cannot be neglected for localization algorithms. In the case of distance outliers, a consistency check method [14] has been proposed to filter out the malicious beacon signals. The signals contain measured distance outliers on the basis of the "consistency". However, if the attackers do not revise the measured distances randomly, but make the modified distances be consistent, the strategy mentioned above will be failed under this scenario. In literature [15], linear equations are used to describe the localization problem. Hence, the norm and linear programming are applied to detect the outliers and avoid the wild measurements in the final solution. To deal with noisy and outlier ranging results, a theoretical foundation [16] has been built to identify distance outliers based on graph embeddability and rigidity theory. However, rigidity theory requires high ranging accuracy and it is computationally intensive. By applying the rigidity theory, the concept of verifiable edges [17] has been presented and the conditions for an edge to be verifiable have been derived. On this basis, the paper designs outlier detection method which explicitly eliminates ranges with large errors. However, facing with the undetected small outliers, the method would lose efficacy. In summary, based on the detection target, the literatures [14–17] mentioned above ignore the influence of anchor outliers. Therefore, these methods are one-sided.

Regarding anchor outliers, a scheme named Localization Anomaly Detection (LAD) [18] is put forward to detect malicious anchor node. The scheme attempts to perform compromise resistant localization without removing the malicious anchors. To monitor and timely detect anchor outliers in large-scale WSNs, a rule-based anomaly detection system, called RADS [19], has been proposed. In conclusion, the results of the algorithms [18,19] which are committed to eliminate anchor outliers are not comprehensive without analyzing the influence of ranges with large errors.

With respect to both outliers, an innovative modular solution [20], featuring two lightweight modules, has been developed. One is attack detection module that harnesses simple geometric triangular rules and an efficient voting technique. The other is secure localization module that computes and clusters certain reference points to estimate the coordinate of the unknown nodes. In [21], a novel algorithm, called neighbor constraint assisted distributed localization (NCA-DL), has been proposed. The method introduces the geometric constraints to detect outliers. To make localization attack-tolerant, a robust statistical method [22] has been presented. By using an adaptive least squares and Least Median Squares (LMS) position estimator, the method is capable of switching to a robust mode when the outliers exist. As a summary of the foregoing, the methods [20–22] could dispel either the anchor outliers or distance outliers. If two kinds of outliers both exit, the geometric constraints and statistical method will become invalid to filter out malicious colluding beacons or the beacon whose measured distance and coordinates change at the same time. In addition, to reduce the impact of both outliers simultaneously, Jin et al. [23] has put forward a trilateral localization algorithm for outliers suppression. However, the study of the paper focuses on the error of the algorithm itself, and discusses the stability of equations. This outliers excluded are just a portion of malicious beacons. Reference [24] designs a Beta Reputation System-based Localization (BRSL) algorithm to mainly detect and eliminate both outliers, but the Taylor-series least squares localization algorithm utilized after trust evaluation phase can't reach high accuracy.

3. Preliminaries

A WSN consists of two types of nodes, namely anchor nodes and unknown nodes. The anchor nodes are specially equipped and aware of their coordinates after deployment. The unknown nodes, whose positions are yet to be discovered, estimate their locations by measuring distances to neighboring anchor nodes. All the nodes are randomly deployed in a 2D spatial region. The communication radius of unknown or anchor nodes is R. Every node is capable of measuring the distance to any of its immediate neighbors through measurement techniques such as RSS, ToA or TDoA. As shown in Figure 1, when the unknown node N_u gets enough measured distances d'_{ui} to anchor nodes $N_i\,(i = 1, 2, \ldots, m)$, $m \geqslant 3$, a system of Euclidean equations can be set up according to the trilateration:

$$\begin{cases} (x_1 - x_u)^2 + (y_1 - y_u)^2 = d'^2_{u1} \\ (x_2 - x_u)^2 + (y_2 - y_u)^2 = d'^2_{u2} \\ \vdots \\ (x_m - x_u)^2 + (y_m - y_u)^2 = d'^2_{um} \end{cases} \tag{1}$$

where $X_u = [x_u, y_u]^T$ is N_u's coordinates that need to be estimated, $X_i = [x_i, y_i]^T$ is anchor node N_i's declared position, and d'_{ui} is the measured distance between the anchor node and the unknown node.

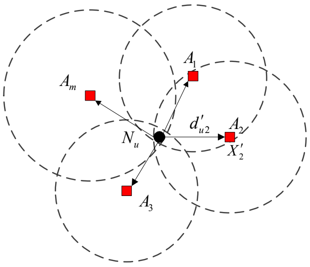


Figure 1. The measured distance d'_{u2} and declared anchor position X'_2 is normal.

Generally, X_u should be located in the intersection of m circles, of which the centers and radiuses are X_i and d'_{ui}, respectively. The smaller the intersection is, the more accurately X_u can be pinpointed. When both X_u and d'_{ui} are accurate, the X_u can be well estimated by solving Equation (1). However, if the distance outliers or the anchor outliers exist, the system would incorrectly estimate the X_u to a location that deviate far from its physical position.

The measured distance can be expressed as $d'_{ui} = d_{ui} + e_r$, and the declared anchor position can be expressed as $X'_i = \left[x_i + e_{px}, y_i + e_{py}\right]$, where e_r and e_p are the ranging error and the anchor position error, respectively. Ranging error e_r is the difference between real distance and measured distance between two sensor nodes. Position error e_p is the difference between the real position of the anchor node and the received position of the anchor node. If the distance-measuring process is disturbed, the measured distance d'_{u2} between anchor node A_2 and unknown nodeN_u, as well as the measured distance between anchor node A_2 and A_1, will be enlarged or reduced. Take the enlarged case for example. As shown in Figure 2a, the distance outlier is $d''_{u2} = d_{u2} + e_r + d_a = d'_{u2} + d_a$, where d_a is the enlarged distance, and d_{u2} is the real distance between anchor node A_2 and unknown node N_u. Meanwhile, the Euclidean distance between A_1 and A_2, i.e., $||A_1A_2||$, is different from their measured distance $||A_1A'_2||$. In addition, if the measured distance d'_{u2}, as well as the measured distance between anchor node A_2 and A_1, is reduced, the computed distance between A_1 and A_2, i.e. $||A_1A_2||$, is also different from its measured distance $||A_1A'_2||$. As shown in Figure 2b, if anchor node A_2 is malicious, the declared anchor position may deviate far from the true position. The anchor outlier is defined as $X''_2 = \left[x_2 + e_{px} + d_x, y_2 + e_{py} + d_y\right]$, where $\left[d_x, d_y\right]$ is the offset distance. In addition, the Euclidean distance between A_1 and A_2, i.e. $||A_1A'_2||$, is not equal to their measurement distance $||A_1A_2||$. Based on the above discussion, these outliers will severely degrade the localization accuracy. Therefore, it is necessary to eliminate the outliers in localization systems.

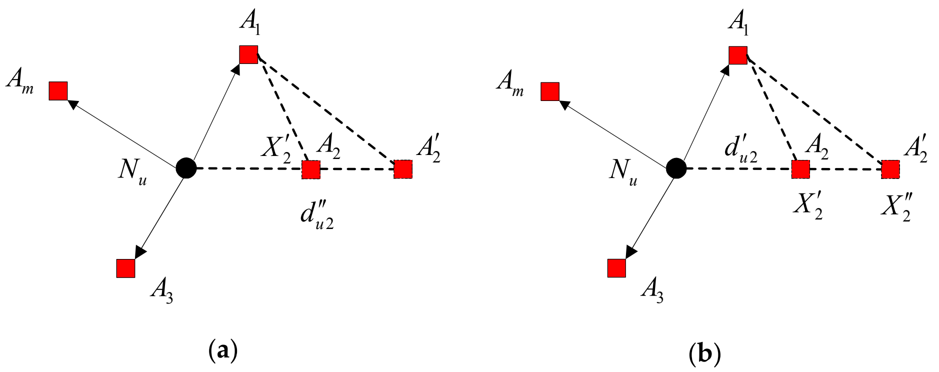


Figure 2. (**a**) The distance-measuring process is disturbed; (**b**) the declared position of the anchor node is inaccurate.

The measured distance and anchor positions exist in pairs in the localization systems. Considering two neighboring anchor nodes around an unknown node, no matter the measured distance between the two anchor nodes is enlarged or the declared position of one anchor node deviates far from the true position, the Euclidean distance will be different from the measured distance between the two anchor nodes. Therefore, no matter the distance outliers or the anchor outliers exist, or how they are generated, the difference between the Euclidean distance and the measured distance, as well as the cooperation of anchor nodes, can be utilized to detect the outliers.

4. Outlier Detection Method

In this section, firstly, the uncertain value of measured distances is calculated based on maximum entropy theory by using the ranging error priori information. Then a trust evaluation model is constructed based on the fuzzy set theory by using the uncertain value and the difference between the

Euclidean distance and the measured distance. In the trust evaluation model, the trust value of each anchor node can be obtained.

4.1. Calculation of the Entropy Uncertainty

Based on the maximum entropy theory, the uncertain value of the measured distance can be obtained by utilizing the mean and standard deviation of ranging error in this section. The information entropy $H\left(e_r\right)$ [25] of ranging error can be calculated as Formula (2),

$$H\left(e_r\right) = -\sum_{i=1}^{n} p\left(e_{ri}\right) \ln p\left(e_{ri}\right) \tag{2}$$

where $p\left(e_r\right)$ is the probability density function of ranging error.

The ranging error e_r is assumed to appear in $\left[e_{r1}, e_{r2}\right]$ with equal probability before measuring. Then, after measurement, the estimated ranging error e_r' with bias $\pm U$ is obtained, where U is the entropy uncertainty of ranging error. Hence the true value of ranging error appears in $\left[e_r' - U, e_r' + U\right]$. The information entropy of e_r' can be calculated as Formula (3).

$$H\left(e_r'\right) = -\int_{e_r'-U}^{e_r'+U} \frac{1}{2U} \ln \frac{1}{2U} \mathrm{d}e_r = \ln 2U \tag{3}$$

As we all know, the probability density function of Gaussian distribution $N\left(0, \sigma^2\right)$ is

$$p\left(e_r\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{{e_r}^2}{2\sigma^2}\right) \tag{4}$$

Hence the information entropy of ranging error can be calculated as Formula (5).

$$H\left(e_r\right) = \int_{-\infty}^{+\infty} p\left(e_r\right) \ln p\left(e_r\right) \mathrm{d}e_r = \frac{1}{2} \ln\left[2e\pi\sigma^2\right] \tag{5}$$

Let Formula (3) equal to Formula (5), the entropy uncertainty of Gaussian distribution $N\left(0, \sigma^2\right)$ can be derived as Formula (6).

$$U = \frac{\sqrt{2\pi e}}{2} \sigma = 2.07\sigma \tag{6}$$

The entropy coefficient of Gaussian distribution is 2.07. In general, let $U = k\sigma$, where k is called entropy coefficient and σ is the standard error deviation. The value of k depends on the error distribution. In this paper, the distribution of ranging error is unknown, so that k cannot be calculated directly. Definition 1 illustrates how to choose the value of k.

Definition 1. *Based on the maximum entropy principle [26] and the obtained partial information of the unknown distribution, the distribution with the maximum entropy should be selected. In all the distributions, Gaussian distribution has the maximum information entropy. Thus, the entropy coefficient of Gaussian distribution can be used to calculate the entropy uncertainty of ranging error in this paper. It is a relatively conservative but reasonable choice.*

Proof of Definition 1. Given the probability distribution of $p\left(x\right)$ and $q\left(x\right)$, the in Equation (7) can be obtained by using the inequation of $\log x \leqslant (x-1)$,

$$\int p\left(x\right) \log \frac{q\left(x\right)}{p\left(x\right)} dx \leqslant \int p\left(x\right) \left(\frac{q\left(x\right)}{p\left(x\right)} - 1\right) dx = \int q\left(x\right) dx - \int p\left(x\right) dx = 0 \tag{7}$$

Meanwhile, since $\log\frac{q(x)}{p(x)} = \log q\,(x) + \log\frac{1}{p(x)}$, $\int p\,(x)\log\frac{q(x)}{p(x)}dx$ can be changed into another format.

$$\int p\,(x)\log\frac{q\,(x)}{p\,(x)}dx = -\int p\,(x)\log p\,(x)dx + \int p\,(x)\log q\,(x)\,dx \tag{8}$$

Through Formulas (7) and (8), the in Equation (9) can be obtained.

$$H\,(p) \leqslant -\int p\,(x)\log q\,(x)\,dx \tag{9}$$

The Formula (9) is a famous conclusion that entropy of a probability distribution is always less than the relative entropy in the information theory. Only when $q\,(x) = p\,(x)$ can the equality hold in Formula (9).

Let $q\,(x) = N\left(u, \sigma^2\right)$, when $p\,(x)$ is under the given condition of mean value u and variance σ^2, then the Formula (9) can be derived as follows.

$$\begin{aligned} H\,(p) \quad &\leqslant -\int p\,(x)\log\left\{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-u)^2}{2\sigma^2}}\right\}dx \\ &= \int p\,(x)\left\{\frac{(x-u)^2}{2\sigma^2} + \log\sqrt{2\pi}\sigma\right\}dx \\ &= \frac{1}{2\sigma^2}\int p\,(x)\,(x-u)^2\,dx + \log\sqrt{2\pi}\sigma \end{aligned} \tag{10}$$

Under the limit of the mean value and the variance of $p\,(x)$:$\int p\,(x)\,(x-u)^2\,dx = \sigma^2$, the inequation of $H\,(p) \leqslant \frac{1}{2\sigma^2}\sigma^2 + \log\sqrt{2\pi}\sigma = \frac{1}{2} + \log\sqrt{2\pi}\sigma$ can be obtained. When $p\,(x) = N\left(u, \sigma^2\right)$, the equality of Formula (9) holds. Hence, the conclusion mentioned above in Definition 1 is verified.

Since the Gaussian distribution has the maximum information entropy, it has the maximum entropy coefficient. Choosing the entropy coefficient of Gaussian distribution is relatively conservative. However, based on the maximum entropy principle, the reasonable inference of the unknown distribution is the distribution which is most random and is in accord with the known information. Because this is the only choice which could be made impartially, and any other options mean that other constraints and assumptions would be added, which cannot be obtained based on the known information. Gaussian distribution is the most random distribution in nature, as we all know. Thus, the entropy coefficient of Gaussian distribution is a reasonable choice.

4.2. Foundation of the Trust Evaluation Model

After obtaining the entropy uncertainty of ranging error, the uncertain value of distance estimation can be written as $U_d = b + 2.07\sigma$, where b is the mean of ranging error. More specifically, b is calculated as Formula (11),

$$b = \frac{\sum\limits_{u=1}^{n}\left(d'_{ui} - \sqrt{\left(x'_u - x_u\right)^2 + \left(y'_u - y_u\right)^2}\right)}{n} \tag{11}$$

where d'_{ui} is the measured distance between the anchor node and the unknown node, $\sqrt{\left(x'_u - x_u\right)^2 + \left(y'_u - y_u\right)^2}$ is the computed distance between the anchor node and the unknown node, and n is the number of unknown nodes. Because the real distance between the unknown node and anchor node cannot be known in the localization process, the distance outlier needs to be detected by using the cooperation of the neighboring anchor nodes around the unknown node.

The difference between Euclidean distances and measured distances is $D_{ij} = \left|d_{ij} - d'_{ij}\right|$, where d_{ij} is the Euclidean distance between the anchor nodes and d'_{ij} (i = 1, 2, 3, . . . , m; j = 1, 2, 3, . . . , m; $i \neq j$) is the measured distance between the anchor nodes. d_{ij} is defined as $d_{ij} = \sqrt{\left(x'_i - x'_j\right)^2 + \left(y'_i - y'_j\right)^2}$ (i = 1, 2, 3, . . . , m; j = 1, 2, 3, . . . , m; $i \neq j$), where $X'_i = \left[x'_i, y'_i\right]$ (i = 1, 2, 3, . . . , m), $X'_j = \left[x'_j, y'_j\right]$ (j = 1, 2, 3, . . . , m)

are the declared coordinates of anchor nodes, and m is the number of anchor nodes. Based on the fuzzy set theory and the neighboring anchor nodes, a trust evaluation model is constructed. In the model, the fuzzy membership function is shown as Formula (12).

$$T_{ij} = \begin{cases} 1 & D_{ij} \leqslant U_d \\ 0 & D_{ij} > U_d \end{cases} \tag{12}$$

Define T_{ij} as the trust value of anchor node A_i from anchor node A_j. All these mutual trust values calculated by Formula (12) comprise the fuzzy relation matrix T.

$$T = \begin{bmatrix} 0 & T_{12} & \cdots & T_{1m} \\ T_{21} & 0 & \cdots & T_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ T_{m1} & T_{m2} & \cdots & 0 \end{bmatrix} \tag{13}$$

Then give a weight matrix W to calculate the trust value of each anchor node.

$$W = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1m} \\ w_{21} & 0 & \cdots & w_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & 0 \end{bmatrix} \tag{14}$$

where $w_{ij} = 1/m{-}1$.

Through data fusion, an evaluation result vector S can be obtained $S = W \circ T = \begin{bmatrix} s_1 & s_2 & \cdots & s_m \end{bmatrix}$, where s_i is the trust value of anchor node A_i and $s_i = \sum\limits_{i \neq j, j=1}^{m} w_{ij} T_{ij}$.

Definition 2. *Based on the majority principle, if the trust value of an anchor node is larger than 0.5, it can be concluded that this anchor node is trusted or normal. If the trust value of an anchor node is smaller than or equal to 0.5, it can be determined that the position of the anchor node is an outlier or the corresponding measured distance is an outlier. Discard the outliers and only use the trustable data to estimate the locations of unknown nodes.*

Since the presence of moving obstacles and other special situations could generate outliers temporarily, the corresponding trust values will decrease at the same time. Throughout the lifetime of the network, this kind of trust value is not credible. Therefore, all the trust values are not stored into the sensor nodes in this paper. In every localization process, the trust values will be recalculated.

5. MEF-Based Location Estimation Method

5.1. Formulation of the Localization Problem

Overall, the localization process against outliers consists of two steps. Firstly, in order to eliminate both kinds of outliers simultaneously, an outlier detection method is proposed based on the maximum entropy principle and fuzzy set theory. The first step of the localization process, named as the initial localization phase or the detecting phase is the foundation of the follow-up positioning process. Then, since not all the outliers can be detected in the detection process, the Maximum Entropy Function (MEF) method is utilized to tolerate the errors and calculate the optimal estimated locations of unknown nodes. Both steps of the localization algorithm are indispensable. Only by utilizing both steps can the localization accuracy be highly improved.

In a word, the detection method mentioned above should be applied to eliminate distance and anchor outliers in the initial localization phase. After the detecting phase, the unknown nodes utilize all their multihop communication anchor nodes to estimate their coordinates. Based on the Formula (1) in Section 3, the nodes localization problem is shown as Formula (15),

$$\min F\left(X_u\right) = \min \sum_{i=1}^{m} \left|f_i\left(X_u\right)\right| = \min \sum_{i=1}^{m} \left|\sqrt{\left(x_i - x_u\right)^2 + \left(y_i - y_u\right)^2} - d_i'\right| \tag{15}$$

where $X_u = [x_u, y_u]^T$ is the coordinate of the estimated unknown node N_u in the localization, $X_i = [x_i, y_i]^T$ is the anchor node N_i's declared position, and d_i' is the measured distance between the anchor node N_i and the unknown node N_u.

From Formula (1), $f_i\left(X_u\right)$ is assumed equal to zero. Due to the presence of errors and outliers, $f_i\left(X_u\right)$ is not equal to zero actually. By obtaining the minimum sum of $f_i\left(X_u\right)$ (i = 1, 2, 3, . . . , m), the impact of the comprehensive error on the localization will be minimized. Therefore, the estimated coordinate of unknown node with the minimum sum of $f_i\left(X_u\right)$ (i = 1, 2, 3, . . . , m) can be as the optimal estimated coordinate in the localization.

Note that $F\left(X_u\right)$ is a non-smooth function and is difficult to be minimized from Formula (15). Therefore, the MEF method, which is the least biased estimate possibility on the given information and mainly used to solve the non-smooth minimum optimization problem [27], is used to estimate the locations of unknown nodes in this paper. Using the MEF method, $F\left(X_u\right)$ can be changed into the entropy function $F_p\left(X_u\right)$, which is smooth and obtained by the following formula,

$$F_p\left(X_u\right) = \frac{1}{p}\sum_{i=1}^{m} \ln\left[\exp\left(pf_i\left(X_u\right)\right) + \exp\left(-pf_i\left(X_u\right)\right)\right] \tag{16}$$

where p is called the maximum entropy factor.

Based on [27], the following properties of the entropy function $F_p\left(X_u\right)$ are listed as follows.

Theorem 1. *For any estimated coordinate X_u of unknown node, (1) when $p \to +\infty$, $F_p\left(X_u\right) \to F\left(X_u\right)$; (2) For any p, $F\left(X_u\right) \leqslant F_p\left(X_u\right) \leqslant F\left(X_u\right) + (\ln m)/p$.*

Proof of Theorem 1.

(1) Given $X_u' \in R^2$, if there is a vector-valued function $V\left(X_u'\right)$ with components $v_i\left(X_u'\right) = \exp\left[f_i\left(X_u'\right)\right]$, where $1 < i < m$, then the l_p-norm of $V\left(X_u'\right)$ is.

$$\left\|V\left(X_u'\right)\right\|_p = \left\{\sum_{i=1}^{m}\left[v_i\left(X_u'\right)\right]^p\right\}^{1/p} = \left\{\sum_{i=1}^{m}\exp\left[pf_i\left(X_u'\right)\right]\right\}^{1/p} \tag{17}$$

Hence, $F_p\left(X_u'\right) = \ln\left\|V\left(X_u'\right)\right\|_p$. Consequently,

$$\lim_{p\to\infty}\left\|V\left(X_u'\right)\right\|_p = \max_{1\leqslant i\leqslant m} v_i\left(X_u'\right) = \exp\left[F\left(X_u'\right)\right] \tag{18}$$

That is $\lim\limits_{p\to\infty} F_p\left(X_u'\right) = F\left(X_u'\right)$.

(2) With the properties of l_p-norm, for $X_u' \in R^2$, $F_p\left(X_u'\right)$ is a monotonically decreasing function in terms of p, hence

$$\begin{aligned} 0 \leqslant F_p\left(X_u'\right) - F\left(X_u'\right) &= \ln\left\|V\left(X_u'\right)\right\|_p - \ln\left[\exp F\left(X_u'\right)\right] \\ &= (1/p)\ln\sum_{i=1}^{m}\exp\left\{p\left[f_i\left(X_u'\right) - F\left(X_u'\right)\right]\right\} \\ &\leqslant (1/p)\ln m \end{aligned} \tag{19}$$

The theorem mentioned above describes the relationship between entropy function $F_p\left(X_u\right)$ and original function $F\left(X_u\right)$ when p changes. $F_p\left(X_u\right)$ converges to $F\left(X_u\right)$ point wisely on X_u, as p tends to infinity. Theoretically, under the given conditions, as long as p is sufficiently large, the error between the optimal solution of $F\left(X_u\right)$ and the optimal solution of $F_p\left(X_u\right)$ can be made arbitrarily small. However in terms of numeral calculations, when p is fairly large, the value of entropy function $F_p\left(X_u\right)$ is overflow. Therefore, in case of the overflow, Equation (16) is transformed into the following modus.

$$F_p\left(X_u\right) = \sum_{1}^{m} \left|f_i\left(x_u\right)\right| + \frac{1}{p}\sum_{1}^{m} ln\left[1 + \exp\left(-2p\left|f_i\left(x_u\right)\right|\right)\right] \tag{20}$$

Derivation steps as follows.

$$\begin{aligned} F_p\left(X_u\right) &= \tfrac{1}{p}\sum_{1}^{m} ln\left[e^{pf_i\left(x_u\right)} + e^{-pf_i\left(x_u\right)}\right] \\ &= \sum_{1}^{m} ln\left[e^{pf_i\left(x_u\right)} + e^{-pf_i\left(x_u\right)}\right]^{\frac{1}{p}} \\ &= \sum_{1}^{m} ln\left[e^{p\left|f_i\left(x_u\right)\right|}\left(1 + e^{-2p\left|f_i\left(x_u\right)\right|}\right)\right]^{\frac{1}{p}} \\ &= \sum_{1}^{m} \left|f_i\left(x_u\right)\right| + \tfrac{1}{p}\sum_{1}^{m} ln\left[1 + e^{-2p\left|f_i\left(x_u\right)\right|}\right] \end{aligned} \tag{21}$$

Hence, summarizing all results, the nodes localization problem can be described as Formula (22) when $p \to +\infty$,

$$\min F_p\left(X_u\right) = \min\left\{\sum_{1}^{m} \left|f_i\left(x_u\right)\right| + \frac{1}{p}\sum_{1}^{m} \ln\left[1 + \exp\left(-2p\left|f_i\left(x_u\right)\right|\right)\right]\right\} \tag{22}$$

where X_u is the estimated unknown node coordinate in the localization, and p is called the maximum entropy factor. Through minimizing the entropy function $F_p\left(X_u\right)$, the estimated coordinate of the unknown node can be regarded as the optimal estimated coordinate in localization.

5.2. MEF-Based Localization Process

After removing the detected outliers, the MEF-based method, which has good error tolerance and calculation accuracy, is used to estimate the locations of unknown nodes in this paper. Meanwhile, it can also rapidly converge to the global optimal value by only iterating twice or three times. Based on the above discussion, the following definition can be concluded about the localization process:

Definition 3. *The entropy function $F_p\left(X_u\right)$ is the overall approximation to the localization function $F\left(X_u\right)$. In the localization systems, by minimizing the entropy function $F_p\left(X_u\right)$ and increasing p, the minimum $F\left(X_u\right)$ can be indirectly obtained under certain accuracy after several iterations. Thus the optimal estimated locations of unknown nodes are obtained.*

The detailed procedures of the MEF-based method for estimating the optimal locations of unknown nodes are presented in Table 1.

Table 1. The Maximum Entropy Function (MEF)-based method.

```
set maximum entropy factor $p = 10$, multiple (iteration step length) $l$ = 3, threshold $\varepsilon$ = 1e-6
calculate the lower limit of the unknown node's coordinate $L_l = \left[max\left(x'_a - d'_a\right), max\left(y'_a - d'_a\right)\right]^T$
calculate the upper limit of the unknown node's coordinate $L_u = \left[min\left(x'_a + d'_a\right), min\left(y'_a + d'_a\right)\right]^T$
calculate the initial coordinate of unknown node $X_u^{(0)} = (L_l + L_u)/2$
while 1
    minimize $F_p\left(X_u^{(j)}\right)$ and get the next iterative coordinate $X_u^{(j+1)}$ $(j = 0, 1, 2, \ldots)$
    //determine whether $X_u^{(j+1)}$ is the optimal solution
    if $\left|F_p\left(X_u^{(j+1)}\right) - F\left(X_u^{(j+1)}\right)\right| \leqslant \varepsilon$
        get the optimal estimated coordinate $X_u^{(j+1)}$
        break
    end if
    change the iterative number: $j = j + 1$
    change the maximum entropy factor: $p = l \times p$
end while
```

6. Performance Evaluation

In this section, simulation results are presented and discussed. For all of the simulations, the sensor nodes are uniformly distributed in a 150 m × 150 m square field. We assume a fixed transmission range R = 30 m for both anchor nodes and unknown nodes. The measured distance of the sensor nodes consists of two sections. One is the real distance between two nodes and the other is the measurement error. The measurement error obeys a Gaussian distribution with the mean of 0 and the variance of 1. Thus, the ranging error is set $e_r \sim N(0, 1)$. The distance outliers, which are the measured distance attacked or disturbed by external factors, can be described as $d'' = d'(1 + \alpha)$, where d' is the measured distance without attacks or disturbance, and α is the disturbed distance percentage. In each simulation, the sensor nodes of the network are deployed 100 times to compute the average localization accuracy. The default parameters of the simulation are shown in Table 2.

Table 2. Default simulation parameters.

Parameters	Values
Network size	150 m × 150 m
Number of sensor nodes	150
Percent of anchor nodes	30%
Communication radius (R)	30 m
Hop count	2

LMS [22] and BRSL [24] are used to compare with proposed localization method. They are both aimed at solving the problem of locating the unknown nodes in the presence of outliers. LMS is an outlier tolerance method and BRSL is an outlier detection and elimination method. Compared with these two different methods, the advantage of our method is revealed clearly in the simulations. The Average Localization Error (ALE) using in the experiment is calculated as Formula (23).

$$ALE = \frac{\sum\limits_{u=1}^{n} \sqrt{\left(x'_u - x_u\right)^2 + \left(y'_u - y_u\right)^2}}{nR} \tag{23}$$

where n is the number of unknown nodes, R is the network communication radius.

In the initial localization phase, no matter the measured distance between an anchor node and unknown node or the declared anchor position is outlier, the detection result is the same. Hence, the scenarios in which distance outliers are only considered are simulated. In the outlier detection phase of the simulation, based on the maximum entropy principle and fuzzy set theory, if the trust value of an anchor node is smaller than or equal to 0.5, it can be determined that the corresponding measured distance is an outlier. Discard the outliers and only use the trustable data to estimate the locations of unknown nodes. If the trustable data left are not enough to estimate the locations of unknown nodes, the information of the neighboring unknown nodes which have been located are used. In this section, all simulations are executed in MATLAB.

6.1. Impact of the Number of Distance Outliers

Figure 3 shows the ALE of our localization method and the compared methods under different numbers of distance outliers. In this simulation, set α = 50% and ranging error $e_r \sim N(0,1)$. Simulation results show that all the detected percent of distance outliers are almost equal to 100%. With the increase of number of outliers, the ALE of LMS rises obviously, while that of MEF and BRSL remain stable, which declares that our localization is robust to the variation of distance outliers. Meanwhile, under the same numbers of distance outliers, our method can greatly improve the average localization accuracy than BRSL.

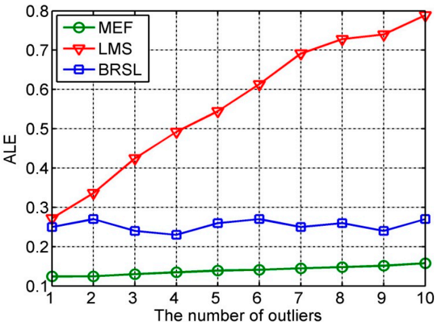
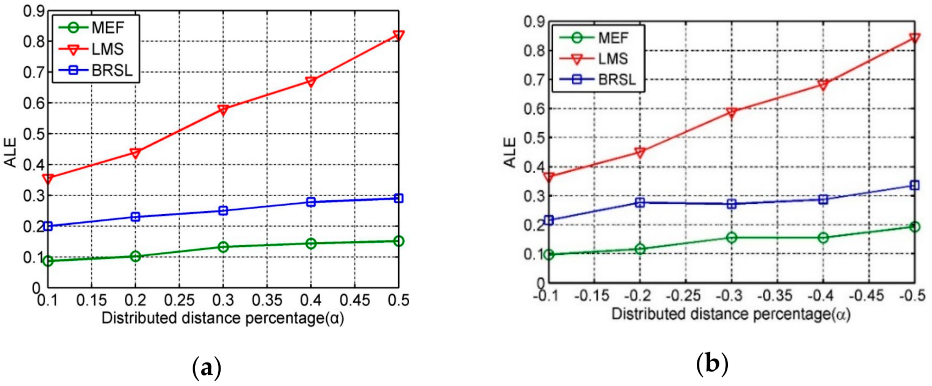


Figure 3. Average Localization Error (ALE) under different numbers of distance outliers.

6.2. Impact of Disturbed Distance Percentage

Figure 4 presents the ALE of our localization method and the compared methods under different disturbed distance percentage. In this simulation, set the number of distances outliers as 10 and ranging error $e_r \sim N(0,1)$. In this case, the detected percent is increased to almost 100% when $\alpha = \pm 50\%$. As shown in Figure 4, no matter when $\alpha > 0$ or $\alpha < 0$, the absolute difference $|d'' - d'|$ increases with the rise of $|\alpha|$. In conclusion, the localization accuracy of our method decrease slowly under different disturbed distance percentages, which shows that our localization can effectively inhibit aggressive behaviors of malicious nodes and improve the localization accuracy of unknown nodes.

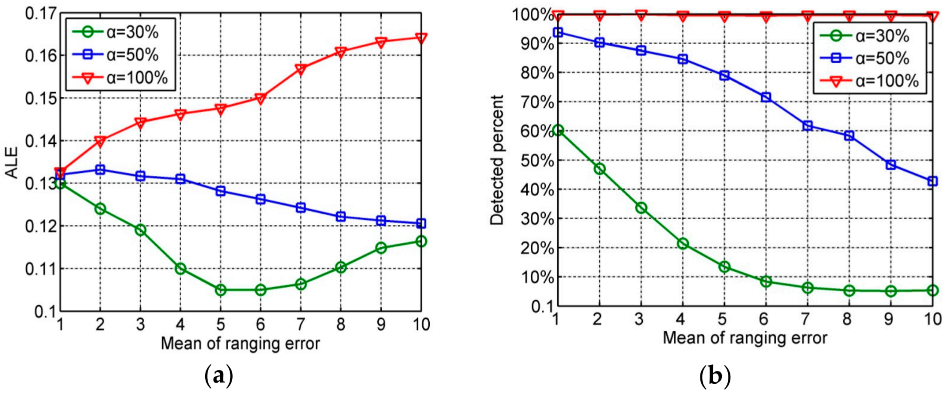


(**a**) (**b**)

Figure 4. (**a**) ALE under different disturbed distance percentages when α > 0; (**b**) ALE under different disturbed distance percentages when α < 0.

6.3. Impact of the Mean of Ranging Error

Figure 5 illustrates the performance of our method under different means of ranging error and disturbed distance percentages. In this simulation, set the number of distances outliers as 10. Because the distance outlier detection is based on the uncertain value of distance estimation, the detected percent of distance outliers is decreased with the increasing mean of ranging error when the variation in distance is small. From the Figure 5, the localization accuracy is increased with the decreasing detected percentand the increasing mean of ranging error in the curves of $\alpha = 30\%$ and $\alpha = 50\%$. Thus, it can be concluded that when the difference between the distance outlier and estimated distance is small, our localization method is error-tolerant to the undetected distance outliers; when the estimated distance is large, our localization method can detect the distance outliers.

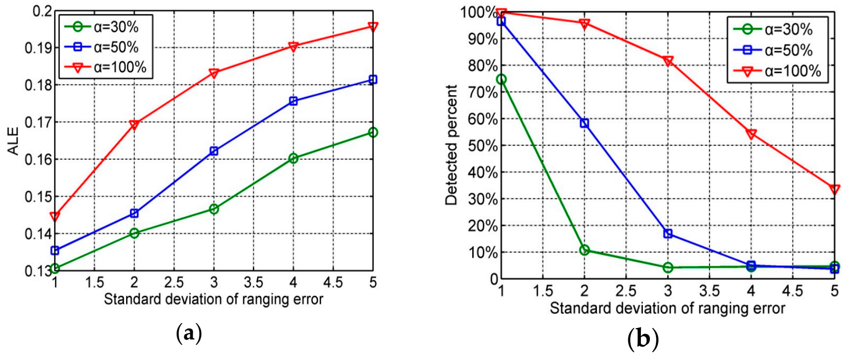


(**a**) (**b**)

Figure 5. (**a**) ALE under different means of ranging errors and disturbed distance percentages; (**b**) Detected percent of distance outliers under different means of ranging errors and disturbed distance percentages.

6.4. Impact of the Standard Deviationof Ranging Error

Figure 6 presents the performance of our localization method under different standard deviations of ranging errors and disturbed distance percentages. In the simulations, also set the number of distance outliers as 10. Compared with Figure 5, the standard deviation of ranging error has a larger impact on localization accuracy.



(**a**) (**b**)

Figure 6. (**a**) ALE under different standard deviations of ranging error and disturbed distance percentages; (**b**) Detected percent of attacked distance estimations under different standard deviation of ranging errors and distance attacked percentages.

6.5. Impact of the Iteration Step Length

Note that MEF-based algorithm contains a variable parameter, i.e. the iteration step length l, the value of which will affect the performance of the algorithm. To this end, the simulation is applied to analyze the influence of step length value on algorithm performance and explain the rationality of the parameter value selection.

In this simulation, set the number of distances outliers to 10, p = 10, α = 50% and ranging error $e_r \sim N(0,1)$. Figure 7 illustrates the change trend of the average localization error and average iteration times with l increases. The increasement of iteration step length contributes to improve the efficiency of the localization. However, on the contrary, it also results in the reduction of the localization accuracy. In summary, the small value of l will reduce the efficiency of iteration. Meanwhile, the quite large value of l will decrease the localization accuracy. It should be taken into consideration that the effect of value l to the localization accuracy and the localization efficiency, when deciding the appropriate value of l in the MEF-based iteration method.

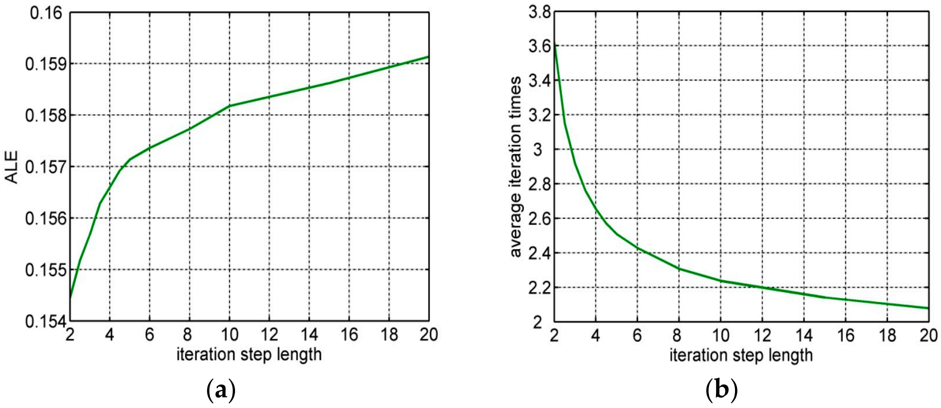


Figure 7. (**a**) ALE under different iteration step length; (**b**) The average iteration times under different iteration step length.

7. Conclusions

This paper develops an error-tolerant localization method against distance outliers and anchor outliers. First, an outlier detection method is proposed based on the maximum entropy principle and fuzzy set theory. With the cooperation of the neighboring anchor nodes of unknown node, the outliers can be detected effectively. In order to tolerate the undetected outliers and achieve high localization accuracy, MEF method is used to estimate the locations of unknown nodes. Compared with the BRSL method and LMS method, simulation results show that our localization method has higher localization accuracy.

Acknowledgments: This research was supported by the National Natural Science Foundation of China under Grant No.61371135 and Beihang University Innovation & Practice Fund for Graduate under Grant YCSJ-02-2015-07.

Author Contributions: Dandan Wang contributed to whole idea for this paper and mathematical development and paper written. Jiangwen Wan and Meimei Wang was involved in the simulations as well as drafting of the paper. Qiang Zhang was involved in data analyzing and critically reviewed the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tomic, S.; Beko, M.; Dinis, R. Distributed RSS-based localization in wireless sensor networks based on Second-order cone programming. *Sensors* **2014**, *14*, 18410–18432. [CrossRef] [PubMed]
2. Yin, J.H.; Wan, Q.; Yang, S.W.; Ho, K.C. A sample and accurate TDOA-AOA localization method using two stations. *IEEE Signal Process. Lett.* **2016**, *23*, 144–148. [CrossRef]
3. Gui, L.Q.; Val, T.; Wei, C.; Dalce, R. Improvement of range-free localization technology by a novel DV-hop protocol in wireless sensor networks. *Ad Hoc Netw.* **2015**, *24*, 55–73. [CrossRef]

4. Zhang, S.G.; Liu, X.; Wang, J.X.; Cao, J.N.; Min, G.Y. Accurate range-free localization for anisotropic wireless sensor networks. *ACM Trans. Sens. Netw.* **2015**, *11*, 1–28. [CrossRef]
5. Han, G.J.; Xu, H.H.; Doung, T.Q.; Jiang, J.F.; Hara, T. Localization algorithms of wireless sensor networks: A survey. *Telecommun. Syst.* **2013**, *52*, 2419–2436. [CrossRef]
6. Shahid, N.; Naqvi, I.H.; Qaisar, S.B. Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: A survey. *Art. Intell. Rev.* **2015**, *43*, 193–228. [CrossRef]
7. Zhong, S.; Jadliwala, M.; Upadhyaya, S.; Qiao, C. Towards a theory of robust localization against malicious beacon nodes. In Proceedings of the 27th Conference on Computer Communications, Phoenix, AZ, USA, 13–18 April 2008; pp. 2065–2073.
8. Garg, R.; Varna, A.L.; Wu, M. An efficient gradient descent approach to secure localization in resource constrained wireless sensor networks. *IEEE Trans. Inf. Forensic Secur.* **2012**, *7*, 717–730. [CrossRef]
9. Yu, N.; Zhang, L.R.; Ren, Y.J. A novel D-S based secure localization algorithm for wireless sensor networks. *Secur. Commun. Netw.* **2014**, *7*, 1945–1954. [CrossRef]
10. Xiao, Q.J.; Bu, K.; Wang, Z.J.; Xiao, B. Robust localization against outliers in wireless sensor networks. *ACM Trans. Sens. Netw.* **2013**, *9*, 1–26. [CrossRef]
11. So, H.C.; Chan, Y.T.; Chan, F.K.W. Closed-form formulae for time-difference-of-arrival estimation. *IEEE Trans. Signal Process.* **2008**, *56*, 2614–2620. [CrossRef]
12. Li, B.; Cui, W.; Wang, B. A robust wireless sensor network localization algorithm in mixed LOS/NLOS scenario. *Sensors* **2015**, *15*, 23536–23553. [CrossRef] [PubMed]
13. Boukerche, A.; Oliveira, A.B.F.; Nakamura, E.F.; Loureiroet, A.A.F. Secure localization algorithms for wireless sensor networks. *IEEE Commun. Mag.* **2008**, *46*, 96–101. [CrossRef]
14. Liu, D.G.; Ning, P.; Liu, A.; Wang, C.; Du, W.L. Attack-resistant location estimation in wireless sensor networks. *ACM Trans. Inf. Syst. Secur.* **2008**, *11*, 1–22. [CrossRef]
15. Picard, J.S.; Weiss, A.J. Bounds on the number of identifiable outliers in source localization by linear programming. *IEEE Trans. Signal Process.* **2010**, *58*, 2884–2895. [CrossRef]
16. Yang, Z. Beyond triangle inequality: Sifting noisy and outlier distance measurements for localization. *ACM Trans. Sens. Netw.* **2013**, *9*, 1–26. [CrossRef]
17. Yang, Z.; Jian, L.R.; Wu, C.S.; Liu, Y.H. Detecting outlier measurements based on graph rigidity for wireless sensor network localization. *IEEE Trans. Veh. Technol.* **2013**, *62*, 374–383. [CrossRef]
18. Du, W.L.; Fang, L.; Peng, N. LAD: Localization anomaly detection for wireless sensor networks. *J. Parallel Distrib. Comput.* **2006**, *66*, 874–886. [CrossRef]
19. Sarigiannidis, P.; Karapistoli, E.; Economides, A.A. Detecting Sybil attacks in wireless sensor networks using UWB ranging-based information. *Expert Syst. Appl.* **2015**, *42*, 7560–7572. [CrossRef]
20. Zhu, W.T.; Xiang, Y.; Zhou, J.Y.; Deng, R.H.; Bao, F. Secure localization with attack detection in wireless sensor networks. *Int. J. Inf. Secur.* **2011**, *10*, 155–171. [CrossRef]
21. Wen, L.F.; Cui, L.G.; Chai, S.C.; Zhang, B.H. Neighbor constraint assisted distributed localization for wireless sensor networks. *Math. Problem Eng.* **2014**, *2014*. [CrossRef]
22. Li, Z.; Trappe, W.; Zhang, Y.Y.; Nath, B. Robust statistical methods for securing wireless localization in sensor networks. In Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, Los Angeles, CA, USA, 24 April 2005; pp. 91–98.
23. Jin, R.C.; Che, Z.P.; Xu, H.; Wang, Z.; Wang, L.D. An RSSI-based localization algorithm for outliers suppression in wireless sensor networks. *Wirel. Netw.* **2015**, *21*, 2561–2569. [CrossRef]
24. Yu, N.; Zhang, L.R.; Ren, Y.J. BRS-based robust secure localization algorithm for wireless sensor networks. *Int. J. Distrib. Sens. Netw.* **2013**, *2013*. [CrossRef]
25. Edwards, S. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 2008; pp. 400–401.
26. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
27. Li, X.S.; Fang, S.C. On the entropic regularization method for solving min-max problems with applications. *Math. Methods Op. Res.* **1997**, *46*, 119–130. [CrossRef]