International Journal of Innovative
Computing, Information and Control
Volume **17**, Number **3**, June **2021**

ICIC International ©2021 ISSN 1349-4198
pp. **973–989**

PRIVACY PROTECTION ALGORITHM FOR SOURCE NODE LOCATION BASED ON PHANTOM ROUTING IN THE INTERNET OF THINGS ENVIRONMENT

Yihong Li

School of Remote Sensing and Information Engineering
Wuhan University
No. 129, Luoyu Road, Wuhan 430079, P. R. China
lyhleon@whu.edu.com

Received November 2020; revised March 2021

Abstract. *Generally speaking, different source location protection strategies have played an important role in protecting location security of source nodes, and prolonged attackers' time to locate source nodes. Different strategies have different requirements for the network, and the degree of security protection for the source node location is different. In order to ensure the diversification of geographic location for phantom nodes, this paper proposes a privacy protection algorithm for source node location based on phantom routing in the Internet of Things environment. The algorithm more effectively resists attacks from strong visual attackers and strengthens the privacy protection of source location. Firstly, a network model is constructed, base stations are used to broadcast to the entire network for obtaining the hop count of each node and base station, and then a hop count table is constructed. Subsequently, two nodes away from base stations are randomly selected from the hop count table to form a set of candidate phantom nodes. And the data packet with two candidate phantom node IDs is broadcast to the whole network. Further, the phantom node positions are scattered away from source nodes by the twice selection of phantom nodes. In addition, the transmission path avoids nodes that easily cause failure paths, thereby achieving a balance of network security requirements. Finally, the simulation experiment proves that the proposed algorithm has a good privacy protection effect and is better than several comparison algorithms.*

Keywords: Source location privacy protection, Phantom routing, Internet of Things, Network security, Wireless sensor network, Attacker

1. **Introduction.** Since 21st century, with the development of perceptual recognition technology, automated information production equipment such as sensors and recognition terminals can perceive physical world in real time and accurately. At the same time, the development of network technology makes it possible to use information in the physical world. In order to realize the integration and intercommunication between physical world and information world, people have proposed and developed Internet of Things technology [1,2].

In the architecture of Internet of Things, perception recognition is the core technology of Internet of Things and the link between physical world and information world. The wireless sensor, as a device for automatically generating information at perception and recognition layer, plays a key role in the entire informatization process. In practical applications, a large number of sensors need to be deployed in the sensing area in order to obtain accurate information. The sensors form a multi-hop self-organizing network system by wireless communication, which is called a wireless sensor network. Wireless sensor

DOI: 10.24507/ijicic.17.03.973

networks have three functions: computing, communication and perception. It is mainly used for the detection and tracking of scene targets, including the acquisition of target content data and location information. Specifically it is related to industrial production, smart transportation, agricultural production, medical care, smart home, military, environment and other fields. However, wireless sensor nodes are often deployed in remote and unguarded complex environments, and wireless sensor networks generally use wireless multi-hop communication. This can easily lead to attacks by attackers. Therefore, the privacy and security issues of entire network need to be fully considered when deploying wireless sensor networks [3-5].

The privacy security of wireless sensor networks can be roughly divided into content-based privacy security and context-based privacy security [6,7]. Content-based privacy security mainly solves the problem of data privacy caused by communication between nodes. The main technologies include data encryption, data fusion and user authentication. Context-based privacy security mainly solves the problem of source location privacy security and base station location privacy security related to context. This paper will mainly study the location privacy security of source nodes in context [8]. The source node is a key node in the wireless sensor network. If it is not protected, it may bring major security risks to the monitoring target. For example, sensor nodes are deployed in the wild to monitor precious animals or are scattered on the battlefield to obtain sensitive military information. The location information of these monitoring targets is very important. Once the information is leaked, precious animals may be captured and emergency military information may be leaked [9,10]. Thus, it is of great significance to study how to strengthen the protection of location privacy for source nodes.

The following chapters of this paper are arranged as follows. Chapter 2 reviews some related research results and expounds the research motivation of this paper. The third chapter introduces the process of model building, including "network model" and "attack model". Chapter 4 introduces the implementation of the algorithm. The fifth chapter presents the simulation experiment and result analysis, which verifies the proposed algorithm and proves its effectiveness. Chapter 6 summarizes and prospects the research.

2. **Related Research.** In location privacy protection, the most widely used model is k-anonymous model. [11] proposed an anonymous algorithm based on fake location and Stackelberg game based on the structure of semi-trusted anonymous server. It introduced a pseudonym server to store user privacy separately, which effectively avoids the problem of leaking users' complete privacy information when the anonymous server is attacked.

In order to avoid communication bottleneck caused by the central server and the problem of not being completely credible, scholars have proposed a method of group collaboration to build k anonymous groups. [12] proposed a point-to-point space camouflage algorithm for user cooperation. The user formed a point-to-point group with surrounding users by single-hop or multi-hop communication, and then expanded the location area into a point-to-point anonymous group. They used anonymous groups to replace users' real locations for location service query to protect users' privacy. In [13], a method of location privacy protection without anonymity area for user collaboration was proposed in order to improve users' service quality and anonymity system performance. An anonymous group was constructed by user collaboration, and anonymous group center replaces the user's real location to initiate incremental queries, which improves the quality of service. However, the above solutions all assume that collaborative users are credible, and do not consider the untrustworthy state of real environment.

[14] used a safe summation method to solve the problem of dishonesty among collaborative users on the basis of [13]. However, complex cryptography techniques make anchor

calculation algorithms low. When there are many dishonest collaborative users, multiple recalculations can easily lead to an endless loop. In addition, considering the untrustworthy behavior of collaborative users, [15] proposed a location privacy protection method based on query fragmentation user collaboration. This method divided the request information into several fragments according to security level, and then randomly distributed them to other users in groups. Only when the request fragments of all users in group are collected will they be sent to service providers to protect the privacy of users. [16] proposed a location privacy protection scheme based on reputation incentive mechanism. They set a threshold for users, and only when users' reputation reaches the threshold, can you get help from people around you. This scheme considered the honest behavior of requesting users and cooperating users. The disadvantage was that the reputation incentive mechanism is stored on cloud servers, and it is assumed that the third-party cloud server is semi-trusted. Block chain had the characteristics of decentralization, difficulty in tampering, and incentive mechanism. Some scholars have combined block chain technology and distributed k-anonymous location privacy protection for research. [17] combined with block chain technology to improve the k-anonymity incentive mechanism. A security deposit system was designed to prevent malicious users from joining to a certain extent and increase the success rate of the anonymous zone. [18] used block chain technology for the first time to regard requesting users and cooperating users' anonymous area generation process as a transaction, which was stored in the block chain. As soon as the requesting user who divulges the location of the cooperative user and the cooperative user who provides the false location are found to have fraud, they are punished by multiple rounds of prohibiting the construction of anonymous zone. However, none of the above considers the behavior of requesting users to disclose location information of collaborative users.

[19] proposed a location privacy protection method based on Privacy Region Replacement (PRR). First, they generated a privacy zone according to the density of people and privacy requirements. Second, according to the distribution of people in privacy zones, the privacy zone was replaced with an anonymous zone. Then they calculated the coverage between anonymous areas and user query areas, and used a new query area for online query finally. [20] proposed a location privacy protection method that satisfies differential privacy constraints. It protected the privacy of location data and maximized the utility of data and algorithms in Industrial Internet of Things. Aiming at the high-value and low-density characteristics of location data, a multi-level location information tree model was established by combining practicability and privacy. Besides, the index mechanism of differential privacy was used to select data according to the access frequency of tree nodes. Finally, Laplace method was used to add noise to the access frequency of selected data. These two methods can resist stronger attackers. However, because the number and positions of injected false nodes are randomly distributed, unnecessary communication overhead is inevitably brought.

Aiming at the defects of existing source location privacy protection schemes against local traffic attackers, this paper proposes a privacy protection algorithm for source node location based on phantom routing in the Internet of Things environment. The algorithm can effectively resist attacks from attackers with strong visual capabilities and strengthen the protection of source location privacy. The main contributions of this paper are as follows.

1) Use base stations to broadcast to the entire network for obtaining the number of hops from each node to base stations and construct a hop table. Two nodes away from base stations are randomly selected from the hop count table to form a set of candidate phantom nodes and broadcast to the entire network. Source nodes randomly select a node

from the set as phantom nodes after receiving it, which ensures that the distribution of phantom nodes is diverse and has better security.

2) Two choices of phantom nodes make the phantom node positions scattered. This keeps phantom nodes away from the real source node, and at the same time does not cause a failure path from phantom nodes to sink nodes, which improves network security.

3. Model Establishment.

3.1. **Network model.** For the research of source node location privacy protection technology, the panda-hunter model [21] is usually used as shown in Figure 1. The hunter in the model has high computing power and sufficient storage space. This can locate the panda and capture it by tracking data packets hop-by-hop in the reverse direction. The network model of existing source location privacy protection scheme is basically modified based on panda-hunter model. For example, increase the number of source nodes, base stations or attackers, and change the distribution of nodes and the characteristics of nodes in the network.

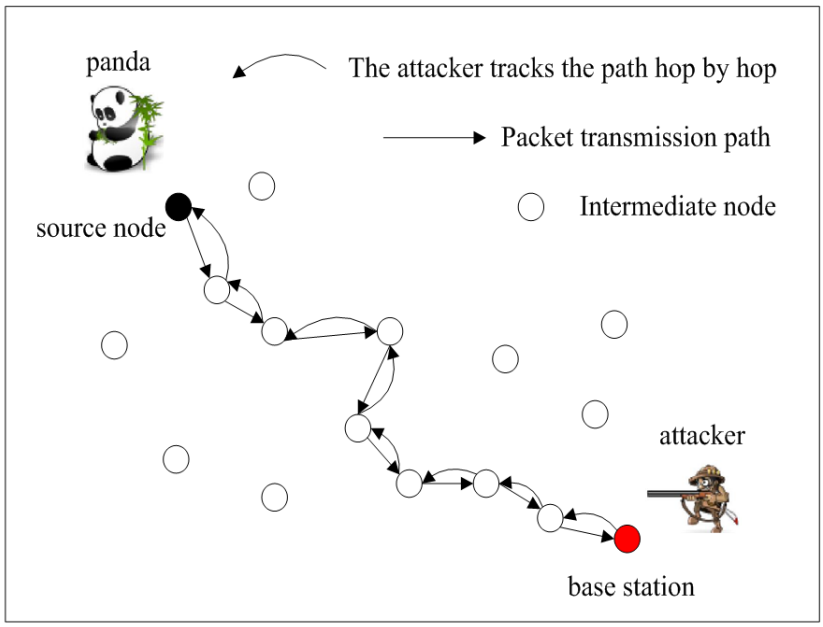


FIGURE 1. Panda-hunter model

The network model used in this program is similar to "panda-hunter model". The nodes in network are uniformly and randomly distributed in two-dimensional grid area. Nodes become source nodes after discovering the panda, and continue to send messages to sink nodes until the panda leaves [22,23]. We make following assumptions about the network model.

1) Network interconnection. The node knows its location and sink node location, and knows the relative location of its neighboring nodes. Data can be transmitted between any nodes in a single-hop or multi-hop manner. The communication radius of sensor nodes is the distance of one hop.

2) The location of source nodes is random. The monitored target appears randomly in the network. The node closest to panda monitors panda's information and becomes source nodes, and sends the panda's location information to sink nodes. However, there is only one source node in the network.

3) There is only one convergence node in the network, which is located in the center of monitoring network. The aggregation node has good computing, storage and data processing capabilities.

4) The communication between nodes is confidential. The content of data packet is encrypted. The key generation, distribution and update between nodes are beyond the scope of this paper.

A homogeneous sensor network contains N sensor nodes $\{v_i|1 \leq i \leq N\}$. Each sensor node has the same computing, storage and energy consumption resources. And each node v_i knows its own position (x_i, y_i) and sink node position (x_s, y_s).

Assuming that the sensor network is deployed in an obstacle-free flat space, the distance between sensor nodes is Euclidean distance. If the positions of node v_1 and node v_2 are (x_1, y_1) and (x_2, y_2) respectively, the distance between two points is

$$d(v_1, v_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{1}$$

In a sparse network, due to the sparse number of neighbor nodes, it is easy for attackers to locate the direct sender and receiver of data packets [24,25]. Therefore, this paper assumes that the network is densely connected. In the eyes of an external attacker, the size and format of each data packet are the same. The node ID information is encrypted. The security encryption mechanism can ensure that attackers cannot decrypt the specific content of data packet, nor can it distinguish between true and false messages [26].

Assume that the recognition module of sensors can determine the location $H(x_h, y_h)$ of attackers. We introduce an authorization mechanism in nodes to eliminate interference in the process of attacker identification. Unauthorized moving objects (with broadcast signals) are considered as attackers.

3.2. **Attack model.** In Internet of Things environment, due to the limited communication range of each sensor node, data transmission adopts a hop-by-hop transmission method. The attacker traces base station or data source according to the time correlation of data packet transmission and traffic pattern of different communication nodes. This paper considers 2 types of attackers.

1) Patient attacker. When capturing new data, they move to the sending direction of data packet. Otherwise, it has been waiting in place.

2) Curious attacker. If no data packet is received during a node's waiting time, they will walk randomly.

These two types of attackers are more typical and more representative. In practical applications, there is no clear distinction between the strength of these two attackers [27,28]. Although a curious attacker will be more flexible when he does not receive any data, a patient attacker may have greater attack capabilities than a curious attacker. For example, routing based on the shortest path, patient attackers can capture more data packets [24,29].

Assume that the attacker here has the following characteristics.

1) Local. That is, the surveillance range of attackers is its neighboring sensor nodes.

2) Passive. The attack method is monitoring and unable to control or destroy sensor nodes, and will not have any functional impact on the network.

3) Mobile. They start from sink nodes to find the location of source nodes.

The attack trajectory of attackers is shown in Algorithm 1.

Every time an attack is launched, the attacker starts from sink nodes (lines (1) $\sim$ (3)). Before source nodes are captured, every time a new data packet is captured (lines (4) and (5)), it is based on the sending angle and signal strength of received data packet. Determine the direction of location for direct sender of data packet, and move to the

Algorithm 1 The attack trajectory of attackers

```
1)  hunter = sink; //Attacker starts from sink
2)  pre_hunter = sink;
3)  next_hunter = sink;
4)  while (next_hunter ≠ source & time < Time)
5)     msg = ListenMessage();
6)     if (TimedListen() < T & IsNewMessage(msg))
7)        next_hunter = alculateImmediateSender(msg);
8)        pre_hunter = hunter;
9)        hunter = next_hunter;
10)    else if (TimedListen() ≥ T)
11)       next_hunter = ran_hunter;
12)       hunter = next_hunter;
13)       ran_hunter = GetRandomHunter(hunter);
14)    end if
15) end while
```

direct sender (lines (6) $\sim$ (9)). If the attacker does not listen to data packet within a certain period of time T (line (10)), the random walk method is used to find node that is sending data. And they continue to monitor (lines (11) $\sim$ (13)) until source nodes are found or the algorithm is not found but the time runs out. If T is relatively short, he is a curious attacker. If T is infinite, it is a patient attacker.

The attacker moves at a constant speed V_A, where $V_A \leq V_m$. V_m is the transmission speed of data packets between adjacent nodes. The attacker's monitoring range is not greater than the node's communication range, namely $D \leq r$. Among them, D and r are the listening radius of attackers and the communication radius of sensor nodes respectively.

4. Algorithm Implementation.

4.1. Privacy protection strategy for source location based on phantom routing.

The source privacy protection strategy based on phantom routing is mainly divided into four stages: phantom node selection, hop limited flooding, phantom routing and directional random routing. The strategy implementation is based on the node's knowledge of sink node location and its relationship with neighbor nodes. We initialize the network first.

4.1.1. *Network initialization.* The sink node sets initial hop value to 0. When the entire network is flooded, attackers' visible area radius V is informed to all nodes. After the node receives flooding message, it adds 1 to the hop value. Record current hop value, and broadcast the flooding message to its neighbor nodes. After the flooding is over, each node chooses the smallest one among recorded hop values, and uses it as the shortest path hop number to the sink node. Update its neighbor list. This paper considers that the nodes with same number of hops in the shortest path from sink nodes are located on a gradient, and the number of hops in the shortest path is used to represent its gradient. The symbols used in this section are shown in Table 1.

Deflection angle: the angle between the connection between common nodes and sink nodes and the connection between source nodes and sink nodes. As shown in Figure 2, the deflection angle of node P_1 is $\angle P_1 Ds$.

Failure path: The path through visible areas is called the failure path. The area where the source node of failed path is located is failed path area. The shortest path from nodes P_2 and P_3 to sink node passes through the visible area of source s. The path in shaded

TABLE 1. Symbol description

s	Source node
r	Node transmission radius
V	Viewing area radius
ΔT	Packet sending interval of source nodes
h	Flood hop count
u, v	Node
c	The number of random routing hops
H	The number of shortest routing hops from source node to sink node

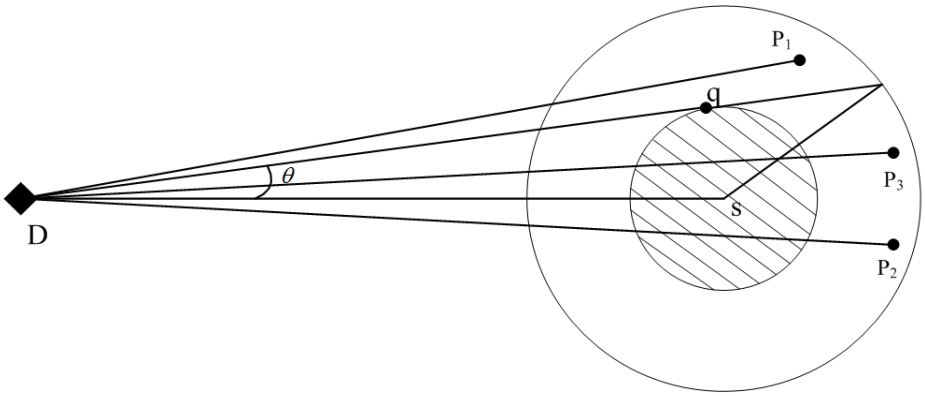


FIGURE 2. Deflection angle and failure path

area becomes failed path, and the area where nodes P_2 and P_3 are located is the failed path area.

4.1.2. *Phantom node selection.* First, base stations broadcast data packet Sink_Msg = {Base_bro, Node_ID, Coun_b, Info_Coor} by the network. Base_bro represents the message type; Node_ID represents the ID of message senders. Coun_b represents the hop count of messages, initial value is 0, and the value of Coun_b is increased by 1 each time the message is forwarded. Info_Coor stands for coordinate system information. For any intermediate node u, receiving Sink_Msg for the first time, node u adds 1 to the value of Coun_b and updates $\mathrm{Hop_{u,b}} = \mathrm{Coun\_b}$. For any Sink_Msg received, node u must store Node_ID, Info_Coor and Coun_b in its neighbor node information list. After base stations are broadcast on the entire network, each node in network knows the number of hops from base stations. In order to facilitate the subsequent selection of phantom nodes, each node needs to inform the base station of its relative position information and the number of hops. Each node replies Res_Sink = {Source_ID, Coun_b, Pos_info} response data packet to base stations. Among them, Source_ID represents the ID of node itself; Coun_b represents the number of hops; Pos_info represents the relative position information of nodes. After receiving the response data packet, base stations create a hop table with Source_ID, Coun_b and Pos_info as attributes.

After base stations end the whole network broadcast, it randomly selects two nodes P_1 and P_2 from hopTable as candidate phantom nodes. In order to ensure the diversity and effectiveness of selected phantom nodes, candidate phantom nodes need to meet the following conditions: 1) $\mathrm{Hop_{p_1,b}}$ and $\mathrm{Hop_{p_2,b}}$ are either greater than or equal to N; 2) the distance between P_1 and P_2 is greater than L. The two conditions ensure that two candidate phantom nodes are far away from base stations and are not within the visual

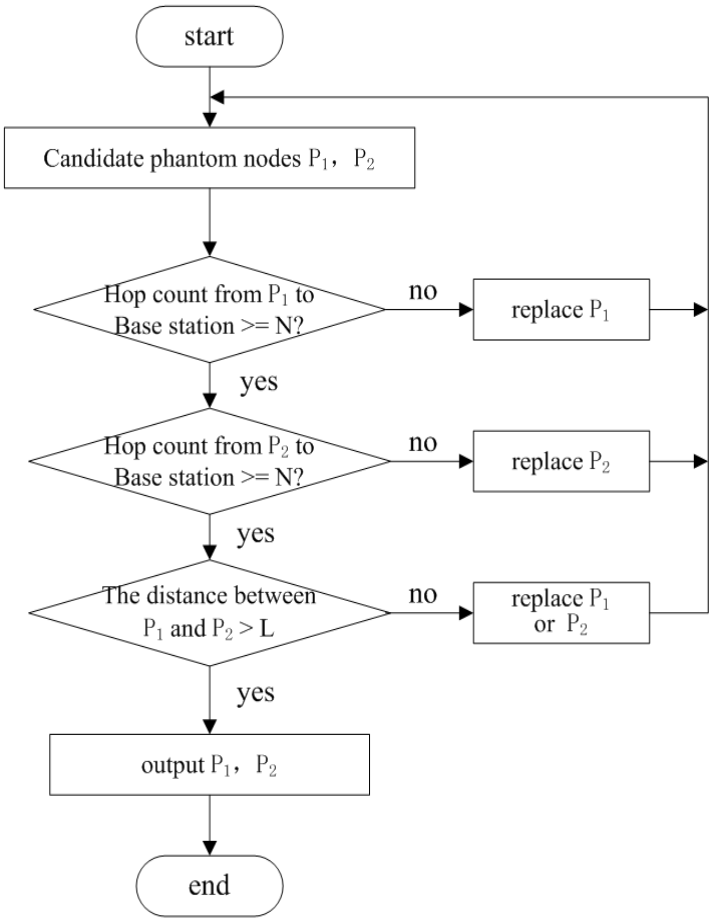


FIGURE 3. Flow chart of candidate phantom node selection

area of opponent's attackers. Once one of candidate phantom nodes is used as the source node, the other can also be used as phantom node. This ensures the effectiveness of selected candidate phantom nodes and increases the difficulty of attacker's tracking. The specific process is shown in Figure 3.

After candidate phantom nodes are selected, the base station broadcasts a data packet carrying ID and relative position of P_1 and P_2 to the entire network, so that each node knows the information of candidate phantom nodes.

4.1.3. *H-hop limited flooding of source nodes.* The source node generates a broadcast message BM = {Bro_Source, id, h_s, θ, flag, s_x, s_y}, where Bro_Source represents the message type, and id represents the number of the nodes that sent messages. h_s represents the count value of messages and is initialized to 0. θ is the largest angle among the deflection angles of nodes in visible area. As shown in Figure 2, when node q is tangent to the circle of visible area, its deflection angle is the largest, and θ is the deflection angle of node q. flag is a flag bit indicating whether the node with node number id will cause a failed path, and the initial value is 0. (s_x, s_y) represents the position coordinate of source nodes. Then broadcast the message BM within h hops of source nodes. When the message BM arrives at each forwarding node, h_s increases by 1, and at the same time, according to the position of source node (s_x, s_y), the position of sink nodes and its own position are calculated according to the law of cosines and compared with θ. If the deflection angle is smaller than θ, set flag to 1, and broadcast the message to its neighbor nodes. When h_s counts to h, the node no longer broadcasts messages. After h hop limited flooding,

nodes within h hop range obtain the minimum hop value between themselves and their neighbor nodes from source nodes. At the same time, you can determine which of your neighbor nodes will cause failure paths.

4.1.4. *$h + c$ hop phantom routing.* The source node s sends a data packet $\mathrm{PK} = \{\mathrm{E_k(m)}, h\_r, nx\_id\}$ every ΔT time. $\mathrm{E_k(m)}$ represents the result of encrypting message content m with the key k, h_r represents the count variable of the number of packet forwarding hops, and nx_id represents the id number of next hop node. h_r is initially 0, and h_r increases by 1 every time a data packet is forwarded. Until the count is h, then stop forwarding. The node classifies neighbor nodes whose hops from source nodes are greater than the hops from source nodes to the remote node set. Then each time a data packet is forwarded, a node that does not cause a failure path is selected as the next hop node in the set. In this way, each hop of data packet is performed in the direction away from source nodes, and at the same time, it will not enter the failure path area. The source node 3 divides neighbor nodes into left and right sets according to the direction of the shortest route to sink nodes. Two consecutive data packets are sent out by different sets. That is, if source nodes select the next hop sending node in the left set at this time, it will select the next hop sending node in the right set when sending a data packet next time.

The node that data packet arrives after being forwarded by the h hop route is called phantom node 1. Then the phantom node 1 selects the node deviating from the source direction of data packet as down-hop sending node among its gradient neighbor nodes. The node receiving data packet performs $c - 1$ hop routing and forwarding in the same way, and finally reaches phantom node 2.

The schematic diagram of $h + c$ hop phantom routing is shown in Figure 4. The h hop phantom routing reaches phantom node $\mathrm{P_1}$. The minor arc AB will cause a failure path, and phantom node 1 will not be selected on the minor arc AB. After the hop, it reaches phantom node $\mathrm{P_2}$ with the same gradient as $\mathrm{P_1}$, and then $\mathrm{P_2}$ acts as a pseudo source node to forward data packet to sink nodes.

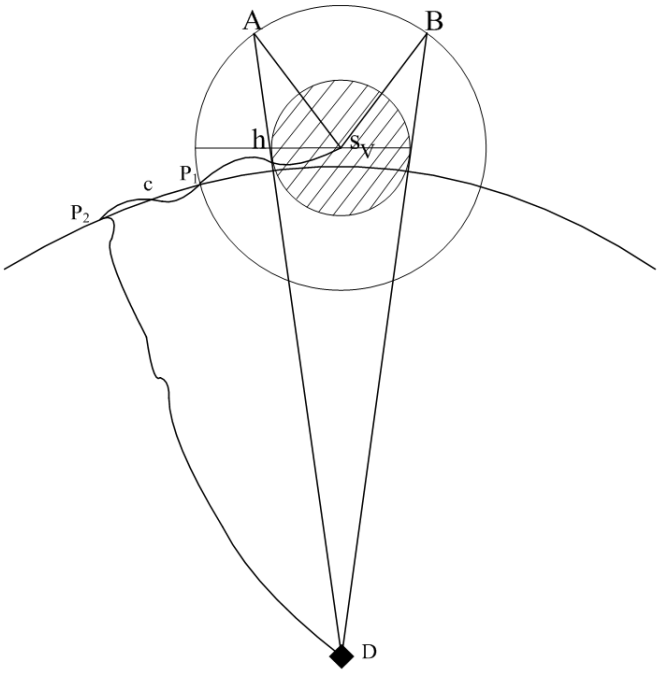


FIGURE 4. $h + c$ hop phantom routing

4.1.5. *Directed random routing.* If data packet has completed the above mentioned $h + c$ hop routing process, the node that receives data packet randomly selects among its neighbor nodes – a node whose hop count from sink nodes is less than the hop count from itself to sink nodes is taken as the next hop send node. Repeat this process until the data packet is transmitted to sink nodes.

4.2. **Performance analysis.** In order to facilitate the analysis of location distribution for phantom nodes, the following theorem is introduced.

Theorem 4.1. *Within the finite flooding range of source node s, the absolute value of the difference between minimum number of hops from any node u to source nodes and minimum number of hops from its neighbors to source nodes is less than or equal to 1.*

Proof: Suppose the neighbor node of node u is denoted by v, and the minimum number of hops from node u to source nodes is h_{u-s}. The minimum number of hops from node v to source nodes is h_{v-s}. The hop count of node u and node v is h_{u-v}. When the source node performs limited flooding, the node adds 1 to h_s in broadcast message BM data packet and forwards it to its neighbor nodes. That is, if nodes u and v are neighbor nodes, then $h_{u-v} = 1$.

When $h_{u-s} = h_{v-s}$, obviously Theorem 4.1 holds.

If $h_{u-s} > h_{v-s}$, then there must be a path through node v in the path from node u to source nodes. Node u reaches node v through 1 hop, and then reaches source nodes through the shortest path from node v. Therefore, the number of hops passed by this path is $1 + h_{v-s}$. $1 + h_{v-s}$ must be greater than or equal to the minimum hop count h_{u-s} of node u, that is, $1 + h_{v-s} \geq h_{u-s}$, then $h_{u-s} - h_{v-s} \leq 1$. Similarly, when $h_{v-s} > h_{u-s}$, $h_{v-s} - h_{u-s} \leq 1$. Therefore, Theorem 4.1 holds.

Theorem 4.2. *If the transmission radius of node u is r and the minimum routing hop from node to source node s is h_{u-s}, then the actual distance from node u to source node s is less than or equal to $h_{u-s} \times r$.*

Proof: 1) Node u is the neighbor node of source node s, and then the distance from s to u is only one hop:

$$d_{s-u} \leq 1 \times r = h_{s-u} \times r \tag{2}$$

2) Node u is not a neighbor node of source node s, and then source node s to node u passes through transit node $a_1, a_2, \ldots, a_{h_{s-u}}$. Because the distance between any two nodes that can communicate directly is less than r,

$$|su| \leq |sa_1| + |sa_2| + \cdots + \left|sa_{h_{s-u}}\right| \leq h_{u-s} \times r \tag{3}$$

In summary, the relationship between source node s and node u satisfies Theorem 4.2.

Theorem 4.3. *In a barrier-free environment, data packets are routed in h hop direction according to the maximum number of hops from neighbor nodes to source nodes, and the arriving phantom nodes u are distributed with source nodes as the center of circle. The inner radius is $(h-1) \times r$ and outer radius is $h \times r$ in annular area.*

Proof: The data packet selects next hop based on the maximum number of hops from neighbor nodes to source nodes. According to Theorem 4.1, the minimum number of hops from phantom nodes to source nodes is h. We use mathematical induction to reason.

1) When $h = 1$, the phantom node is located with source nodes as the center of circle. On a circle with r being the radius, Theorem 4.3 holds.

2) Suppose that when $h = k$, Theorem 4.3 holds. Then when $h = k + 1$, Theorem 4.3 does not hold. Let the minimum number of hops from node v to source node s be $k + 1$. From Theorem 4.2, the actual distance between node v and the source node is less

than $(k+1) \times r$. If node v is located on a ring with source nods as the center, the inner radius is $(k-1) \times r$, and the outer radius is $k \times r$, it must be centered on source nodes. Find a point in a circular area with a radius of $(k-1) \times r$ so that the distance to node v is less than r, that is, this node and node v are neighbor nodes. Then from Theorem 4.1, the absolute value of difference between this node and the minimum number of hops from node v to source nodes is less than or equal to 1. However, the minimum number of hops from this point to source nodes is $k-1$. The minimum number of hops from node v to source node s is $k+1$, and the absolute value of difference is 2. It contradicts the assumption, so Theorem 4.3 holds.

5. **Simulation Experiment and Result Analysis.**

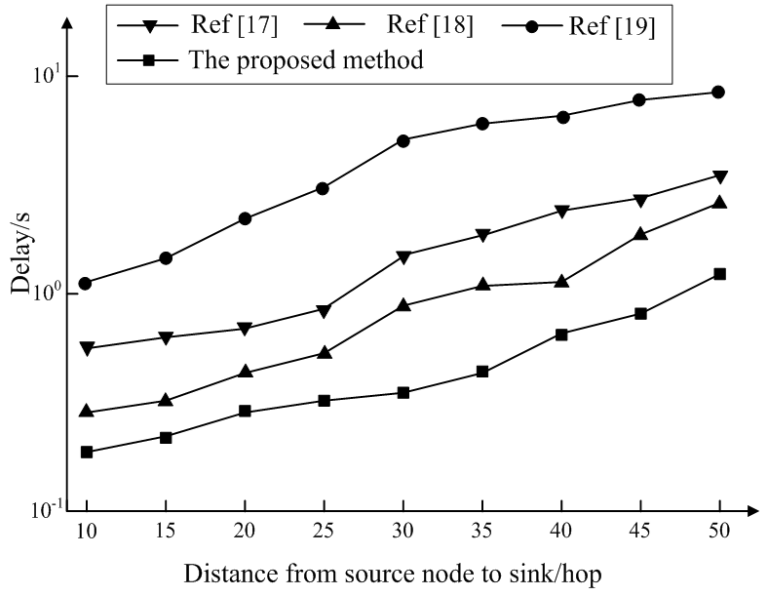
5.1. **Experimental setup.**

In order to verify the performance of algorithm in this paper in large-scale networks, the experiment chooses a simulator based on OMnet++, Castalia, and deploys 100×100 sensors in a square flat network. The nodes are evenly distributed randomly, and sink nodes are randomly placed in the center of network. Suppose there is only one attacker in the network, and source nodes are placed at a different location from sink nodes. The MAC layer protocol is based on IEEE 802.15.4, and the heartbeat packet load carries signal information. When a node detects an attacker, it sends a heartbeat packet. In order to show the performance of proposed strategy, reference [17] strategy, reference [18] strategy and reference [19] strategy are selected for comparison. The simulation is performed 100 times, and each time 500 new data packets are sent from source nodes.

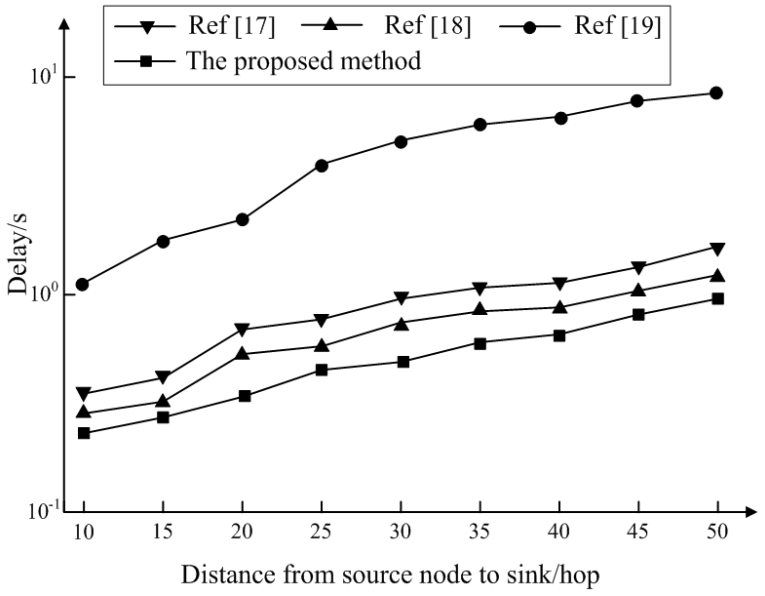
5.2. **Experimental results and analysis.**

5.2.1. *Comparative analysis of transmission delay.* As shown in Figure 5, as the distance from source nodes to sink increases, the delay caused by the strategy in [19] also increases significantly, which is much higher than other three methods. The reference [19] strategy relied on heartbeat packet data to update routing table before each data delivery; otherwise security protection cannot be achieved. This caused a great time delay. In [19], the path length of this strategy is related to attackers' attack mode. For a patient attacker who stays near the sink, the strategy in [19] is based on the shortest path. However, they always choose to deviate from the farthest node of attackers, until attackers capture a new data packet. For a curious attacker, if a new data packet is not captured, he will walk randomly. Far away from the shortest path, the path will not be offset. Therefore, compared with a more curious attacker, the strategy in [19] will produce a little more time delay when facing a patient attacker.

Our proposed strategy is the same as the strategies in [17-19], and the path offset is also affected by attackers' location. As shown in Figure 5(a), for patient attackers, the proposed strategy mechanism causes the delivery delay of data packets to be much higher than the shortest path. This is because when attackers approach sink nodes, it will cause the path to drift. This causes more than 90% of data packets to fail to be routed to sink nodes. In the simulation experiment, some data packets were forwarded 145 times before reaching sink nodes. As shown in Figure 5(b), when a curious attacker is encountered, the data packet delay of the proposed strategy is greatly reduced. This is because curious attackers cannot receive data packet and walk randomly, away from the shortest path. Therefore, in practical applications, the minimum safety distance can be reduced near sink nodes. When it is far away from sink nodes, adjust it back to the communication radius accordingly.



(a) Patient attacker

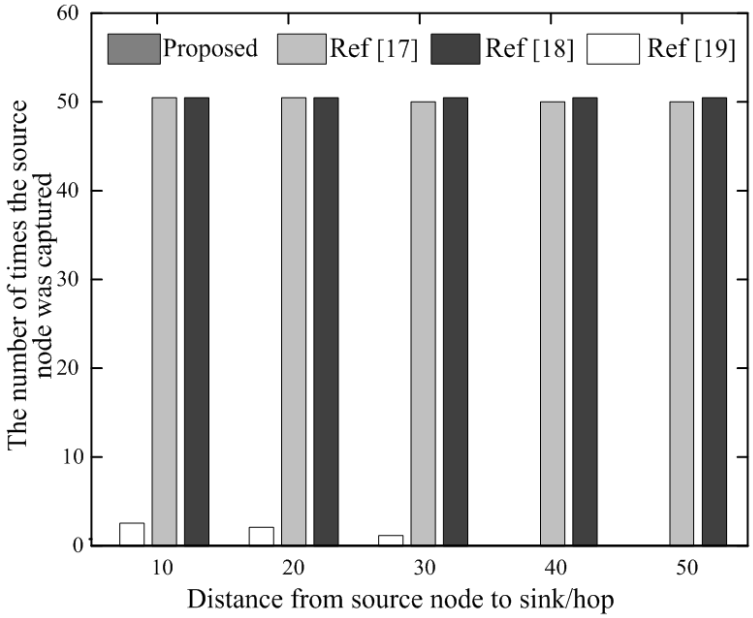


(b) Curious attacker

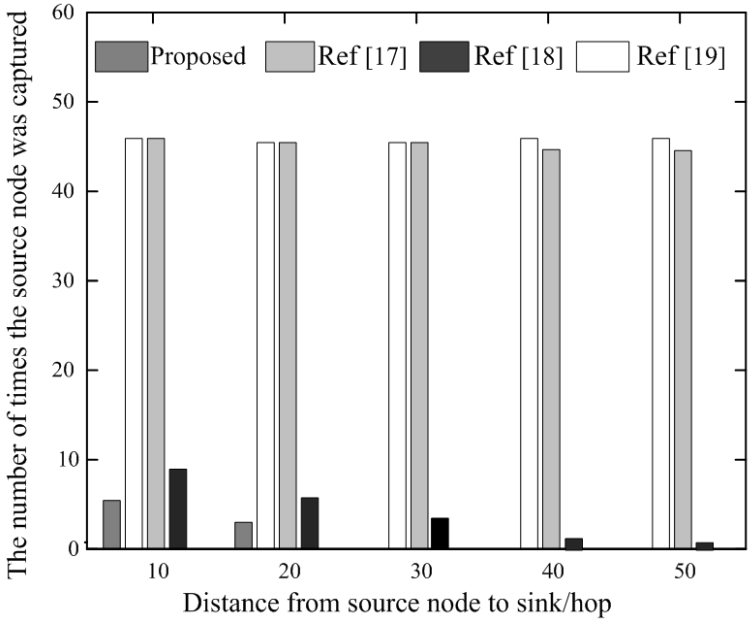
FIGURE 5. The relationship between time delay and s-d distance

5.2.2. *Comparative analysis of privacy security.* As shown in Figure 6, with the increase of s-d, the security of the shortest path does not increase, and is the worst. After attackers capture a data packet, it will follow the shortest path to find source nodes. It can be seen from the experimental results that curious attackers have captured source nodes fewer times than patient attackers. Curious attackers start from sink and walk randomly without waiting for any data packets. Thus, some data packets are missed and the real path is shifted. Thus, patient attackers have a higher capture rate than curious attackers.

The essence of the reference [19] strategy is greedy shortest path, but every time it chooses the farthest away from attackers. Moreover, the strategy in reference [19] is related to attackers' attack radius setting. In the simulation, when attack radius is equal to communication radius, attackers can receive some real data packets near the sink. So



(a) Patient attacker



(b) Curious attacker

FIGURE 6. The relationship between the number of captured nodes and s-d distance

patient attackers can also capture some data sources. A curious attacker walks randomly when not receiving information, and captures some source nodes when s-d is small. The strategy in [17] does not perform better than the shortest path when it encounters a patient attacker in simulation. It is just that as s-d increases, the number of random walks increases. A curious attacker will deviate from the route because it has not received data packet temporarily, and lose part of the opportunity to capture source nodes.

As shown in Figure 6(a), under the proposed strategy, patient attackers stay in place because he has not received data packets. The data source cannot be tracked continuously, and the privacy security is close to 100%. When source nodes are close to the sink, curious attackers are in a random walk state because it basically cannot receive data packets. It

will be captured when source nodes are very close. As shown in Figure 6(b), when s-d = 10 hops and the proposed strategy encounters a curious attacker, the proposed strategy is much more private than the strategy in [17]. This is because the distance s-d is shorter in the actual experiment, and the speed and range of attackers' random walk in the real environment are smaller than those in the simulation environment.

5.2.3. *Comparative analysis of safety time.* As shown in Figure 7, the security time of the strategy in [17] is higher than that of other three comparative strategies. This is because the data packet transmission path of the strategy in [17] avoids attackers' visible area, which prolongs the time for attackers to discover the location of source nodes. The strategies of [18] and [19] selected phantom nodes too close to speed up the exposure time of source nodes. The security time of the proposed strategy is set according to the plan. If security time at this time cannot meet the time required for location privacy security of current source node, the area of low-latency area can be expanded accordingly. This allows the transmission delay and safety time to achieve an optimal match.

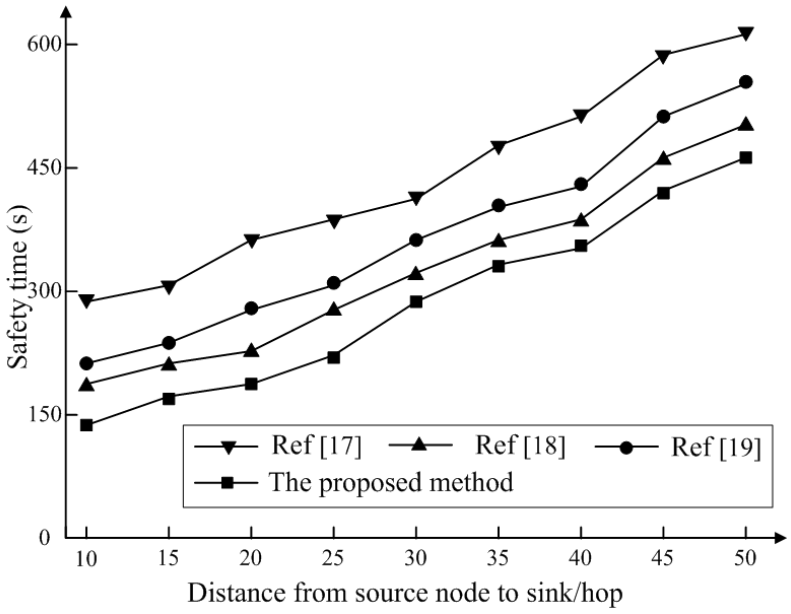


FIGURE 7. Comparison of security time of several strategies

5.2.4. *Comparative analysis of information integrity.* In order to further verify the performance of the algorithm in Internet of Things privacy protection, a loss channel model is constructed. When information passes by the channel, packet loss will occur. With the gradual increase of packet loss rate, the comparison results of our proposed strategy, reference [17] strategy, reference [18] strategy and reference [19] strategy to protect the integrity of information are described in Figure 8.

Analyzing Figure 8, it can be seen that when the information passes by loss channel, as the packet loss rate gradually increases, the integrity of our proposed strategy, reference [17] strategy, reference [18] strategy and reference [19] strategy is all in a certain degree of decline. However, the proposed strategy has the lowest decline and integrity curve has been kept at the lowest level. This shows that the proposed strategy can still effectively ensure the integrity of information in the presence of packet loss, and the privacy protection effect is very good.

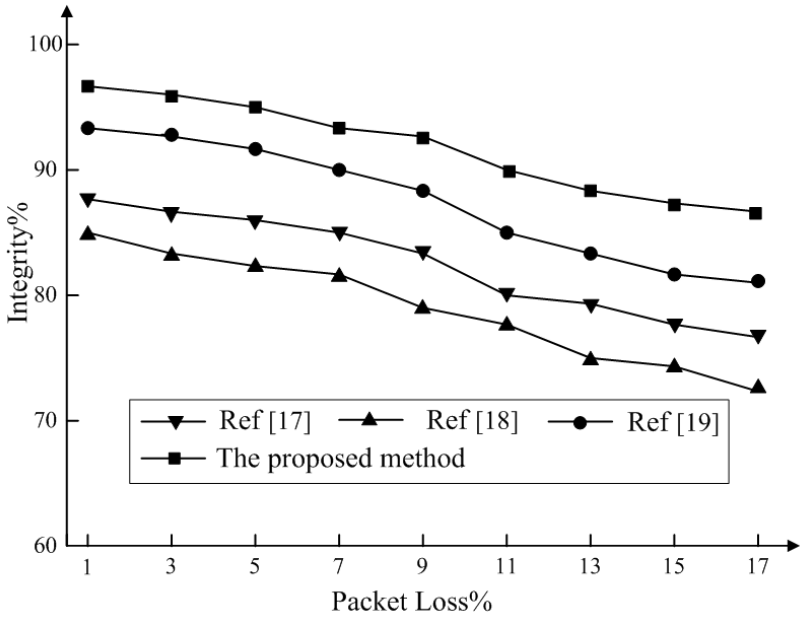


FIGURE 8. Comparison of integrity of several strategies under different packet loss rates

6. **Conclusion.** Researchers at home and abroad have made a series of progress in the research on the privacy protection of source node location in wireless sensor networks, and they proposed many classic source location privacy protection strategies. In order to ensure the diversification of geographic location for phantom nodes, this paper proposes a privacy protection algorithm for source node location based on phantom routing in the Internet of Things environment. This algorithm can more effectively resist attacks from strong visual attackers and strengthen the protection of source location privacy. By selecting the phantom nodes twice, phantom nodes are scattered and far away from source nodes. At the same time, the transmission path from source nodes to sink nodes avoids the node that causes failure path to achieve the balance of network security requirements. Experiments proved that the proposed strategy can effectively guarantee the integrity of information even in the presence of packet loss, and the privacy protection effect is very good. Under the proposed strategy, patient attackers stay in place because they have not received data packets and cannot continue to track data sources. Therefore, the privacy security is close to 100%.

At present, the strategy proposed in this paper can simulate and run well in the simulation environment. In actual application scenarios, source nodes may face global attackers or even multiple types of attackers attacking together. Thus, the proposed strategy inevitably has certain limitations when applied to real environments. In the next work, attackers' attack conditions will be relaxed. Thus, a source location privacy protection scheme is designed to resist more complex attackers.

REFERENCES

[1] S. Cha, T. Hsu, Y. Xiang and K. Yeh, Privacy enhancing technologies in the Internet of Things: Perspectives and challenges, *IEEE Internet of Things Journal*, vol.6, no.2, pp.2159-2187, 2019.

[2] C. Li and B. Palanisamy, Privacy in Internet of Things: From principles to technologies, *IEEE Internet of Things Journal*, vol.6, no.1, pp.488-505, 2019.

[3] J. Xiong et al., Enhancing privacy and availability for data clustering in intelligent electrical service of IoT, *IEEE Internet of Things Journal*, vol.6, no.2, pp.1530-1540, 2019.

[4] Y. Chen and L. Wang, Privacy protection for Internet of Drones: A network coding approach, *IEEE Internet of Things Journal*, vol.6, no.2, pp.1719-1730, 2019.
[5] W. Liu, Y. Jia and Z. Wang, Location-based random key predistribution scheme of wireless sensor networks, *ICIC Express Letters*, vol.11, no.2, pp.365-372, 2017.
[6] M. Wang, D. Xiao, Y. Xiang and H. Wang, Privacy-aware controllable compressed data publishing against sparse estimation attack in IoT, *IEEE Internet of Things Journal*, vol.6, no.4, pp.7305-7318, 2019.
[7] D. Li, Q. Yang, D. An, W. Yu, X. Yang and X. Fu, On location privacy-preserving online double auction for electric vehicles in microgrids, *IEEE Internet of Things Journal*, vol.6, no.4, pp.5902-5915, 2019.
[8] J. Xiong et al., A personalized privacy protection framework for mobile crowdsensing in IIoT, *IEEE Transactions on Industrial Informatics*, vol.16, no.6, pp.4231-4241, 2020.
[9] P. Yang, X. Kang, Q. Wu, B. Yang and P. Zhang, Participant selection strategy with privacy protection for Internet of Things search, *IEEE Access*, vol.8, pp.40966-40976, doi: 10.1109/ACCESS.2020.2976614, 2020.
[10] S. R. Pokhrel, Y. Qu and L. Gao, QoS-aware personalized privacy with multipath TCP for industrial IoT: Analysis and design, *IEEE Internet of Things Journal*, vol.7, no.6, pp.4849-4861, 2020.
[11] X. Xia, Z. Bai, J. Li et al., A location cloaking algorithm based on dummy and Stackelberg game, *Chinese Journal of Computers*, vol.42, no.10, pp.2216-2232, 2019.
[12] C. Y. Chow, M. F. Mokbel and X. Liu, A peer-to-peer spatial cloaking algorithm for anonymous location-based service, *Proc. of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems*, pp.171-178, 2006.
[13] Y. Huang, Z. Huo and X. Meng, CoPrivacy: A collaborative location privacy-preserving method without cloaking region, *Chinese Journal of Computers*, vol.34, no.10, pp.1976-1985, 2011.
[14] Y. Chen, X. Liu and B. Li, Collaborative position privacy protection method based on game theory, *Computer Science*, vol.40, no.10, pp.92-97, 2013.
[15] J. Jiang and C. Fu, Location privacy protection method based on query fragment and user collaboration, *Journal of Chinese Mini-Micro Computer Systems*, vol.40, no.5, pp.935-940, 2019.
[16] X. Li, M. Miao, H. Liu et al., An incentive mechanism for Kanonymity in LBS privacy protection based on credit mechanism, *Soft Computing*, vol.21, no.14, pp.3907-3917, 2017.
[17] J. Xu, M. Wen and K. Zhang, An improved k-anonymous incentive mechanism scheme combined with blockchain technology, *Computer Engineering and Application*, vol.56, no.6, pp.111-116, 2020.
[18] H. Liu, X. Li, B. Luo et al., Distributed k-anonymity location privacy protection scheme based on blockchain, *Chinese Journal of Computers*, vol.42, no.5, pp.942-960, 2019.
[19] Y. Ji, R. Gui, X. Gui, D. Liao and X. Lin, Location privacy protection in online query based-on privacy region replacement, *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, pp.742-747, 2020.
[20] C. Yin, J. Xi, R. Sun and J. Wang, Location privacy protection based on differential privacy strategy for big data in Industrial Internet of Things, *IEEE Transactions on Industrial Informatics*, vol.14, no.8, pp.3628-3636, 2018.
[21] P. Kamat, Y. Zhang, W. Trappe et al., Enhancing source-location privacy in sensor network routing, *IEEE International Conference on Distributed Computing Systems*, pp.599-608, 2005.
[22] I. S. Acharyya, A. Al-Anbuky and S. Sivaramakrishnan, Software-defined sensor networks: Towards flexible architecture supported by virtualization, *2019 Global IoT Summit (GIoTS)*, Aarhus, Denmark, pp.1-4, doi: 10.1109/GIOTS.2019.8766429, 2019.
[23] B. S. Awoyemi, A. S. Alfa and B. T. Maharaj, Network restoration in wireless sensor networks for next-generation applications, *IEEE Sensors Journal*, vol.19, no.18, pp.8352-8363, 2019.
[24] R. Maheswar, A. R. Maria, N. Sheriff, V. Mahima, G. R. Kanagachidambaresan and M. Lakshmi, Mobility Aware Next Hop Selection Algorithm (MANSA) for wireless body sensor network, *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, pp.1-5, 2019.
[25] C. Gao, Z. Wang and Y. Chen, On the connectivity of highly dynamic wireless sensor networks in smart factory, *2019 International Conference on Networking and Network Applications (NaNA)*, Daegu, Korea, pp.208-212, 2019.
[26] O. Dagdeviren and V. K. Akram, The effect of random node distribution and transmission ranges on connectivity robustness in wireless sensor networks, *2019 International Symposium on Networks, Computers and Communications (ISNCC)*, Istanbul, Turkey, pp.1-5, 2019.

[27] M. S. Azizi and M. L. Hasnaoui, Software defined networking for energy efficient wireless sensor network, *2019 International Conference on Advanced Communication Technologies and Networking (CommNet)*, Rabat, Morocco, pp.1-7, 2019.

[28] Y. Kimura, E. Nii and Y. Takizawa. Cooperative detection for falsification and isolation of malicious nodes through inter-node vote for wireless sensor networks in open environments, *2019 Global Information Infrastructure and Networking Symposium (GIIS)*, Paris, France, pp.1-3, 2019.

[29] M. Sana and L. Noureddine, Multi-hop energy-efficient routing protocol based on minimum spanning tree for anisotropic wireless sensor networks, *2019 International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, Hammamet, Tunisia, pp.209-214, 2019.