Journal of Information Hiding and Multimedia Signal Processing ©2021 ISSN 2073-4212
Ubiquitous International Volume **12**, Number **4**, December **2021**

X-strip Y-band Location Clustering Algorithm for Network Topology Inference

Jian-Liang Xu[1,2], Hua-Xun Lou[1,2], Zhe Wang[3] and Zhe-Ming Lu*[3]

[1]The 36th Research Institute of China Electronics Technology Group Corporation
Jiaxing, 314033, P. R. China

[2]Key Laboratory of Communication Information Control and Security Technology
Jiaxing, 314033, P. R. China

[3]School of Aeronautics and Astronautics
Zhejiang University
Hangzhou, 310027, China
*Corresponding Author: zheminglu@zju.edu.cn

Received August 2021;Revised October 2021
(Communicated by Zhe-Ming Lu)

Abstract. *In modern information warfare, on the battlefield where networked information systems are becoming more and more critical, if you want to attack the enemy's network, you must first obtain the structure of the enemy's network and understand its network characteristics. However, it is impossible for the enemy's network to actively tell us its topology. We can only infer its topology by detecting and intercepting the enemy's communication signals. The network topology inference method mainly solves the problem of how to infer network structure information such as the enemy's network scale, node composition, and communication relationship based on the reconnaissance and intercepted enemy battlefield wireless communication network signal information under non-collaborative conditions. This paper proposes a topology inference technique based on location clustering and the temporal connection of data frames. Among them, for the information source location clustering, this paper proposes a novel "X-strip Y-band" location clustering algorithm. The simulation results show that the location clustering algorithm proposed in this paper greatly accelerates the algorithm speed while maintaining good clustering accuracy, and can be used in real-time dynamic location clustering scenarios.*
Keywords: Networked information system, network topology inference, source location clustering, data frame, real-time dynamic location clustering

---

1. **Introduction.** Network topology is the basic premise of network structure analysis and network control [1]. In the information battlefield, if we can detect and infer the network topology of the enemy's key nodes through technical means, we can understand the opponent's network composition, structure, network communication and other important information, and at the same time, we can obtain the ability to counter opponent's cyber attacks [2]. Therefore, if you want to implement an efficient network confrontation, you must first obtain the topology map of the target network nodes. However, because the enemy's network is usually a non-collaborative network, it can only detect and intercept radio signals through unauthorized access, so it is difficult to obtain network topology information through traditional topology discovery techniques [3, 4]. The fundamental reason is that from the perspective of a third party, the non-collaborative party in wireless communication only intercepts the radio signal of the node and obtains the binary information on the node's time series (1: send a data signal or confirmation signal; 0: no signal), and without any prior information, this makes the current traditional topology discovery methods unsuitable. Due to the complexity of the electromagnetic environment, there are

serious challenges in inferring network topology from a small amount of observational data. It is still a difficult problem to reconstruct a network with random processes from a limited time series of nodes [5].

According to different applications, wireless network topology inference algorithms can be divided into the following two types: (1) Topology discovery algorithms for network management [6]. It is to discover the activity of the nodes in the network, the link status between nodes, the remaining energy of the nodes and other information, and then provide conditions for maximizing the life of the entire network system. (2) Attack-oriented topology discovery algorithm [7]. Based on the trend characteristics of the data in the wireless network application as a whole summarized to the base station, the topology of the target network is analyzed by passively monitoring the communication messages of the target network.

Obviously, we are aiming at the second method, which is to use the data in the wireless network application as a whole to summarize to the command center and then issue commands to the weapon node as the basis, and combine the communication signals obtained by reconnaissance, MAC layer communication messages, and location information methods. To analyze the topology of the target network. In wireless sensor networks, an attacker can usually infer the topology information of the target network by monitoring the communication flow and data forwarding path. The closer to the base station, the more packets that need to be forwarded, and therefore the greater the communication traffic of the network. The attacker tracks the location of the base station along the direction of data convergence according to the characteristics of packets converging to the base station in multiple hops. However, due to the large number of wireless sensor networks, there are many forwarding times in the process from the source node of the data to the base station. It is impractical to use the forwarding path of mobile tracking packets; on the other hand, an attacker can only obtain partial information of the network, and it is not enough to rely on this information to infer the network topology. In addition, the attacker can also monitor the target network cooperatively by deploying a large number of monitoring nodes. A certain number of monitoring nodes are deployed in the coverage area of the target network to form a monitoring network. Pass more monitoring nodes to passively monitor the communication situation of the target area, and analyze the target network in a centralized manner based on the characteristics of message forwarding during the route establishment process and the convergence of the message routing tree to the base station in the process of data reporting. This method can monitor the communication situation of the target network on a larger scale, and thus can better characterize the topology of the target network.

Different network topology inference methods and approaches are proposed in the literature. Some of them rely on access to the contents of data packets, which is not always feasible and may increase network overhead [8], [9]. In [10], a path inference method using routing information in data packets is proposed. Others belong to the category of network tomography and need to access the information of endpoints [11], [12]. For example, in [13], a low-complexity inference algorithm based on Kullback-Leibler (KL) divergence that requires link rate estimation is developed. Without accessing the contents of the data packet, the solution proposed in [14] uses the spectral coherence based on the Lomb-Scargle periodogram as a measure of the causal relationship between two signals. This approach relies on the concept of correlation, which in principle does not necessarily imply causality. In [15], a Bayesian non-parametric model is proposed to learn the topology of an unknown ad-hoc network, and the solution is based on the hidden semi-Markov model (HSMM) to segment node transmission activities. The different research areas that contribute to topology inference are represented by graph signal processing (GSP) applied to networks. Graph learning as an edge subset selection problem or neighborhood-based sparse linear regression is proposed in [16]. In [17], a nonlinear structural equation model is studied to detect the topological structure of the graph from the observation of the process propagating through it. In [18], a new method based on an elastic network solver [19] is proposed, which can perform well even when the data is highly correlated. However, topology inference in GSP assumes that the network is affected by the diffusion process, that is, a piece of information is propagated between all nodes.

Considering that the wireless communication network we are targeting is biased towards self-organizing networks, the above-mentioned inference technique is not entirely applicable. Based on this, this paper studies the topology inference technology based on location clustering and the temporal connection of data frames.

2. The whole Framework.

For the scene where all the source locations and signals are obtained or the enemy's communication signal is obtained by deploying reconnaissance nodes, the node identification and type determination method based on the clustering of the source location and the chain based on the time sequence relationship between the data frame and the response frame can be adopted. The path identification method solves the analysis problem of the network physical topology, and the process is shown in Fig. 1.

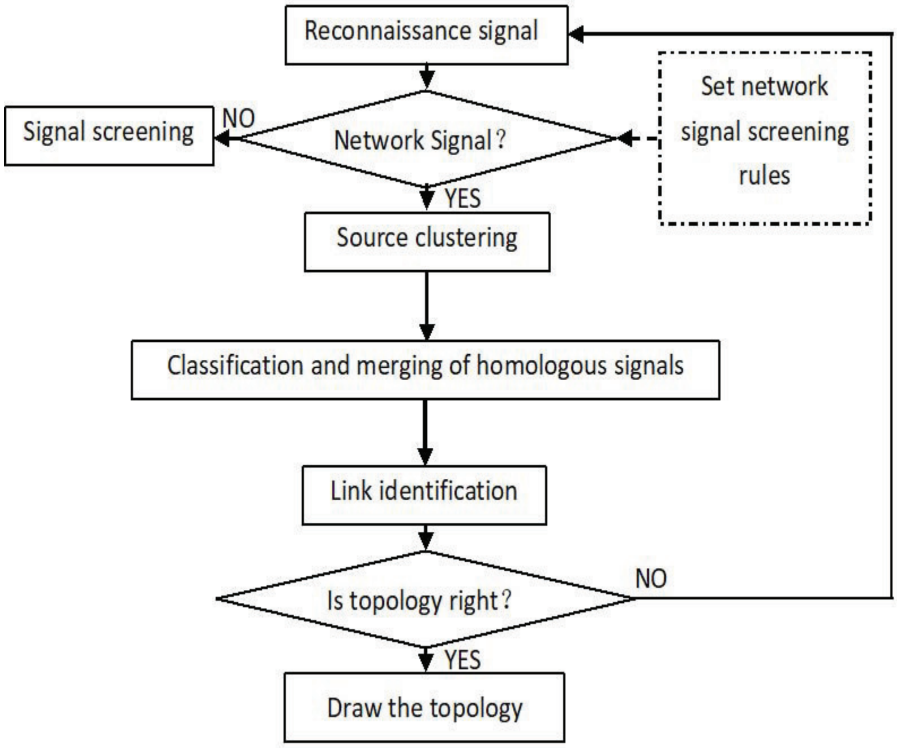

FIGURE 1. Block diagram of topology inference technology based on location clustering and time continuity of data frames

(1)Signal reconnaissance. Physical topology inference requires signal reconnaissance to provide signal appearance time, duration, source location, working frequency, modulation pattern, information rate, demodulation result, etc.

(2)Signal screening. Signal screening is a process of screening out target network signals from a large number of reconnaissance results according to specific criteria. The signal screening criteria need to be set reasonably according to the specific situation. Such as filtering according to the location of the source, filtering according to the signal pattern, filtering according to the information rate, and so on. This step is equivalent to the signal reconnaissance phase that has been completed, it can be considered that the required signals have been screened according to the information rate, that is, all the signals obtained are already network signals.

(3)Source clustering. Cluster the filtered network signals according to the location of the signal source, and the signals sent from the same location are classified into one category, with the purpose of identifying network nodes. The specific algorithm of source clustering will be introduced later.

(4)Classification and merging of homologous signals. According to modulation style, information rate, time of appearance, frequency, etc., the homologous signals are classified. After classification, the same frequency, same rate, same formula, and simultaneous signals are merged and processed, and the one with the best quality is retained. The purpose of merging is to eliminate multiple reconnaissance results on a single signal. When using reconnaissance nodes to obtain reconnaissance signals, different nodes may detect the same frequency, same speed, same pattern, and simultaneous signals.

(5)Link identification. Number the identified nodes, and analyze the wireless signals of different nodes (sources) according to the signal pattern, information rate, time relationship, communication logic, etc., identify the communication links of the network, and record the links between different nodes The frequency of use is used as the weight of the link. The specific algorithm will be introduced later.

(6)Topology confirmation. According to the results of subsequent signal reconnaissance and specific topology determination rules, the correctness of the inferred network topology is verified. If the inference result is not stable, continue the inference. Otherwise, confirm the inference result and form network topology data.

(7)Draw a topology map. According to the result data of topology inference, appropriate nodes and link symbols are selected to represent network nodes and communication links, and a topology diagram

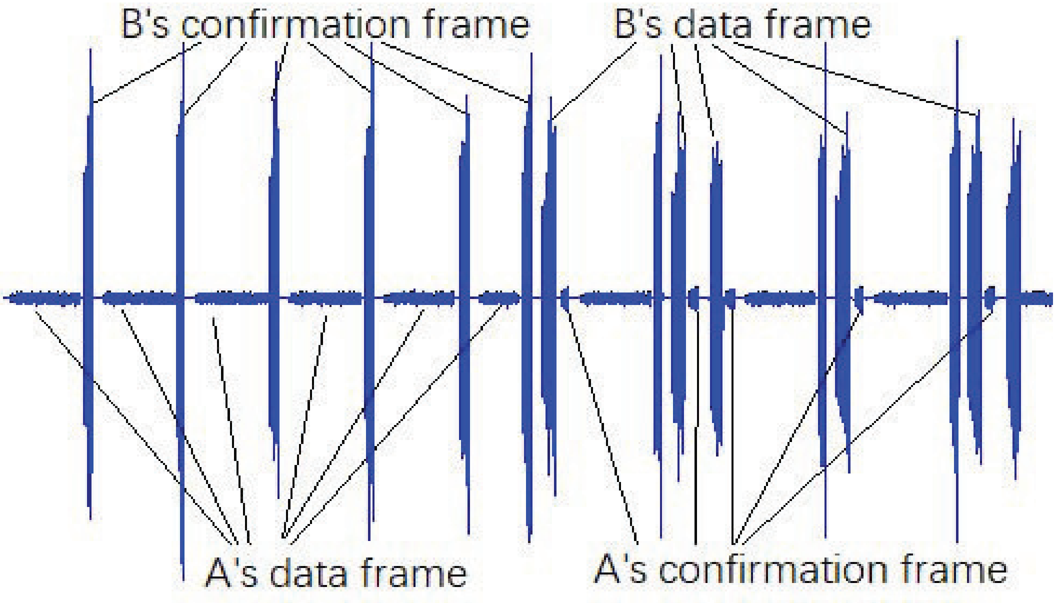

FIGURE 2. Schematic diagram of link recognition principle based on the time connection relationship between data frame and response frame and the position of the source

is drawn in combination with link weights. According to actual needs, the characteristic parameters of different links can also be marked.

According to the above process to complete the network physical topology inference, it is necessary to effectively solve the source clustering and link identification methods. For the problem of information source clustering, see the following description. The problem of link identification can be solved by a link identification method based on the time connection relationship between data frames and response frames. Wireless network communication needs to follow a specific link layer protocol. In the wireless channel, the channel condition is relatively poor, and the response to the data frame is very frequent. There is an obvious time connection relationship between the data frame and the response frame. Therefore, according to the time connection relationship between the data frame and the response frame, combined with the source location Identifying trunk links has wide applicability. Based on the time connection relationship between the data frame and the response frame, the principle of identifying the trunk link in combination with the position of the source is shown in Fig.2. This figure shows the time-domain monitoring results of a certain communication link signal, that is, the interactive communication process of the two sources of A and B. The signal from source A is relatively weak (small amplitude), and the signal from source B is relatively strong (large amplitude). For each data frame sent by A, B responds with an "confirmation" signal. At the same time, for each data frame sent by B, A also responds with an "confirmation" signal. This time connection relationship between the data frame and the response frame can be used as an important basis for link identification. The identification of the two end points of the link can be solved by locating the source. This link identification method based on the time connection relationship between the data frame signal and the response signal has nothing to do with the signal frequency. It can be used to identify the same-frequency link for receiving and sending, and it can also be used for identifying the link between the receiving and transmitting signals.

Here, we don't care about the real physical signal, but assume that the type of frame can be obtained. According to the MAC layer implementation scheme specified in the 802.11 protocol, we believe that an adhoc node must first send a control frame to solve the hidden station problem before sending a data frame. The general communication process between two nodes is as follows: First station A sends a control frame RTS, and station B responds with a control frame CTS after receiving the RTS frame. When station A receives the CTS frame and then sends the data frame DATA, station B receives the data frame from station A, and confirms with the control frame ACK. Our link inference process is based on this communication process. Algorithm idea: Under normal communication conditions, a control frame CTS that responds to node B will inevitably exist between the RTS and DATA frames sent by node A with the shortest time interval. Among all the CTS frames that appear in the interval between A's RTS and DATA, only the CTS of the responding node B is the one closest to the position of the sending node. The proof is as follows: Suppose the latest CTS frame is sent by node C which is different from B, which means that the distance between C and A is shorter than the distance between B and A, then C must be within the communication range of A. According to the definition of the 802.11 control frame RTS, after A successfully sends the RTS, all nodes in the communication range of A except B cannot send any more signals. This contradicts the assumption, so the closest CTS frame to A can only be sent from B.

3. **Location Clustering Problem and Proposed Algorithm.** To solve the problem of information source clustering, the method of information source location clustering and node feature analysis can be used. According to the positioning accuracy, set the clustering reference distance. According to the distribution of the information source and the reference distance, the number of clustering centers is determined, and the number and approximate location of the network nodes are determined according to the clustering results.

Clustering is to divide a data set into different classes or clusters according to a certain standard (such as distance criterion), so that the similarity of data objects in the same cluster is as large as possible, and the difference between data objects that are not in the same cluster is also as great as possible. That is, after clustering, data of the same type are gathered together as much as possible, and different data are separated as much as possible. At present, a large number of clustering algorithms have appeared. For specific applications, the choice of clustering algorithm depends on the type of data and the purpose of clustering. If cluster analysis is used as a tool for description or exploration, you can try multiple algorithms on the same data to discover the results that the data may reveal.

The main clustering algorithms can be divided into the following categories[20]: partitioning methods[21], hierarchical methods[22], density-based methods[23], grid-based methods[24], and model-based methods[25]. There are widely used algorithms in each category, such as the k-means clustering algorithm [21] in the partitioning method, and the agglomerated hierarchical clustering algorithm [22] in the hierarchical method. The study of clustering problems is not limited to the above-mentioned hard clustering, that is, each data can only be classified into one category. Fuzzy clustering [26] is also a branch of cluster analysis that is widely studied. Fuzzy clustering uses the membership function to determine the degree to which each data belongs to each cluster, rather than rigidly classifying a data object into a certain cluster. At present, many fuzzy clustering algorithms have been proposed, such as the famous FCM algorithm [27]. Here we mainly examine the k-means clustering algorithm [21], agglomerated hierarchical clustering algorithm[28], neural network clustering algorithm SOM [29] and fuzzy clustering FCM algorithm [27]. Finally, it is designed based on only the plane coordinates of the node and the limited mobility range of the node. A simple and feasible "X-band Y-band" clustering algorithm based on a given threshold is proposed. These clustering algorithms are introduced below:

3.1. **K-means Clustering Algorithm.** k-means is one of the most classic clustering algorithms in partitioning methods. Because of the high efficiency of this algorithm, it is widely used when clustering large-scale data. At present, many algorithms are expanded and improved around this algorithm. The k-means algorithm takes k as a parameter, divides n objects into k clusters, so that the clusters have a higher degree of similarity, but the similarity between clusters is lower. The processing process of the k-means algorithm is as follows: First, randomly select k objects, each object initially represents the average value or center of a cluster; for each remaining object, according to its distance from the center of each cluster, It is assigned to the nearest cluster; then the average value of each cluster is recalculated. This process is repeated until the criterion function converges. Usually, the square error criterion is used to make the generated clusters as compact and independent as possible, and the distance metric used is Euclidean distance, of course, other distance metrics can also be used. The process of k-means clustering algorithm [21] is as follows: input a database containing n objects and the number of clusters k; output k clusters to minimize the square error criterion. The steps are as follows:

(1) arbitrarily select k objects as the initial cluster center;

(2) repeat:

(2.1) assign each object to the most similar cluster according to the average value of the objects in the cluster;

(2.2) update the cluster and calculate the average value of objects in each cluster;

(3) until there is no longer changes.

Obviously, since the value of k cannot be determined in advance, the k-means method is not very suitable.

3.2. **Hierarchical Clustering Algorithm.** According to whether the order of hierarchical decomposition is bottom-up or top-down, hierarchical clustering algorithms are divided into agglomerative hierarchical clustering algorithms and split hierarchical clustering algorithms. The strategy of agglomerative hierarchical clustering is to first treat each object as a cluster, and then merge these atomic clusters into larger and larger clusters, until all objects are in one cluster or a certain end condition is met. Most hierarchical clustering methods belong to agglomerated hierarchical clustering methods, and they differ only in the definition of similarity between clusters. Some are based on the minimum distance between classes, some are based on the maximum distance between classes, some are based on the distance between class averages, and some are based on the average distance between classes. For example, the process of agglomerative hierarchical clustering algorithm with minimum distance [28] is as follows: (1) treat each object as one class and calculate the minimum distance between two; (2) merge the two classes with the smallest distance into a new class; (3) Recalculate the distance between the new class and all classes; (4) Repeat (2) and (3) until all classes are finally merged into one class. This algorithm needs to be given a termination condition, otherwise it is grouped into a class, which is not applicable.

3.3. **SOM clustering algorithm.** The self-organizing map neural network (SOM) was proposed by the Finnish neural network expert Professor Kohonen. The algorithm assumes that there are some topological structures or sequences in the input object, and can achieve a dimensionality reduction mapping from the input space (n-dimensional) to the output plane (2-dimensional). The mapping has the property of maintaining topological features, which is very similar to actual brain processing with strong theoretical connection. The SOM network includes an input layer and an output layer. The input layer corresponds to an high-dimensional input vector, and the output layer is composed of a series of ordered nodes organized on a 2-dimensional grid. The input nodes and output nodes are connected by weight vectors. In the learning process, we find the output layer unit with the shortest distance from it, that is, the winning unit, and then update it. At the same time, the weights of the neighboring areas are updated, so that the output nodes maintain the topological characteristics of the input vector. The algorithm flow is as follows: (1) Initialize the network, assign an initial value to the weight of each node in the output layer; (2) Randomly select the input vector from the input sample, find the weight vector with the smallest distance from the input vector; (3) Define the winning unit, the adjacent area of the winning unit adjusts the weight to make it closer to the input vector; (4) Provides new samples and trains; (5) Shrinks the neighborhood radius, reduces the learning rate, repeats until it is less than the allowable value, and outputs the clustering result. The algorithm is complicated and not very applicable.

3.4. **FCM Clustering Algorithm.** In 1965, Professor Zade of the University of California at Berkeley proposed the concept of fuzzy set for the first time. After dozens of years of development, fuzzy set theory has gradually been applied to various practical aspects. In order to overcome the shortcomings of either-or classification, cluster analysis based on fuzzy set theory emerged. The method of using fuzzy mathematics to perform cluster analysis is Fuzzy C-means (FCM). The FCM algorithm [27] is an algorithm that uses the degree of membership to determine the degree to which each data point belongs to a certain cluster. The clustering algorithm is an improvement of the traditional hard clustering algorithm. The algorithm flow is as follows: (1) Standardize the data matrix; (2) Fuzzy similarity matrix is established, and membership matrix is initialized; (3) The algorithm starts to iterate until the objective function converges to a minimum; (4) According to the results of the iteration, the final membership matrix determines the class to which the data belongs, and displays the final clustering result. The algorithm is complicated and not applicable.

3.5. **Proposed Location Clustering Algorithm.** In this paper, we propose a novel "X-Strip Y-Band" Position Clustering Algorithm. The idea of clustering algorithm of "X-strips Y-band" algorithm is shown in Fig. 3. The input of this algorithm is the node position information that has been sorted from small to large (X coordinates first, then Y coordinates), and the output is the node position information after clustering. The basic idea is to use the prior knowledge of position, that is, the movement of each node does not exceed the circle of radius R centered on the initial position, and the position between different nodes is much larger than R (at least $4R$) to speed up The speed of clustering calculation when a large amount of location data is input. The algorithm first selects the leftmost node, and then marks the stripe area with the X coordinate of this node as the left end and $X + 2R$ as the right end (as shown in the schematic diagram). In this strip area, if the difference between the Y coordinate of the newly retrieved node and the Y coordinate of all the nodes that have been recorded in the strip is greater than $2R$, it means that this node is a new unrecorded node and record it. Otherwise, it means that this node is a recorded node and associate its location information to the recorded node. After traversing a strip area,

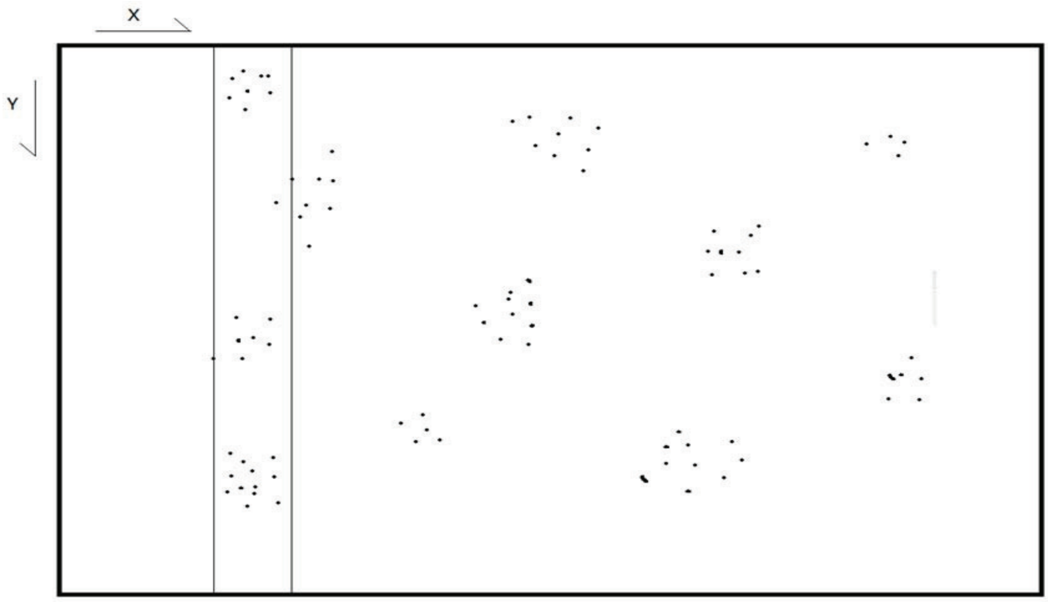

Figure 3. Schematic diagram of location clustering algorithm

calculate a new strip area according to the position of the next node, and repeat the previous process. The time complexity of this algorithm is $O(N)$, and N represents the number of location information. It should be noted that when the moving distance of a node is much smaller than the distance between nodes, there will be a small amount of position information across two X areas, so that the position information of the same node will be clustered into two distances less than different nodes of $2R$. In this case, in order to simplify the calculation, only the positions of these two nodes are averaged to merge them into one node.

4. **Experimental Results.** To evaluate the effectiveness of the proposed algorithm, we do several experiments based on simulated data. We also compare our algorithm with the K-means algorithm, agglomerative hierarchical clustering (AHC) algorithm, SOM clustering algorithm and FCM clustering algorithm. Here, our goal is to clustering 2-D coordinates. Therefore, the data used for test in our paper is a set of 2-D coordinates. We do two experiments to demonstrate the effectiveness of our algorithm. First, we use a wireless network communication scenario to demonstrate the effectiveness of our algorithm in topology inference. Second, we use simulated data to test the performance of our algorithm and compare our algorithm with other algorithms in terms of clustering performance and clustering time.

4.1. **Effectiveness of Our Algorithm in Topology Inference.** In order to demonstrate the effectiveness of our algorithm in topology inference, a wireless network communication scenario with 34 nodes is designed. The scenario is deployed within a range of $200km \times 100km$, including three outposts plus one NCS node, four sensor nodes, twenty common forwarding nodes, one command center node, Five attack platform nodes, the scenario deployment plan based on OPNET is shown in Fig. 4. The three nodes at the far left front_32, front_33, and front_34 are outposts, which are responsible for transmitting the information collected by the radar to the back-end network. In the figure, link11_gtw and the node_1 node connected to it are a simulation of the network control station (NCS), which can be regarded as a node in the simulation. The NCS is responsible for retrieving the data information of the outpost according to the fixed time slot and forwarding it to the back-end network. The link 11 protocol is used for communication between NCS and the outpost, and 802.11 and adhoc protocols are mainly used between the back-end networks. The four nodes node_2, node_3, node_4 and node_5 on the right side of node_1 are the modeling of sensor nodes. They are responsible for collecting other intelligence information in the battlefield and transmitting their own information according to the requests of other nodes. The nodes from node_6 to node_26 (except node_16) on the right side of the sensor node are models of a series of forwarding nodes in the battlefield. Such nodes do not generate information by themselves, and only forward information from other nodes. The five nodes node_27 to node_31 on the far right in Fig. 4 are modeling of the attack platform. The function is to receive the radar information from the outpost node

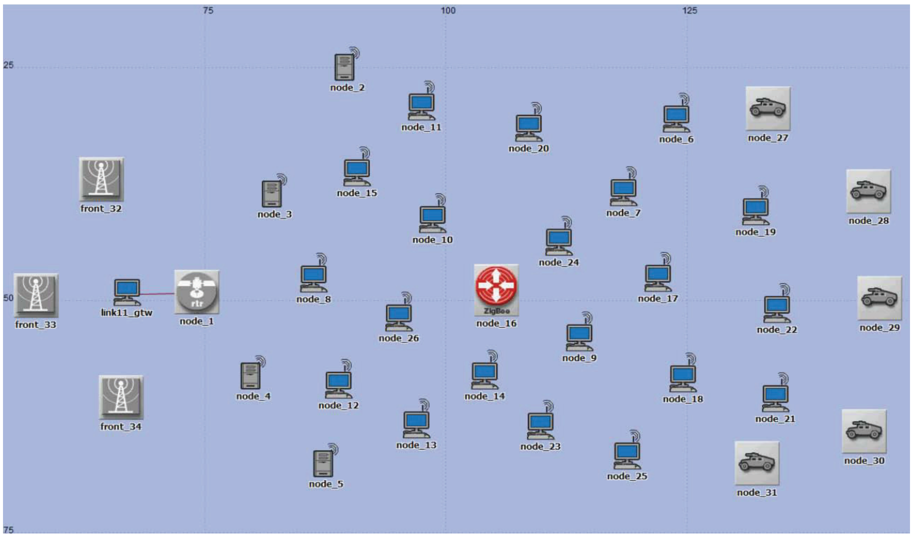

FIGURE 4. Schematic diagram of OPNET scenario deployment

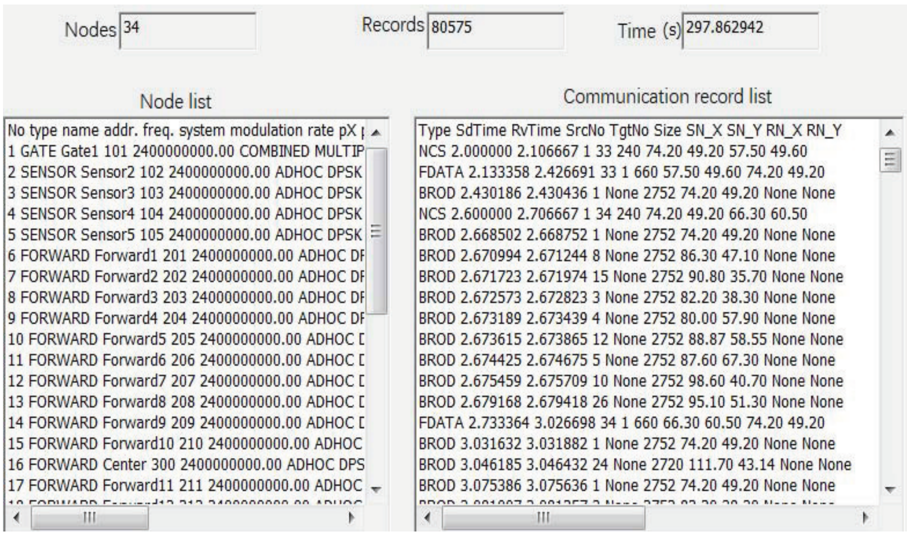

FIGURE 5. The data obtained

and the command information of the node_16 node. The node_16 node located in the center of the figure is a modeling of the battlefield command unit. It is responsible for obtaining information from sensor nodes and sending information to attack platform nodes.

The NCS node actively sends the radar data to each weapon platform node after acquiring the radar data. The business of the sensor node is simulated by the OPNET application layer database, and the commander node regularly sends database query requests to the sensor node. The information sent by the commander node to the weapon platform is also simulated by the OPNET application layer database. The weapon node sends a database query request to the commander node on time.

The network simulation time is 5 minutes, and all data link layer frame information in the network is intercepted as the original data to be analyzed. The unit of distance in the simulation scene is kilometers. The distance between outpost and NCS is 35-40 kilometers, the distance between outpost nodes is greater than 30 kilometers; the distance between nodes in the back-end network is between 10-15 kilometers, and the effective communication of the back-end network nodes The radius does not exceed 20 kilometers. The forwarding node, the commander node and the weapon platform node will all move within a radius of 200m centered on their initial position. All records are assumed to be detected, and the results of the reconnaissance are saved. An example of the data obtained is shown in Fig. 5.

Based on above data, we use our method to clustering the locations(since the locations of a node are changing a little with the time, the goal of clustering is to find the centroid location of each node). Fig. 6 shows the topology inference results of the obtained data based on the reconnaissance source location and signal pattern. It can be seen that a weighted network can be obtained with this method, and the thickness of each link reflects the weight (equal to the frequency of communication), indirectly see the importance of the link. By changing the starting record and record width, the topology can be analyzed by time period, and by clicking the comparison of the original topology at the same time segment, it can be compared with the original real topology and the topology similarity can be calculated. For example, the top half of Fig. 6 is based on the topology inference results of the entire simulation period, and the similarity with the original topology in the bottom half is 98.9

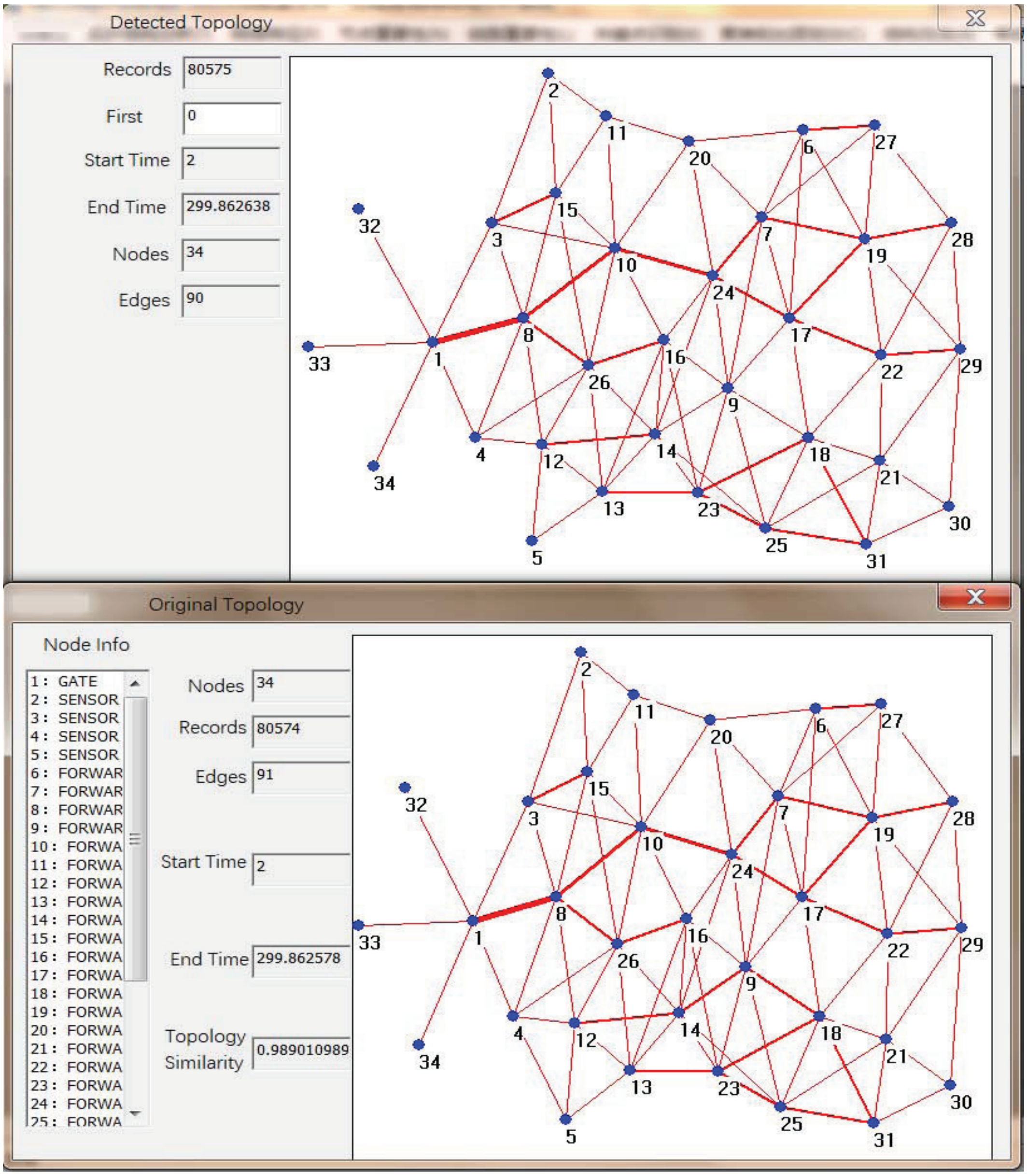

Figure 6. The topology inference results

4.2. **Comparison with Other Algorithms in Terms of Clustering Performance.** In order to demonstrate the superiority of our algorithm in clustering, we compare our algorithm with four traditional clustering algorithms, including k-means algorithm (KM in short),agglomerative hierarchical clustering (AHC) algorithm, SOM clustering algorithm and FCM clustering algorithm. We generate four datasets of 2D coordinates, their sizes are 1000, 2000, 3000, 4000 respectively, and the numbers of clusters are

Table 1. Comparisons of our scheme with other four schemes in terms of precision and time based on different datasets.

Algorithm	Dataset 1, size 1000, clusters 100			Dataset 2, size 2000, clusters 200		
	k	time	precision	k	time	precision
KM	set to 100	56.7	87.5%	set to 200	56.7	86.6%
AHC	set to 100	119.1	89.1%	set to 200	117.8	88.3%
SOM	set to 100	1.71	97.8%	set to 200	1.72	96.3%
FCM	set to 100	152.2	91.6%	set to 200	139.7	90.9%
proposed	output 100	1.0	98.2%	output 200	1.0	96.7%
Algorithm	Dataset 3, size 3000, clusters 300			Dataset 4, size 4000, clusters 400		
	k	time	precision	k	time	precision
KM	set to 300	56.7	86.1%	set to 400	56.7	85.2%
AHC	set to 300	123.2	88.5%	set to 400	125.7	87.1%
SOM	set to 300	1.68	95.2%	set to 400	1.67	94.8%
FCM	set to 300	140.9	90.4%	set to 400	145.8	90.1%
proposed	output 300	1.0	96.5%	output 400	1.0	95.1%

100,200,300,400 respectively. since traditional algorithms cannot know the actual number of clusters, thus we compare them under the condition that we have known the number of clusters. The comparison result is shown in Table 1. From this table, we can see that our algorithm outperforms other algorithms. First, our algorithm doesn't need know the number of clusters in advance. Second, our algorithm is very fast. Third, our algorithm has great clustering precision.

5. **Conclusion.** This paper proposed a novel "X-strip Y-band" location clustering algorithm. The simulation results show that, compared other traditional clustering algorithms, the location clustering algorithm proposed in this paper greatly accelerates the algorithm speed while maintaining good clustering accuracy, and can be used in real-time dynamic location clustering scenarios. Our algorithm can be used in topology inference with a perfect performance.

REFERENCES

[1] F. Guo, I. P. Brown. Simultaneous magnetic and structural topology optimization of synchronous reluctance machine rotors,*IEEE Transactions on Magnetics*, vol. 56, no. 10, pp. 1-12, 2020, Art no. 8101612.

[2] N. B. Gaikwad, H. Ugale, A. Keskar, N. C. Shivaprakash. The Internet-of-Battlefield-Things (IoBT)-based enemy localization using soldiers location and gunshot direction, *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11725-11734, 2020.

[3] Y. Liu, J. Feng, O. Simeone, J. Tang, Z. Wen, A. M. Haimovich, M. Zhou. Topology discovery for linear wireless networks With application to train backbone inauguration, *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2159-2170, 2016.

[4] R. Diamant, R. Francescon, M. Zorzi. Topology-efficient discovery: a topology discovery algorithm for underwater acoustic networks, *IEEE Journal of Oceanic Engineering*, vol. 43, no. 4, pp. 1200-1214, 2018.

[5] G. Cavraro, R. Arghandeh. Power distribution network topology detection with time-series signature verification method, *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3500-3509, 2018.

[6] A. Takada, N. Hayashi, M. Nakamura, T. Seki, K. Yamagoe. Topology discovery method using network equipment alarms, *2020 16th International Conference on Network and Service Management (CNSM)*, pp. 1-5, 2020.

[7] E. Testi, A. Giorgetti, Blind wireless network topology inference,*IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1109-1120, 2021.

[8] J. Wenli, G. Teng, J. Meiyin, L. Dan. Researching topology inference based on end-to-end date in wireless sensor networks, *Fourth International Conference on Intelligent Computation Technology and Automation*, pp. 683–686, Mar. 2011.

[9] T. Kontos, G. S. Alyfantis, Y. Angelopoulos, S. Hadjiefthymiades. A topology inference algorithm for wireless sensor networks, *IEEE Symposium on Computers and Communications*, pp. 479–484, Jul. 2012.

[10] Y. Gao, W. Dong, C. Chen, J. Bu, W. Wu, X. Liu. IPath: Path inference in wireless sensor networks,*IEEEACM Transactions on Networking*, vol. 24, no. 1, pp. 517–528, 2016.

[11] C.-K. Yu, K.-C. Chen, S.-M. Cheng. Cognitive radio network tomography, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1980–1997, 2010.

[12] Y. Vardi. Network tomography: Estimating source-destination traffic intensities from link data, *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 365–377, 1996.

[13] Y. E. Sagduyu, Y. Shi, A. Fanous, J. H. Li. Wireless network inference and optimization: Algorithm design and implementation,*IEEE Transactions on Mobile Computing*, vol. 16, no. 1, pp. 257–267, 2017.

[14] C. Partridge, D. Cousins, A. W. Jackson, R. Krishnan, T. Saxena, W. T. Strayer. Using signal processing to analyze wireless data traffic, Proceedings of the 1st ACM workshop on Wireless Security, pp. 67–76, 2002.

[15] S. Kokalj-Filipovic, C. B. Acosta, M. Pepe. Learning structural properties of wireless ad-hoc networks non-parametrically from spectral activity samples. *IEEE Global Conference on Signal and Information Processing*, pp. 1092–1097, Dec. 2016.

[16] G. Mateos, S. Segarra, A. G. Marques, A. Ribeiro. Connecting the dots: Identifying network structure via graph signal processing, *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[17] Y. Shen, B. Baingana, and G. B. Giannakis. Nonlinear structural equation models for network topology inference, *Annual Conference on Information Science and Systems*, pp. 163–168, Mar. 2016.

[18] P. A. Traganitis, Y. Shen, G. B. Giannakis. Network topology inference via elastic net structural equation models, *25th European Signal Processing Conference*, pp. 146–150, Aug. 2017.

[19] H. Zou, T. Hastie. Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society*, vol. 67, no. 2, pp. 301–320, 2005.

[20] G. Ahalya, H. M. Pandey. Data clustering approaches survey and analysis, *International Conference on Futuristic Trends on Computational Analysis and Knowledge Management*, pp. 532-537, 2015.

[21] Z.-M. Fang, X.-D. Jiang, Z.-M. Lu. Improved k-means algorithm using initialization technique based on edge-mean grid for image vector quantizer design, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no.5, pp. 1050-1057, 2018.

[22] J. Liu, B. Ni, C. Li, J. Yang, Q. Tian. Dynamic points agglomeration for hierarchical point sets learning, *2019 IEEE/CVF International Conference on Computer Vision*, pp. 7545-7554, 2019.

[23] X. Wang, D. Huang. A novel density-based clustering framework by using level set method, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 11, pp. 1515-1531, 2009.

[24] C. Tsai, S. Huang. An effective and efficient grid-based data clustering algorithm using intuitive neighbor relationship for data mining, *International Conference on Machine Learning and Cybernetics*, pp. 478-483, 2015.

[25] C. Xiong, D. M. Johnson, J. J. Corso. Active clustering with model-based uncertainty reduction, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 5-17, 2017.

[26] T. Bhowmik, I. Banerjee, A. Bhattacharya. An improved PSO based fuzzy clustering algorithm in WSNs, *IEEE 16th India Council International Conference*, pp. 1-4, 2019.

[27] J. K. Parker, L. O. Hall. Accelerating fuzzy-c means using an estimated subsample size, *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 5, pp. 1229-1244, 2014.

[28] Nisha,P. J. Kaur. Cluster quality based performance evaluation of hierarchical clustering method, The 1st International Conference on Next Generation Computing Technologies, pp. 649-653, 2015.

[29] X. Zou, H. Sun. Clustering analysis of micro-array data based on the SOM algorithm, *Ninth International Conference on Computational Intelligence and Security*, pp. 308-312, 2013.