Interference Sources Localization and Communication Relationship Inference With Cognitive Radio IoT Networks

ZHAO NIU, TAO MA, NINA SHU, AND HONG SHAN

College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China

Corresponding author: Zhao Niu (bingzhenjie@163.com)

ABSTRACT With the widespread application of Internet of things (IoT), the interference problem becomes more and more serious, which results in not only the poor network performance but also the increased energy consumption of IoT nodes. Therefore, in this paper, we investigate the problem of how to locate the interference sources and infer their communication relationships in the cognitive radio IoT (CR-IoT) deployment scenario. First, we utilize MDS-MAP(P) algorithm with dynamic power control to realize cooperative self-localization of the CR-IoT nodes, which is more energy-efficient than all the IoT nodes equipped with global positioning system (GPS) receivers. Then, we propose a non-cooperative localization method to determine the inference sources with the angle of arrivals (AoAs) measured by the CR-IoT nodes. Finally, the communication relationship between interference sources can be inferred based on the association rule of signals. The network simulation results validate that the proposed methods can locate the interference sources accurately with low energy consumption and correctly infer their communication relationships, which is helpful for the interrupted CR-IoT nodes to take a specific opportunistic transmission policy to reduce their energy consumption.

INDEX TERMS The Internet of Things (IoT), green, localization, communication relationship inference.

I. INTRODUCTION

In recent years, Internet of things (IoT) has been widely used in all walks of life and became the focus of 5G and beyond wireless application [1]–[4]. Due to the scarcity of the spectrum resources, many communication devices work at close frequencies or even overlapped Industrial Scientific and Medical (ISM) frequency bands to compete for limited available channel, thus incurring complicated interference situations [5]. By applying cognitive radio (CR) capabilities to IoT, the emerging cognitive radio IoT (CR-IoT) network provides a novel solution for the sensor nodes to efficiently utilize spectrum resources. The CR-IoT network is a dedicated ad-hoc distributed IoT network, which consists of a massive number of self-configured sensor nodes with cognition capabilities to perform a coordinated task in severe environments [6], [7]. Therefore, besides traditional operations like detecting available spectrum channels on a single CR node, CR-IoT networks should be able to figure out more detailed interference map including the locations of the interference sources and their communication relationships in the CR-IoT nodes deployment area, such that smarter opportunistic spectrum channel access could be achieved to maximize the spectrum and energy efficiencies of CR-IoT networks [8], [9].

The premise of locating the interference sources is to know the positions of the CR-IoT nodes themselves. And for battery powered terminals, energy-efficient is very important which has a tremendous influence on the life cycle of the whole network [10]. In order to realize cooperative localization in the IoT, attaching a global positioning system (GPS) receiver to every sensor node is one option for determining the node location. However, this option will increase the cost and complexity of the sensor node. Moreover, it will increase the energy consumption because GPS is based on one of the energy-hungry communication technologies. A typical GPS module consumes energy in range of 143 to 166 mW [11], and it is reported in [12] that GPS incurs a serious energy cost that can drain a fully charged phone battery in 8.5 hours. Therefore, equipping all the nodes with GPS receivers is not
a practical solution. An alternative approach is to attach only a few sensor nodes with GPS receivers, which are termed as “beacons” [13]. Thus, a number of localization methods based on multidimensional scaling (MDS) could be applied, such as ranging-based MDS [14], MDS-MAP [15], MDS-MAP(P) [16] and ranging and angle of arrival (AoA)-based SMD [17]. MDS is a technique projecting high-dimensional data onto a low-dimensional space to obtain the relative coordinates of objects, which has the advantages of high localization accuracy, especially in the case of small number of beacons, and robust to ranging errors. However, MDS is a centralized algorithm and requires the network nodes being fully connected, which is difficult to be always satisfied in practice. In order to make MDS be able to work in partially connected network, Yi Shang et al. proposed the improved MDS-MAP algorithm [15], where shortest path algorithm is used when the distances between directly connected terminals are unknown. Although MDS-MAP provides a trade-off between availability and solution accuracy, it can not work well in irregularly-shaped networks, where the shortest path distance between two nodes does not correlate well with their true Euclidean distance. In addition, MDS-MAP is still a centralized algorithm with the computational complexity $O(N^3)$. Thereafter, Yi Shang et al. extend MDS-MAP to a distributed algorithm MDS-MAP(P) in [16]. The main idea of MDS-MAP(P) is to build a local map at each node of the immediate vicinity and merge these maps together to form a global map. Since MDS-MAP(P) can effectively deal with the localization problem in irregularly-shaped networks, it is suitable for the energy-efficient self-localization of the CR-IoT network.

For non-cooperative localization, Arian Shoari et al. propose to localize an uncooperative target using binary observations and without the knowledge of the propagation model in [18], where sensor nodes are assumed to be uniformly distributed on the target area around the emitter, the location of the emitter could be estimated by sample mean, minimum enclosing rectangle center, and minimum enclosing circle center respectively. However, the assumption of this method is too ideal and the localization is not accurate enough in practical scenarios. Moreover, for the non-cooperative communication relationship inference, Changkun Liu et al. have intensively explored this direction from the perspective of spectrum analysis, and proposed multiple solutions based on communication rules and clustering [19]. However, during data mining, the signal strength, i.e., the measured power is selected as a key metric of speculating the communication relationship, which is prone to being interfered by the environment.

In this paper, aiming at figuring out more detailed interference map including the locations of the interference sources and their communication relationships for achieving green CR-IoT networks, we first utilize MDS-MAP(P) algorithm with dynamic power control to realize cooperative self-localization of the CR-IoT nodes, which is more energy-efficient than all the IoT nodes equipped with GPS receivers. Then, we propose a non-cooperative localization method to determine the interference sources with the angle of arrivals (AoAs) measured by the CR-IoT nodes. Finally, the communication relationship between interference sources can be inferred based on the association rule of signals.

The remainder of this paper is as follows: Section II describes the system model of the CR-IoT network. Section III illustrates the MDS-MAP(P) algorithm with dynamic power control for self-localization of the CR-IoT nodes. Then, Section IV illustrates the solution for localization of the interference sources, and how to infer the communication relationship between interference sources are presented in Section V. Section VI shows the network simulation results, and conclusions are summarized in Section VII.

**Notations:** Upper (lower) case boldface letters are for matrices (vectors). $\mathbb{R}^{m \times n}$ denotes the $m \times n$-dimensional real Euclidean space. $\|\cdot\|$ denotes the Frobenius norm. $(\cdot)^{\mathrm{C}}$ denotes the $m$th element-wise (Hadamard) power. The operators $(\cdot)^T$ and $(\cdot)^\dagger$ denote the transpose and pseudoinverse of a matrix, respectively. The complex Gaussian distribution with mean $\mu$ and variance $\sigma^2$ is denoted by $\mathcal{CN}(\mu, \sigma^2)$. diag($\cdot$) denotes a diagonal matrix and len($\cdot$) denotes the number of elements of a set. The $N$-dimensional identity matrix is denoted by $\mathbf{I}_N$. $\mathbf{I}_N$ is a $N \times N$ matrix of ones. $\mathbf{0}_M$ is a $M \times M$ matrix of zeros.

**II. SYSTEM MODEL**

In order to facilitate the presentation, the node in the following parts is short for CR-IoT node, and the interference sources are assumed to be licensed (primary) radio stations or other unlicensed (secondary) radio stations.

The scenario with CR-IoT nodes and interference radio stations is shown in Fig. 1, the interference radio stations are surrounded by CR-IoT nodes. The blue dots indicate the CR-IoT nodes, the red squares indicate the interference radio stations, and the dotted lines indicate the wireless links.

![FIGURE 1. The scenario diagram of interference sources and the CR-IoT network.](image-url)
Assume $N$ homogeneous nodes are randomly and uniformly distributed in a two-dimensional target area to form a CR-IoT network. $\mathbf{S} = [s_1, s_2, \cdots, s_N]^{T} \in \mathbb{R}^{N \times 2}$ contains the coordinates of all the nodes, and $s_i = [x_i, y_i]^{T}$, $i = 1, 2, \cdots, N$. Every node is capable of communicating with other nodes within a radius of $r$. There are no isolated nodes or isolated subnets in the network [20]. We assume that only a small fraction of nodes, i.e., beacons are equipped with GPS receivers, all the nodes have the ranging and AoA estimation ability. It is worth noting that ranging has been realized in Semtech LoRa SX1280 chips [21]. Thus, the pairwise distances between nodes can be obtained. Moreover, there are a total of $M$ interference radio stations in the target area. $\mathbf{R} = [r_1, r_2, \cdots, r_M]^{T} \in \mathbb{R}^{M \times 2}$ contains the coordinates of all interference radio stations, and $r_j = [x_j, y_j]^{T}, j = 1, 2, \cdots, M$. It is also reasonable to suppose that the interference radio stations communicate with each other, but they cannot communicate with CR-IoT nodes. Although the nodes cannot crack the information contained in radio signal, we assume that nodes can measure the AoAs of the radio signals for identifying a specific interference radio station within a certain range.

### A. DISTANCE-BASED MODEL

In the self-localization of CR-IoT nodes, there is a distance-based model used in MDS-MAP(P). We assume that the maximum communication range of each node is $r$, if the $i$th node and the $j$th node are all in the communication range of each other, the distance between the two nodes is estimated by the ranging measurement [22]. When the $i$th node and the $j$th node are not directly connected, shortest path algorithm is used to estimate the distance of the two nodes as [23]

$$d_{ij} = \begin{cases} d_{ij}, & \text{if } d_{ij} \leq r \\ d_{h(1)} + \sum_{k=1}^{L-1} d_{h(k), h(k+1)} + d_{h(L), j}, & \text{otherwise} \end{cases}$$

(1)

where $d_{ij} = ||s_i - s_j|| + x_{ij}$ is the Euclidean distance between the $i$th node and the $j$th node, $s_{ij} \sim \mathcal{N}(\mu, \sigma^2)$ is irrelevant noise, $\sigma^2$ is the variance of the noise. $h(1), \cdots, h(L)$ are the indexes for the intermediate nodes between the $i$th node and the $j$th node which are determined by the shortest path algorithm, $L$ is the number of intermediate nodes.

### B. AoA-BASED MODEL

Assume that the interference radio station’s orientation measured by the $j$th node is $\theta_j$, $\theta_j$ takes values in the interval $[-\pi, \pi)$, and has the relation with the positions of CR-IoT nodes as

$$\theta_j = \arctan \left( \frac{y_j}{x_j} \right),$$

$$\hat{\theta}_j = \theta_j + \nu_j,$$

(2)

(3)

where $j = 1, 2, \cdots, N$, $(x, y)$ is the coordinates of the unknown interference radio station, $(x_j, y_j)$ is the coordinates of the $j$th node that could sense interference radio station’s signal, $\nu_j$ stands for AoA estimation deviation with the Gaussian distribution $\nu_j \sim \mathcal{CN}(0, \sigma^2)$.

### C. INFORMATION INTERACTION MODEL

Generally, radio stations usually adopt half-duplex communication, the quality of communication is susceptible to environmental influences. In order to improve the communication quality and prevent the loss of packets or frames during transmission, error control mechanism is usually adopted by various protocols, e.g., the stop-and-wait automatic repeat request (ARQ) mechanism [19]. The process of information interaction is shown in Fig. 2. In the procedure of the information interaction, the transmitter A sends a data frame to the receiver B. After receiving the data frame, receiver B sends an acknowledgement (ACK) frame back to A. Until the ACK frame is received at A, it does not send other data frames to B. If the receiver B has an error during the receiving process, it will send a feedback frame that requires sender A to resend the false or lost data frame. In other words, a data frame sent from radio station A causes radio station B to respond with an ACK frame or other feedback frame. It follows that the correlation between data frame and feedback frame is strong.

![Figure 2. Information interaction process between interference radio stations.](image_url)
to improve the estimation accuracy, a node estimates its position several times. However, the successive-refinement approach increases the energy consumption [13]. Hence, the stop condition for MDS-MAP(P) should be considered, and after the localization, using dynamic power control to deduce energy consumption. Instead of continuing to send the packets using the maximum transmission power to cover the entire range, the central nodes in local network adapt the maximum distance between the central node and the node that used to merge two local networks as the communication distance. In this way, when the network changes due to the location of several nodes, the update of the node location in the whole network can still be completed according to the locations of several beacons.

A. LOCAL NETWORK DIVISION

In the CR-IoT network, every node and its neighboring nodes constitute a local network. In the practical application process, every node periodically broadcasts its own identity (ID) to the surrounding area, and the node receiving the broadcast message stores the ID of the sender to build a neighbor set. For the message stores the ID of the sender to build a neighbor set.

B. CENTRAL NODE SELECTION

During the implementation of this algorithm, the set of central nodes is constructed as \( f \). Select the node with the largest degree value as the initial central node, the degree value of the \( i \)-th node is defined as \( w(i) = \text{len}(u(i)) \). If there are more than one node with the largest degree value, the node with smallest ID is selected.

When \( f(1) \) is selected, the local map \( M' = N'(f(1)) \), the rest of the local networks need to be merged into the local map. When all nodes enter the local map, the local map becomes the global map \( G \).

In the process of selecting the central node after \( f(1) \), assume the new selected central node is \( f(j) \), and \( f(j) \) should meet the following conditions:

1) The common nodes in local network \( N'(f(j)) \) and local map \( M' \) is denoted as \( I(f(j)) \). The number of nodes in \( I(f(j)) \) should not be less than 3, i.e.,

\[
\text{len}(I(f(j))) \geq 3. \tag{4}
\]

2) The difference set of nodes between local network \( N'(f(j)) \) and local map \( M' \) is denoted as \( D(f(j)) \). The maximum number of nodes that have not been divided into \( M' \), i.e.,

\[
f(j) = \arg \max_{f(j) \in N'(f(j-1))} \text{len}(D(f(j))). \tag{5}
\]

C. COOPERATIVE LOCALIZATION IN THE LOCAL NETWORK

Take the local network \( N(f(1)) \) as an example, the relative coordinates between the nodes in the local network are denoted as \( C = [c_1, c_2, \ldots , c_{l(f(1))}]^T \). Given proximity information matrix \( D = [d_{ij}] \in \mathbb{R}^{l(f(1)) \times l(f(1)),} \), the positive semi-definite symmetric matrix \( B \) is constructed as

\[
B = -\frac{1}{2}JD^\odot 2J, \tag{6}
\]

where \( J \) is given as

\[
J = I_{l(f(1))} - \frac{1}{l(f(1))} I_{l(f(1))}. \tag{7}
\]

\( I_{l(f(1))} \) and \( I_{l(f(1))} \) is a \( l(f(1)) \times l(f(1)) \) identity matrix, and \( I_{l(f(1))} \) is a \( l(f(1)) \times l(f(1)) \) matrix of ones. Next, perform eigenvalue decomposition of \( B \) as

\[
B = QAQ^T, \tag{8}
\]

where \( A = \text{diag}(\lambda_1, \lambda_2, \ldots , \lambda_{l(f(1))}), \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{l(f(1))} \) is the ordered diagonal eigenvalue matrix. For two-dimensional space, select the two largest eigenvalues to form a new diagonal matrix \( A^* = \text{diag}(\lambda_1, \lambda_2), Q^* \) indicates the corresponding eigenvector matrix, the dimensionally reduced matrix \( C \) can be expressed as

\[
C = A^{1/2}Q^*T. \tag{9}
\]

D. LOCAL NETWORK MERGING

The process of merging the local network to the local map is actually the process of coordinate fusion. Transform the coordinates of nodes in the local network to the same coordinate system of the local map, which requires the operations of translation, rotation and mirroring.

Take merging local network \( N'(f(2)) \) into local map \( M' \) as an example. Select three nodes from \( I(f(2)) \), and assume the coordinates of the three nodes in the \( M' \) are \( p_1, p_2, p_3 \), the relative coordinates of the three nodes in the local network are \( c_1, c_2, c_3 \). The steps of transformation are as follows:

1) For the \( i \)-th node, if vector \( c_i \) rotates counterclockwise \( \alpha \) angle to get \( c_i^{(1)} \), i.e., \( c_i^{(1)} = Q_1c_i, \) where \( Q_1 = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \).

2) If the vectors \( c_i^{(2)} \) and \( c_i^{(1)} \) are mirrored on the line \( L = \begin{bmatrix} \cos \beta \\ \sin \beta \end{bmatrix} \) (\( \beta \) is the angle between \( L \) and \( c_i^{(1)} \)), then \( c_i^{(2)} = Q_2c_i^{(1)}, \) where \( Q_2 = \begin{bmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{bmatrix} \).

3) If the translation from vector \( c_i^{(2)} \) to \( c_i^{(3)} \) is \( g \), then \( c_i^{(3)} = c_i^{(2)} + g \). Based on the above analysis, When \( p_1, p_2, p_3 \) and \( c_1, c_2, c_3 \) are known, the coordinates of \( l(f(2)) \) nodes in local map which is denoted as \( P = [p_1, p_2, \ldots, p_{l(f(2))}] \) can be calculated, i.e.,

\[
Q_1Q_2 = [p_2 - p_1, p_3 - p_1][c_2 - c_1, c_3 - c_1]^{-1}. \tag{10}
\]
\[ [p_4 - p_1, p_5 - p_1, \ldots, p_{i(f(2))} - p_1] = Q_1 Q_2 \left[ c_4 - c_1, c_5 - c_1, \ldots, c_{i(f(2))} - c_1 \right], \]  
(11)

\[ [p_4, p_5, p_6, \ldots, p_{i(f(2))}] = [p_1, p_1, \ldots, p_1] + Q_1 Q_2 \left[ c_4, c_5, c_6, \ldots, c_{i(f(2))} \right]. \]  
(12)

In two-dimensional space, to realize the localization of nodes in the network at least three beacons are needed. To make MDS-MAP(P) more energy-efficient, we add one beacon to test the accuracy of localization results. Assume the coordinates of the added beacon is \(s_4 = [x_4, y_4]^T\), then the distance between the real coordinates and estimated coordinates is denoted as

\[ p^\Delta = \|s_4 - \hat{s}_4\| = \sqrt{(x_4 - \hat{x}_4)^2 + (y_4 - \hat{y}_4)^2}. \]  
(13)

The threshold for localization accuracy is denoted as \(\zeta\). When \(p \leq \zeta\), the process of cooperative localization stops. Note that the average localization error of GPS could come to 4.9 m [24], so we set \(\zeta = 4.9\) in this paper.

The pseudocode for cooperative localization of nodes in the CR-IoT network is presented in Algorithm 1. We assume that the absolute coordinates of the node 1, 2, 3 and 4 are known and denoted as \(s_1, s_2, s_3\) and \(s_4\), respectively.

### IV. LOCALIZATION OF INTERFERENCE RADIO STATIONS WITH CR-IOT NODES

After obtaining the absolute coordinates of all nodes in the CR-IoT network, we start to calculate the coordinates of the interference radio stations. Assume CR-IoT nodes can obtain angle information by measuring the AoA of the signals from interference radio stations [25]. Then according to the angle information measured by two different nodes, by drawing an extension line which is also called Line of Bearing (LoB) [26], there is a junction point of the extension lines of the two nodes. Combining any two nodes can obtain an estimation value of the radio station’s coordinates. To calculate the coordinates of the radio stations, using mean method to all estimation values is a relatively simple way, however, it causes a large error.

Here, we adopt a least squares (LS) method to estimate the location of interference radio stations based AoA information [26]. Assume \(n\) nodes can obtain AoA from the \(m\)th radio station. All measured AoA information is denoted as \(\theta^{(m)}, \theta^{(m)} = [\theta_1^{(m)}, \theta_2^{(m)}, \ldots, \theta_{n}^{(m)}]\). For the AoA information measured by the \(i\)th node \(\theta_i^{(m)} = \theta_i^{(m)} + \nu_i^{(m)}\), where \(\theta_i^{(m)}\) is the real angle, \(\nu_i^{(m)}\) is the measurement error. The equation between AoA and the measured coordinates \(\hat{s}_i = [\hat{x}_i, \hat{y}_i]^T\) of the \(i\)th node is expressed as

\[ -\hat{x}_i \sin \theta_i^{(m)} + \hat{y}_i \cos \theta_i^{(m)} = \begin{bmatrix} -\sin \theta_i^{(m)} & \cos \theta_i^{(m)} \end{bmatrix} r_m. \]  
(14)

For all the \(n\) nodes, the following over-conditioned system is obtained as

\[ \begin{bmatrix} -\hat{x}_1 \sin \theta_1^{(m)} + \hat{y}_1 \cos \theta_1^{(m)} \\ \vdots \\ -\hat{x}_n \sin \theta_n^{(m)} + \hat{y}_n \cos \theta_n^{(m)} \end{bmatrix} = \begin{bmatrix} -\sin \theta_1^{(m)} & \cos \theta_1^{(m)} \\ \vdots & \vdots \\ -\sin \theta_n^{(m)} & \cos \theta_n^{(m)} \end{bmatrix} r_m, \]  
(15)

which leads to the following matrix-vector notation

\[ b = H r_m. \]  
(16)

Use the least squares method to solve the problem in equation (16), the solution is defined as

\[ r_m = (H^T H)^{-1} H^T b = H^T b. \]  
(17)

The pseudocode for localization of radio stations is presented in Algorithm 2.
Algorithm 2 Localization of Radio Stations.

Input:
\[ \theta' = \{\theta'^{(1)}, \theta'^{(2)}, \ldots, \theta'^{(M)}\} \]

Output:
\[ \tilde{\mathbf{R}} = [\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_M]^T \]

1: \( \hat{\mathbf{R}} \leftarrow \emptyset \)
2: for \( i = 1 : M \) do
3: \( \mathbf{H} \leftarrow \emptyset \)
4: \( \mathbf{b} \leftarrow \emptyset \)
5: \( n \leftarrow \text{len}(\theta'^{(i)}) \)
6: for \( j = 1 : n \) do
7: \( h_j \leftarrow [-\sin\theta'^{(i)}_j \cos\theta'^{(i)}_j] \)
8: \( b_j \leftarrow h_j \cdot \tilde{r}_j \)
9: \( \mathbf{H} \leftarrow [\mathbf{H}; h_j] \)
10: \( \mathbf{b} \leftarrow [\mathbf{b}; b_j] \)
11: end for
12: \( \tilde{r}_j \leftarrow \mathbf{H}^\dagger \mathbf{b} \)
13: \( \hat{\mathbf{R}} \leftarrow [\hat{\mathbf{R}}; \tilde{r}_j] \)
14: end for

V. INFERENCE OF COMMUNICATION RELATIONSHIP

Communication relationship refers to the relationship that the communication entities in the network exchange data on a specific application or service platform. In the process of radio communication, it usually follows certain communication rules. In order to achieve the inference of the communication relationship under the condition that the communication information cannot be cracked, start from the perspective of communication rules, according to the acquired location information, the spectrum analysis of the received signal is completed, and then the communication relationships between the radio stations are inferred through the correlation of the information frame in time.

A. ANALYSIS OF COMMUNICATION RULE

According to the analysis of information interaction model in Section II-C, due to the error control mechanism, a data frame sent from transmitter usually cause receiver to respond with an ACK frame or other feedback frame. The length of the data frame and feedback frame is related to specific protocol, and the length after coding is positively correlated with the length of channel frame. Moreover, the length of channel frame directly affects the duration of signal in burst communication. Therefore, the length of data frame and feedback frame is positively correlated with the duration of signal in burst communication. Under the condition of fixed transmission rate, the data frame length range can be roughly inferred by analyzing the duration of the burst signal, so as to establish the frame time series.

A sample signal waveform of data frame and ACK frame is shown in Fig. 3 [27], radio station A and B communicate with each other, the red rectangle represents the data or ACK frame transmitted from A to B, the yellow rectangle represents the data or ACK frame transmitted from B to A.

The length of data frame is longer than the length of ACK frame obviously.

Each data frame appears as a burst signal on the spectrum. As shown in Fig. 4, the green rectangles represent the duration of the burst signals. And the duration of burst signals are denoted as \( T = \{T_1, T_2, \ldots, T_i, \ldots\} \). The blue rectangles represent the idle time of the channel. The duration of the burst signal and the idle time of the channel are all random in length.

In the process of spectrum analysis, the method proposed in the paper [19] is used as a reference, set the scan cycle \( \delta' = 1/3w' \) to scan the spectrum signal, where \( w' \) is the frequency hopping period of signals. If a signal is continuously transmitted during the scan cycle, the corresponding value in the time series is set to the ID of the radio station, otherwise it is set to 0. \( T' = \{T'_1, T'_2, \ldots, T'_M\} \) indicates the time series converted from the spectrum monitoring data for all \( M \) radio stations in the network. Assume the ID of radio station A is 1, segment of time series \( T'_1 \) is shown in Fig. 5.

B. INFERENCE OF COMMUNICATION RELATIONSHIP

Association rules reflect the interdependence and correlation between one entity and other entities, it is an important
technology of data mining that used to mine the correlation between valuable data items from a large amount of data. The association rules are simplified in the process of inferring the communication relationship. The steps of inference is as follow: first, the time series of different radio stations are merged, and then the string matching is performed using the Boyer-Moore (BM) algorithm, finally, it is determined whether there is a communication relationship between the two radio stations by comparing the support count with the threshold.

For the radio stations in the network, select two as comparison objects, the time series constructed based on the transmitted signals of the two radio stations will be merged, the radio station’s ID is used as the value corresponding to the scan cycle of signal transmission in the spectrum analysis. Assume the ID of radio station A is 1, the ID of radio station B is 2, then the time series for A and B is $T_1$ and $T_2$, respectively. The merged result of $T_1$ and $T_2$ is denoted as $U$. For the specific scan cycle, if the value is not zero in $T_1$, the value of $T_1$ is retained, otherwise the value of $T_2$ is retained. When the value in $T_1$ and $T_2$ are all not zero at the same scan cycle, set the value as $-1$. Under normal circumstances, $-1$ does not appear when the two radio stations are communicating with each other during specific period, because the same frequency table is used between radio stations in the same network, and ARQ mechanism is used to avoid collisions. So the section of $-1$ is skipped directly in the process of communication relationship inference. A sample of merged time series is shown in Fig. 6.

![FIGURE 6. Mergence of time series.](image)

Each character in the time series $U$ corresponds to a transaction item. A set contains zero or more transaction items is called an itemset. $ee000qq$, $ee00qq$, and $ee0qq$ are all defined as itemset $X$. If itemset $X$ appears in $U$ which could also be called as transaction, then $U$ is said to contain itemset $X$. When judging whether itemset $X$ is in $U$, set $ee000qq$, $ee00qq$, and $ee0qq$ as the search term, where $e$ and $q$ indicates the ID of different radio stations. The specific meaning of $X$ is that within a specific period, the two radio stations show strong correlation on the signal. Use BM algorithm to realize the match of the search term and time series, and $qq$ is used as a good suffix. In the process of match, good suffix rules are applied to match firstly, if the good suffix matches successfully, the match continues from the tail. A sample process of good suffix rule is shown in Fig. 7.

![FIGURE 7. Good suffix rule.](image)

(2) If there is no bad character in the search term, move directly after the bad character in the time series.

An important property of an itemset is its support count $\delta(X)$, $\delta(X)$ indicates the number of times $X$ occurs in $U$. Set support count threshold to 2, which means in the period, if $X$ appears more than twice in $U$, there exists communication relationship between the two radio stations.

The pseudocode for inference of communication relationship is presented in Algorithm 3, $F = \{F_1, F_2, \cdots, F_M\}$ is the spectrum monitoring data, where $F_i = \{t^{(i)}_s,t^{(i)}_p\}$, $i = 1, 2, \cdots, M$, and $t^{(i)}_s$ indicates the start time of the signal, $t^{(i)}_p$ indicates the stop time of the signal. $A = [a_{ij}] \in \mathbb{R}^{M \times M}$ is the adjacency matrix of the interference radio stations.

**Algorithm 3 Inference of Communication Relationship.**

**Input:**

$F = \{F_1, F_2, \cdots, F_M\}$

**Output:**

$A = [a_{ij}] \in \mathbb{R}^{M \times M}$

1: $A \leftarrow 0_M$
2: for $i = 1 : M$ do
3: $t'_i \leftarrow t^{(i)}_p - t^{(i)}_s$
4: convert $t'_i$ to time series $T_i$
5: end for
6: $\delta = 0$
7: for $i = 1 : M - 1$ do
8: for $j = i + 1 : M$ do
9: $U \leftarrow$ Merge $T_i$ and $T'_j$
10: BM algorithm match $X$ in $U$
11: if $X$ in $U$ then
12: $\delta \leftarrow \delta + 1$
13: if $\delta > 2$ then
14: $a_{ij} \leftarrow 1$
15: $a_{ij} \leftarrow 1$
16: else
17: continue
18: end if
19: else
20: continue
21: end if
22: end for
23: end for

**VI. SIMULATION EVALUATION**

The network simulator EXata is adopted in our simulations. The size of target area is configured as $(1 \text{ km} \times 1 \text{ km})$. The initial maximum communication range $r$ of the nodes is set to 230 m, and the corresponding connectivity (average number
of neighbor nodes) of the nodes is 7. The inference range of the interference radio stations is set to 400 m. The network scenario consisting of CR-IoT nodes and interference radio stations is shown in Fig. 8. The other parameters of CR-IoT nodes and interference radio stations are listed in Table 1.

![Figure 8. Experiment scenario in EXata.](image)

TABLE 1. Parameters of the CR-IoT nodes and radio stations.

| Parameters     | CR-IoT node | Radio station |
|----------------|-------------|---------------|
| Number         | 50          | 3             |
| Routing protocol | OSPF        | OSPF          |
| Data link protocol | ALOHA      | ARQ           |
| Frequency      | 902-928 MHz | 902-928 MHz   |
| Business type  | CBR         | CBR           |

Because MDS can perform localization only when the network is fully connected, so only MDS-MAP and MDS-MAP(P) combined with the AoA data are used to perform the localization of the radio stations, and the localization results are compared and analyzed during the experiment. The mean method and least squares (LS) method are used for non-cooperative localization respectively. Localization error is adopted to measure the performance of each method, which is denoted as $E$ and calculated by

$$E = \frac{\sum_{i=1}^{M} \| r_i - \hat{r}_i \|}{M}, \quad (18)$$

where $M$ is the number of radio stations, $r_i$ is the real location of the $i$th radio station, $\hat{r}_i$ is the localization result of the $i$th radio station. The standard deviation of range and angle estimations are $\sigma_r^2$ and $\sigma_\theta^2$, respectively. We assume the ranging error of the nodes does not exceed 0.5 m, that is, $3\sigma_r = 0.5$, the angle estimations error does not exceed $2^\circ$, that is, $3\sigma_\theta = 2$. Based on the obtained localization error results, the average and median of localization error are selected to draw graphs according to the change of the connectivity. For each value of connectivity, the experiment was conducted $10^4$ times. The average of localization error changed with the connectivity is shown in Fig. 9, and the median of localization error changed with the connectivity is shown in Fig. 10. The abscissa is the connectivity of the nodes in the network, by increasing the communication range of the nodes, the connectivity of the nodes can be increased. The ordinate is the magnitude of $E$. Experimental results show that the localization error $E$ obtained by using mean method is much larger than the one obtained using LS method.

![Figure 9. Average of localization error changed with connectivity.](image)

![Figure 10. Median of localization error changed with connectivity.](image)

Analyze the localization results of the interference radio stations obtained by using MDS-MAP, MDS-MAP(P) combined with LS method respectively, the experimental results obtained is shown in Fig. 11, the abscissa is the connectivity of the nodes in the network, the ordinate is the magnitude of $E$, fences bound the middle 90 percent while boxes bound the middle 70 percent. Horizontal black lines locate
the population medians. It can be seen from the figure that MDS-MAP does not perform well for localization with nodes are randomly and uniformly deployed. MDS-MAP(P) can effectively deal with the randomness that occurs during deployment, and has good robustness. When the connectivity of the nodes is 14, the communication range of the nodes is 345 m, the average of \( E \) is 3.94 m, and the median of \( E \) is 3.82 m. The average and median of \( E \) are all less than 4.9 m.

Analyze the localization result when communication range \( r = 345 \) m. The cooperative localization result of the nodes based on MDS-MAP is shown in Fig. 12. The circles represent the real locations of the nodes and the red solid lines represent the errors of the estimated locations from the real locations. The longer the line, the larger the error is. The localization result based on MDS-MAP(P) is shown in Fig. 13. As can be seen from the figures, for the results of cooperative localization, MDS-MAP(P) is significantly better than MDS-MAP.

MDS-MAP(P) and GPS are used to localize all nodes in the whole network respectively, and the energy consumed by the whole network is counted by the energy consumption statistics function of EXata. The energy consumption ratio of MDS-MAP(P) and GPS is shown in Fig. 14, although the results of each experiment fluctuate greatly, at least 25% energy is saved by using MDS-MAP(P) than equipping all the nodes with GPS.

The average energy consumption of nodes in the network is calculated from the beginning of simulation to 300 s. In the case that 345 m was used as the maximum communication distance, whether adopt the stop condition and the dynamic power control was compared. The statistical result obtained is shown in Fig. 15. The cycle time for localization is set to 5 s, when MDS-MAP(P) performed about 10 times, the localization error is less than 4.9 m. In the subsequent simulation, the energy consumption of the nodes using of dynamic power control is lower than the nodes using maximum power.
obviously, when simulation time is about 300 s, more than 30% energy is saved. The stop condition and dynamic power control make MDS-MAP(P) more energy-efficient.

The localization result of interference radio stations is shown in Fig. 16. The blue circles indicate the real locations of the radio stations. The red crosses indicate the non-cooperative localization result of MDS-MAP combined with LS method, and the green crosses indicate the non-cooperative localization result of MDS-MAP(P) combined with LS method. The non-cooperative localization result of MDS-MAP(P) combined with LS method is more precise obviously.

Finally, using association rule analysis, the inference result of communication relationship between the radio stations in the network is shown in Fig. 17, where the red dotted line indicates there exist communication relationship between the two radio stations. The inference result is consistent with the actual situation.

VII. CONCLUSION

In this paper, we propose an overall scheme for deriving the locations of the interference sources and their communication relationships from interference signals with the CR-IoT network. We first utilize MDS-MAP(P) algorithm with dynamic power control to realize cooperative self-localization of the CR-IoT nodes, which is more energy-efficient than all the IoT nodes equipped with GPS receivers. Then, we propose a non-cooperative localization method to determine the interference sources with the AoAs measured by the CR-IoT nodes. Finally, the communication relationship between interference sources can be inferred based on the association rule of signals. The proposed scheme is meaningful for the CR-IoT network to take a smart opportunistic transmission policy to achieve good network performance and green energy consumption effect.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, “A survey on 5G networks for the Internet of Things: Communication technologies and challenges,” IEEE Access, vol. 6, pp. 3619–3647, 2018.
[2] N. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen, “Space/aerial-assisted computing offloading for IoT applications: A learning-based approach,” IEEE J. Sel. Areas Commun., vol. 37, no. 5, pp. 1117–1129, May 2019.
[3] R. Chen, Z. Tian, H. Zhou, and W.-X. Long, “OAM-based concentric spatial division multiplexing for cellular IoT terminals,” IEEE Access, vol. 8, pp. 59659–59669, 2020.
[4] R. Chen, H. Zhou, M. Moretti, X. Wang, and J. Li, “Orbital angular momentum waves: Generation, detection and emerging applications,” IEEE Commun. Surveys Tuts., vol. 22, no. 2, pp. 1–30, 2nd Quart., 2020.
[5] K. C. Wiklundh, “Understanding the IoT technology LoRa and its interference vulnerability,” in Proc. Int. Symp. Electromagn. Compat. (EMC Eur.), Sep. 2019, pp. 533–538.
[6] A. A. Khan, M. H. Rehmani, and A. Rachidi, “Cognitive-Radio-Based Internet of Things: Applications, architectures, spectrum related functionalities, and future research directions,” IEEE Wireless Commun., vol. 24, no. 3, pp. 17–25, Jun. 2017.
S. Tomic, M. Beko, and R. Dinis, “3-D target localization in wireless sensor networks,” in *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6748–6759, Aug. 2019.

C.-H. Liao, G. Zhu, D. Kuwabara, M. Suzuki, and H. Morikawa, “Multi-hop LoRa networks enabled by concurrent transmission,” *IEEE Access*, vol. 5, pp. 21430–21446, 2017.

A. Li, G. Han, and T. Ohtsuki, “Energy-efficient channel hopping protocol for cognitive radio networks,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

M. Li, N. Cheng, J. Gao, Y. Wang, L. Zhao, and X. Shen, “Energy-efficient UAV-assisted mobile edge computing: Resource allocation and trajectory optimization,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3424–3438, Mar. 2020.

K. Chen, G. Tan, J. Cao, M. Lu, and X. Fan, “Modeling and improving the energy performance of GPS receivers for location services,” *IEEE Sensors J.*, vol. 20, no. 8, pp. 4512–4523, Apr. 2020.

S. Gaonkar, J. Li, R. Choudhury, L. Cox, and A. Schmidt, “Micro-blog: Sharing and querying content through mobile phones and social participate,” in *Proc. 6th Int. Conf. Mobile Syst., Appl., Services (MOBISTCS)*, Jun. 2008, pp. 174–186.

A. M. Abu-Mahfouz and G. P. Hancke, “ALWadHA localization algorithm: Yet more energy efficient,” *IEEE Access*, vol. 5, pp. 6661–6667, 2017.

I. Borg and P. J. Groenen, “Modern multidimensional scaling: Theory and applications,” *J. Educ. Meas.*, vol. 40, no. 3, pp. 277–280, Jun. 2006.

Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, “Localization from mere connectivity,” in *Proc. 4th ACM Int. Symp. Mobile ad hoc Netw. Comput. (MobiHoc)*, 2003, pp. 201–212.

Y. Shang and W. Ruml, “Improved MDS-based localization,” in *Proc. IEEE INFOCOM*, Mar. 2004, pp. 2640–2651.

G. T. F. de Abreu and G. Destino, “Super MDS: Source location from distance and angle information,” in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2007, pp. 4430–4434.

A. Shouari and A. Seyedi, “Localization of an uncooperative target with binary observations,” in *Proc. IEEE 11th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2010, pp. 1–5.

C. Liu, X. Wu, L. Zhu, C. Yao, L. Yu, L. Wang, W. Tong, and T. Pan, “The communication relationship discovery based on the spectrum monitoring data by improved DBSCAN,” *IEEE Access*, vol. 7, pp. 121793–121804, 2019.

P. Santi, “The critical transmitting range for connectivity in mobile ad hoc networks,” *IEEE Trans. Mobile Comput.*, vol. 4, no. 3, pp. 310–317, May 2005.

A. Pages-Zamora, J. Vidal, and D. H. Brooks, “Closed-form solution for positioning based on angle of arrival measurements,” in *Proc. 13th IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2002, pp. 1522–1526.

H. Yang, G. Pan, and W. Wang, “Topology extrapolation method of battlefield backbone networks,” (in Chinese), *Commun. Countermeas.*, vol. 3, pp. 14–17, Apr. 2009.

**ZHAO NIU** was born in China, in 1990. He received the B.S. and M.S. degrees from the Electronic Engineering Institute, Hefei, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the National University of Defense Technology. His research interests include wireless network security, big data mining, and task collaborative planning.

**TAO MA** received the B.S., M.S., and Ph.D. degrees from the Electronic Engineering Institute, Hefei, China, in 2001, 2005, and 2009, respectively. He is currently an Associate Professor with the National University of Defense Technology. He has published more than 40 papers in journals and conferences. His research interests include wireless network security and system simulation.

**NINA SHU** received the B.S. and M.S. degrees from Information Engineering University, Zhengzhou, China, in 1999 and 2002, respectively. She is currently an Associate Professor with the National University of Defense Technology. She has published more than 40 papers in journals and conferences. Her research interests include wireless networks, cryptography, and information security.

**HONG SHAN** received the B.S. and M.S. degrees from the University of Electronic Science and Technology, Chengdu, China, in 1985 and 1997, respectively, and the Ph.D. degree from the College of Communication Engineering, in 1998. He is currently a Professor with the National University of Defense Technology. He has published more than 80 articles in journals and conferences. His research interests include wireless networks, information systems, and the optimization of protocols for wireless networks.