Research Article

An Efficient Anonymous Communication Scheme to Protect the Privacy of the Source Node Location in the Internet of Things

Fengyin Li ¹, Pei Ren ¹, Guoyu Yang ¹, Yuhong Sun ¹, Yilei Wang ¹, Yanli Wang ¹,
Siyuan Li ¹ and Huiyu Zhou ²

¹School of Computer Science, Qufu Normal University, Rizhao 276826, China

²School of Informatics, University of Leicester, Leicester LE1 7RH, UK

Correspondence should be addressed to Yilei Wang; wang_yilei2019@qfnu.edu.cn

Received 18 December 2020; Revised 22 January 2021; Accepted 15 February 2021; Published 24 February 2021

Academic Editor: Wenjuan Li

Copyright © 2021 Fengyin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Advances in machine learning (ML) in recent years have enabled a dizzying array of applications such as data analytics, autonomous systems, and security diagnostics. As an important part of the Internet of Things (IoT), wireless sensor networks (WSNs) have been widely used in military, transportation, medical, and household fields. However, in the applications of wireless sensor networks, the adversary can infer the location of a source node and an event by backtracking attacks and traffic analysis. The location privacy leakage of a source node has become one of the most urgent problems to be solved in wireless sensor networks. To solve the problem of source location privacy leakage, in this paper, we first propose a proxy source node selection mechanism by constructing the candidate region. Secondly, based on the residual energy of the node, we propose a shortest routing algorithm to achieve better forwarding efficiency. Finally, by combining the proposed proxy source node selection mechanism with the proposed shortest routing algorithm based on the residual energy, we further propose a new, anonymous communication scheme. Meanwhile, the performance analysis indicates that the anonymous communication scheme can effectively protect the location privacy of the source nodes and reduce the network overhead.

1. Introduction

The coming of age of the science of machine learning (ML) coupled with advances in computational and storage capacities have transformed the technology landscape. For example, within the security domain, detection and monitoring systems now consume massive amounts of data and extract actionable information. ML is now pervasive—new systems and models are being deployed in every domain imaginable [1]. Internet of Things is based on the Internet, using RFID, wireless data communication, and other technologies to construct a network covering possible nodes in the world. In the IoT, objects can “communicate” with each other without human intervention. IoT devices are committed to maximizing their utility within a limited capacity and maintaining the security of the IoT system [2]. Wireless sensor networks are an important part of the Internet of Things, and it is a distributed sensor network. Sensor nodes are stationary or mobile, and they constitute a

wireless sensor network in a self-organizing and multihop manner. As a link between the physical world and the virtual world, the WSNs have become one of the most promising technologies. They have the ability to monitor object exits in the network, and realize data collection, processing, and transmission. At present, the WSNs have been widely used in military, transportation, medical, household, and industrial fields. However, due to the large-scale deployment of wireless sensor networks, security and privacy risks become critical. Currently, in the context of the Internet of Things, the WSNs still face security problems. For instance, adversaries can infer the location information of the source node through backtracking attacks, and then obtain the location information of an event, which causes the leakage of sensitive messages.

In the IoT, many existing security schemes can protect message content and contextual information using the traditional cryptography theory, but they cannot solve the problem of the source node location privacy protection.

Adversaries can obtain sensitive messages through traffic analysis [3], hop-by-hop tracking, backtracking attacks, and other methods. Moreover, many schemes fail to take into account the finiteness of the node resources, which result in a large amount of resource consumptions, unsuitable for the WSNs with limited energy. In addition, many schemes cannot resist traffic analysis attacks. A large amount of traffic will be generated near the source nodes due to data packets being transmitted between nodes. Therefore, when an adversary analyzes the traffic in the network to find the hot spots in the network, it will obtain the correct location of the source node and attack it.

In the random routing algorithm, after purely random h hops, the probability that the proxy source node is no more than $h/5$ hops away from the real source node is as follows [4]:

$$P = 1 - e^{-(h/25)}. \quad (1)$$

When h is large enough, the value of P approaches 1. In other words, purely random routing does not guarantee that the selected proxy source node is far enough away from the real source node. However, if the selected proxy source node is very close to the real source node, the position of the real source node cannot be effectively hidden. Therefore, the proxy source node that we select is required to be far enough away from the real source node.

In this paper, we propose an efficient anonymous communication scheme to protect the privacy of the source node location, which not only protects the source location privacy in wireless sensor networks but also guarantees the forwarding efficiency of anonymous messages through the shortest path routing algorithm, and reduces the network overhead.

Specifically, the main contributions of this paper are as follows:

- (1) We propose the mechanism for selecting proxy source nodes based on the candidate region. This proposal selects the nodes that meet the upper and the lower limits of the hop count to construct the candidate region around the source node. It also selects the proxy source node to replace the real source node in the candidate region to forward the messages in order to realize the location privacy protection of the real source node.
- (2) We propose the shortest path routing algorithm based on the residual energy. When forwarding a message from the proxy source node to the sink, each node selects the next hop according to the residual energy of its neighbor nodes and the minimum number of hops from the sink to itself. We propose the shortest path routing algorithm based on the residual energy, which improves the efficiency of anonymous message forwarding.

The rest of the paper is organized as follows. In Section 2, we introduce related work. In Section 3, we first design the system model of the anonymous communication system, and then we propose an anonymous communication scheme based on the proxy source node and the shortest path

routing. In Section 4, the results and discussion of the proposed scheme is given. Finally, we conclude this paper in Section 5.

2. Related Works

In recent years, privacy issues have been a hot issue in machine learning. To avoid privacy issues caused by massive data collection, Mohassel et al. proposed new and efficient protocols for privacy preserving machine learning for linear regression [5], logistic regression, and neural network training using the stochastic gradient descent method. In order to solve the risk of large-scale collection of sensitive data, Bonawitz et al. designed a novel, communication-efficient, failure-robust protocol for the secure aggregation of high-dimensional data [6]. The protocol has good security in an honest but curious and active adversary environment. At the same time, even if a randomly selected subset of users exits at any time, the security is maintained.

Blockchain has been widely discussed and used in the IoT [7, 8], in order to be able to balance between fairness and incentive compatibility. Wang et al. tailored a new bonus reward function by adding random salts to the geometric reward function [9]. Li et al. highlighted the combination of game theory and blockchain [10], including rational smart contracts, game theoretic attacks, and rational mining strategies. In the IoT, privacy issues are also a hot issue of research. The service evaluation model is an important part of the service-oriented Internet of Things (IoT) architecture, but it is vulnerable to various attacks. Li et al. put forth a new service evaluation model named Tesia allowing specific users to submit the comments as a group in the IoT networks [11] to solve the problem. To enhance the privacy of the source location in wireless sensor networks, Zhao et al. conducted a comprehensive investigation on the theory and practice of the SMPC protocol [12], explaining the security requirements and the basic construction technology of the SMPC. It also introduces the research progress of the general SMPC protocol construction technology and its application in the IoT. Wang et al. proposed a trace-cost-based source location privacy protection scheme in wireless sensor networks for a smart city (TCSLP) [13], by constructing a phantom area, and combining shortest path routing and random routing to send packets, whereby the security time of the smart city in the wireless city is extended and the SLP is enhanced. Zhu proposed a method of regional division based on node location information [14], and by using this method, he selects the hop distance between the location nodes. The distance accuracy of the data nodes in the vicinity selected during information transmission is improved, and the location privacy of the source node is better protected. Han et al. proposed a dynamic ring-based routing (DRBR) scheme [15], which solved the balance issues between security and energy consumption and provided efficient source location privacy. Muruganathan et al. proposed a centralized energy-efficient routing protocol for the WSNs (BCDCP) [16], which can evenly distribute energy consumption among all the sensor nodes to improve network life and save energy on average. Mutalemwa and Shin proposed a routing scheme

with stronger source location privacy than the traditional routing scheme [17], providing a highly random routing path between the source and the sink nodes. Randomly send data packets to the sink node through tactically positioned proxy nodes, and implement the stronger source location privacy. In order to protect the privacy of the event and observe the privacy of the source node reporting the event, Chakraborty and Verma proposed a differential privacy framework [18]. By reporting the accumulation of the real and the virtual traffic of the same event, they distinguished the real and virtual events and provided differential privacy protection for nodes in the network.

Wang et al. proposed a data domain partitioning model [19], which is more accurate to choose the grid size. They proposed a uniform grid release method based on this model, and further improved the query accuracy. To solve the problem of privacy leakage caused by data analysis and mining, Spachos and Toumpakaris proposed a source-location privacy scheme that employs randomly selected intermediate nodes based on inclination angles [20], and analyzed the introduced angle-based dynamic routing scheme. However, as this scheme is for the data transmission of the included angle region, it could not adapt this angle for selecting an optimal routing. Furthermore, Liu and Xu proposed a new scheme to dynamically change the included angle in the ADRS—dynamic routing scheme (VADRS) based on the included Angle [21]. The scheme further improves the security performance of the ADRS by selecting the optimal Angle for data transmission at each hop. Aiming at the low security cycle of the existing source location privacy protection algorithm, Bai et al. proposed a source location privacy protection algorithm based on the expected phantom source node [22]. An ellipse is established through the coordinates of the source and the sink nodes, and a node is randomly selected on the ellipse as the expected phantom source node. The source location privacy protection is realized based on the phantom source node. Li et al. proposed a new routing strategy [23]. The routing strategy is divided into three stages to route data packets to the base station: directional random route, H-hop route, and the shortest path route in the ring area. The source location privacy protection is realized when information is sent to the base station in the WSNs.

Lin proposed schemes such as the ant colony algorithm to protect the location information of the source nodes and the multisource and the multipath protection of the source node location, etc. [24], to achieve the protection of node location privacy. To avoid the leakage of user personal information from the IoT devices during data processing and transmission, Li et al. proposed a certificateless encryption scheme to implement a novel anonymous communication protocol [25]. In the protocol, an anonymous communication link establishment method and an anonymous communication packet encapsulation format are proposed. It improves the privacy, security, and efficiency of CPSS anonymous communication. Sharma and Ghosh proposed new technology to prevent active and passive attacks in the mobile base station environment [26]. By deploying mobile sinks in the network, data were collected from sensor nodes and sent to the fixed base station, so as to guarantee the privacy of data in the mobile sink. Tan et al. proposed two effective source node location privacy protection policies [27]:

the enhanced directional random routing protection mechanism (EDROW) and the multilayer ring proxy filtering mode routing protection mechanism (MRPFS). Aiming at the hop-by-hop reverse attackers with local traffic analysis behavior, Zhao et al. proposed the source location privacy protection routing protocol RAPFPR [28] based on the random angle and the probability forwarding. This protocol produces phantom nodes and enables them to be evenly distributed around the real source nodes and adopts the probabilistic forwarding routing mechanism, thus greatly reducing the generation of the overlapping paths. Sheu and Jiang proposed an anonymous path routing protocol (APR) for the wireless sensor networks [29]. This protocol encrypts data based on pair-wise keys, realizes the anonymous message transmission between adjacent nodes and the anonymous information transmission between the source node and the target node in the multi-hop communication path, and protects the data communication in the WSNs. Li and Ren proposed a source-location privacy scheme [30]. In this scheme, an anonymous path is constructed by randomly selecting intermediate nodes far away from the source node to realize the transmission of anonymous messages to the sink node. This solution provides satisfactory privacy of the local source location.

In order to better improve the privacy of source locations in the Internet of Things, we propose an anonymous communication scheme based on the proxy source node and the shortest path routing in this paper. This scheme can prevent the adversary from obtaining the location of the source node and event by means of backtracking attacks and traffic analysis. At the same time, the shortest path routing algorithm in this paper takes into account the residual energy of each node, ensuring the rationality of the energy overhead of the whole network.

3. An Anonymous Communication Scheme Based on the Proxy Source Node and the Shortest Path Routing

In order to realize the privacy protection of source locations in the IoT, we propose an anonymous communication scheme to protect the privacy of the source node location. In this anonymous communication scheme, the privacy protection of the source node location is achieved by setting the candidate region to select the proxy source node; the shortest routing algorithm based on the residual energy is used to achieve efficient anonymous message forwarding.

3.1. System Model

3.1.1. Network Model. First, we make the following assumptions about the network model:

- ① The wireless sensor network is composed of sensor nodes that are uniformly and randomly deployed, which cannot be moved at will after the nodes have been deployed. Any two nodes can communicate through multihop [31].
- ② The appearance of the object is randomly distributed throughout the network, so the probability of each

sensor detecting the object information is equal. The node that detects the object, i.e., the source node, periodically generates data packets and sends them to the base station. There is only one base station in the whole network, the base station is safe, and it cannot be destroyed by adversaries.

③ The adversary cannot attack the object in the area that is one hop away from the base station because this area has powerful surveillance capabilities.

The symbols used in this paper are shown in Table 1.

3.1.2. Adversary Model. The adversary is assumed to be an external, passive, and global attacker [31]:

- ① *External.* An external adversary is an attacker who will not compromise or control any sensor nodes.
- ② *Passive.* Passive means that we assume that the adversary will not conduct any active attacks, such as traffic injection, channel interference, or denial of service attack. The adversary cannot decrypt the data packet and tamper with the contents of the data packet, nor destroy the sensor node.
- ③ *Global.* A global adversary is the one who we assume that an adversary can collect and analyze communications throughout the network.

3.1.3. Energy Consumption Model. In the IoT, no matter what routing strategy is used for data transmission, each node will consume energy to send and receive data. Therefore, here, we only consider the energy consumption generated when sending and receiving a certain number of bits of information [31, 32]. If the sender wants to send n -bit data to the receiver, and the distance between the two parties is l , then for the sender, the energy consumed to send n -bit data is defined as

$$E_{\text{sender}}(n, l) = \begin{cases} nE_{\text{con}} + n\epsilon_{fs}l^2, & l < l_0, \\ nE_{\text{con}} + n\epsilon_{\text{amp}}l^4, & l \geq l_0. \end{cases} \quad (2)$$

For the receiver, the energy consumed to receive n -bit data is defined as

$$E_{\text{receiver}}(n) = nE_{\text{con}}. \quad (3)$$

Among them, E_{con} represents the energy consumption in the sender or the receiver circuit, and the value of E_{con} is related to the distance between the sender and the receiver, i.e., l . We consider two models: for the free space and the multi-path fading channel models, their power losses are l^2 and l^4 , respectively. ϵ_{fs} and ϵ_{amp} are the energies required by the power amplification in these two models, respectively.

3.1.4. DH Key Exchange Algorithm. When two parties communicate, the storage and disclosure of the user keys is a very important issue. We should ensure the identity privacy of both parties and the forward-backward security of the

keys [33, 34]. Diffie-Hellman Key Exchange (D-H) is an algorithm jointly invented by Diffie and Hellman. Both parties in communication are able to generate shared cryptographic numbers only by exchanging publicly available information, and this cryptographic number is used as a key. This key can be used as a symmetric key to encrypt the communication content in subsequent communications.

Specifically, we assume that both Alice and Bob need a symmetric cryptographic key, but the communication line between the two parties has been eavesdropped on by an eavesdropper. At this time, Alice and Bob can generate the shared key by taking the *DH* key exchange in the following way:

- ① Take the prime number p and the integer a , a is a primitive root of p , a and p are disclosed
- ② Alice chooses a random number $X_A < p$, and calculates $Y_A = a^{X_A} \bmod p$
- ③ Bob chooses a random number $X_B < p$, and calculates $Y_B = a^{X_B} \bmod p$
- ④ Each party keeps X secret and Y public to the other party
- ⑤ The way Alice calculates the key is $K = Y_B^{X_A} \bmod p$
- ⑥ The way Bob calculates the key is $K = Y_A^{X_B} \bmod p$

In this way, Alice and Bob have the equal shared key.

3.2. Proxy Source Node Selection Mechanism Based on Candidate Region. The main idea of anonymous communication is to hide the identity or the communication relationship of the two parties through a certain method, so that the adversary cannot directly know or infer the communication relationship between the two parties or the party of the communication.

Anonymous communication in the WSN includes sender anonymity, receiver anonymity, and communication relationship anonymity. In this paper, we mainly focus on the sender's the anonymity and communication relationship anonymity. In order to hide the location information of the real source node, realize the privacy of the source location, and then realize the anonymity in WSN—the anonymity of the communication relationship, there is no identity information involved in the process of message transmission. Each node only knows who its previous hop and next hop are, and does not know the source and the destination of the information. At the same time, we must ensure that the selected proxy source node is far away from the real source node, so as to better protect the location privacy of the real source node.

We use the limited flooding from the real source node to establish an anonymous proxy path, and then establish a candidate region. Before each message is forwarded, a node will be selected from the candidate region as the proxy source node to send the message instead of the real source node. The real source node selects neighbor nodes that meet the energy requirements from the neighbor node list, and sends the detection data packets to the neighbor nodes.

TABLE 1: Definition of notations in this paper.

Notation	Definition
R_s	The real source node
P_s	The proxy source node
$E(\cdot)$	The secure encryption algorithm
$D(\cdot)$	The secure decryption algorithm
$I_e(J)$	Initial energy, value 2
$l_0(m)$	Threshold distance, value 87
$E_{\text{con}}(\text{nJ/bit})$	Energy consumption, value 50
$\varepsilon_{fs}(\text{pJ/bit/m}^2)$	The energy required by the power amplification in the free space channel model, value 10
$\varepsilon_{\text{amp}}(\text{pJ/bit/m}^4)$	The energy required by the power amplification in the multipath fading channel model, value 0.0013
T_u	The neighbor node list of node u
hop_b	The number of hops from the base station to the current node
hop_s	The number of hops from the real source node to the current node
$\text{Minhops}_{i,b}$	The minimum number of hops that the node i is away from the base station
$\text{Minhops}_{i,s}$	The minimum number of hops that the node i is away from the real source node
E_i	Residual energy of node i

3.2.1. H-Hop Limited Flooding Starting from the Real Source Node. After the real source node detects that the object is nearby, it performs a limited flooding with a beacon message $SM = \{ID_s, \text{hop}_s\}$ [4], and the range of the flooding is limited within h hops. The SM contains the ID number of node ID_s that sent the message and the hop value from the real source node to the current node hop_s (the initial value is 0, plus 1 for each hop). When node u receives the beacon message SM from node v , if ID_v already exists in the neighbor node list T_u of node u , then the $\text{Minhops}_{v,s}$ of ID_v in T_u is updated with the smaller value hop_s in SM and the current $\text{Minhops}_{v,s}$ of node v . Otherwise, node u adds a new record to T_u , adds ID_v and hop_s into it, and hop_s at time $\text{Minhops}_{v,s}$.

Then we add 1 to hop_s and compare it with $\text{Minhops}_{u,s}$ for its own basic information, and update $\text{Minhops}_{u,s}$ with the smaller one as the current minimum number of hops from u to the real source node.

Node u replaces the ID in the message SM with its own ID , and forwards SM to its neighbor nodes together with the new hop_s . The neighbor nodes of u perform the same operation as u until hop_s count reaches h . At this point, h -hop limits the flooding process of the beacon message SM . Each node i within the range of h hops from the real source node knows the minimum number of hops $\text{Minhops}_{i,s}$ from itself to the real source node and the minimum number of hops from its neighbor nodes to the source node.

3.2.2. The Source Node Establishes Anonymous Proxy Paths. Definition. A is a node in the sensor network, the current node u selects v from its neighbor nodes as the forwarding node of the next hop. If the node v satisfies $\text{Minhops}_{v,A} - \text{Minhops}_{u,A} \geq 1$, then we can say that the hop forwarding of the data packet from u to v is in a direction away from node A , where $A \neq u$ and $A \neq v$.

In addition, we define an optional set $u.\text{gather}$ for each node u . The nodes in the set are the neighbors of u and satisfy the condition that the minimum number of hops from the source node is greater than the minimum number of hops from node u to the source node.

According to the energy requirements of receiving and sending data packets, the source node selects the neighbor nodes that meet the energy requirements according to the residual energy of each neighbor stored in its neighbor node list, and sends a detection data packet (h' , Q) for possible proxy nodes. The detection data packet includes the number of hops h' from the source node to the node (the initial value of h' is 0) and a node queue Q ; at the beginning, Q only contains the ID of the source node. Each time a detection data packet arrives at a node, the node adds its own ID to the node queue Q , at the same time the hop count h' is added 1.

We first select neighbor nodes that meet the energy requirements from its neighbor node list; then, we verify whether these selected neighbor nodes exist in their own optional set, and forward the detection data packet to the neighbor nodes in the optional set $u.\text{gather}$.

The neighbor node repeats this process until h' reaches h , and the detection process is completed. The node that receives the data packet at the h hop returns the queue Q to the real source node along the original path. Each node queue Q received by the real source node constitutes an anonymous proxy path. During the message forwarding phase, the anonymous proxy path is responsible for forwarding anonymous data packets to the proxy source node.

The detection process is over. At this time, the real source node has obtained several anonymous proxy paths.

3.2.3. Establishment of Candidate Region. For the anonymous proxy paths obtained in the previous section, the real source node chooses the first returned t as the candidate anonymous proxy paths. According to the predetermined upper and lower limits of the number of hops, we select all the nodes between the upper and the lower limits of the number of hops on the candidate anonymous proxy paths to form a candidate region.

For example, we suppose the real source node selects t node queues as follows: $Q_1, Q_2, \dots, Q_t$. The region formed by all the nodes between the l th hop and the h th hop of each queue Q_i ($i = 1, 2, \dots, t$) is called the candidate region. At the same time, we call the region where the node from the

first hop to the m th ($m=l-1$) hop of each queue Q_i ($i = 1, 2, \dots, t$) is located as the visible region. The candidate region and the visible region in this example are shown in Figure 1.

3.2.4. Select Proxy Source Node. We select a node in the candidate region as the proxy source node of this communication. The path from the real source node to the proxy source node constitutes the anonymous proxy path of our anonymous communication.

In Figure 1, if the selected proxy source node P_s is the green node in the figure, then the anonymous proxy path $R_s \rightarrow N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_4 \rightarrow P_s$ from the real source node to the proxy source node is obtained.

3.3. Shortest Path Routing Algorithm Based on Residual Energy. The proxy source node uses the shortest path routing algorithm based on the residual energy to forward the data packets to the sink.

First, the proxy source node P_s obtains the residual energy of its neighbors by looking up the locally stored neighbor node list T_{P_s} , and selects all the neighbor nodes that meet the residual energy requirement, i.e., the residual energy can support the neighbor nodes that can receive and forward the data packets.

Then, the proxy source node searches its neighbor node list T_{P_s} again, selects a neighbor node with the smallest number of hops from the sink from the nodes that meet the remaining energy condition, and sends the data packet to the neighbor node.

Finally, after the neighbor node has received the data packet, it searches its neighbor node list in the same way as the proxy source node, selects the neighbor node that meets the energy requirements and has the smallest number of hops from the sink, which receives the data packet. This forwarding process is repeated till the data packet reaches the sink.

The shortest path routing algorithm based on the residual energy is shown below.

3.4. An Anonymous Communication Scheme Based on Proxy Source Node and Shortest Path Routing. Based on the previously proposed proxy source node selection mechanism, this section presents an anonymous communication scheme to protect the location privacy of the real source node. Our scheme is divided into three stages, namely, network initialization, anonymous path establishment, and anonymous message forwarding.

In the network initialization phase, the sink performs flooding of the beacon message BM in the network. At this stage, the nodes in the network can obtain the minimum number of hops from its own to the sink and the minimum number of hops from its neighbor nodes to the sink.

The anonymous path establishment phase includes four steps: the real source node performs h hop limited flooding, obtains an optional anonymous proxy path, establishes a candidate region, and selects the proxy source node. The real

source node obtains the anonymous proxy paths from the source node to the proxy source node according to Section 3.2.

In the anonymous message forwarding phase, the source node first forwards the data packet from the real source node to the proxy source node via the anonymous proxy path obtained in Section 3.2, and then the proxy source node forwards the data packet to sink via the shortest path algorithm based on the residual energy to complete anonymous forwarding of the data packet.

3.4.1. Network Initialization Phase

① *Deployment of Sensor Networks.* In the sensor network, it includes a real source node, a sink node, and N ($N \in \mathbb{N}^+$) wireless sensor nodes. These wireless sensor nodes communicate wirelessly with each other, and finally deliver the information to the sink node.

The topology of our wireless sensor network is shown in Figure 2, which depicts a random path from the real source node to the sink.

When the sensor network is deployed, each node u establishes a neighbor node list T_u . The neighbor node list contains the ID number ID_i of neighbor node i , the minimum number of hops $Minhops_{i,b}$ from neighbor node i to the base station, the minimum number of hops $Minhops_{i,s}$ from neighbor node i to the real source node, and the residual energy value E_i of neighbor node i . The data structure of the neighbor node list is shown in Table 2. Among them, there is a situation where a neighbor node is of only $Minhops_{i,b}$ but not $Minhops_{i,s}$.

The neighbor node ID of each node u is obtained by the flooding process of the base station and the real source node. The minimum number of hops $Minhops_{i,b}$ from the neighbor node i to the base station is obtained by the flooding process from the base station. The minimum number of hops $Minhops_{i,s}$ from the neighbor node i to the real source node is obtained by the h hop limited flooding process starting from the source node.

Basic information of node u : each node u stores its own basic information through a quadruple $(E_u, h, Minhops_{u,b}, Minhops_{u,s})$. Among them, E_u represents the residual energy value of u . H -parameter is the number of the hops in the limited flooding performed by the real source node, which is initialized to null and is obtained during the flooding process of the base station. $Minhops_{u,b}$ is the minimum number of the hops between node u and the sink node, $Minhops_{u,s}$ is the minimum number of the hops between node u and the real source node. Both $Minhops_{u,b}$ and $Minhops_{u,s}$ are initialized to be the maximum number of the nodes in the network.

② *Flooding of Base Station.* The base station floods the beacon message $BM = \{ID_b, hop_b, h\}$ to the network, which contains the ID number of the node ID_b that sent the message, the hop value from the base station to the current node hop_b (the initial value is 0, plus 1 for each hop) and the number of hops h required for the establishment of the

FIGURE 1: Establishment of candidate region.

```

cur_node  $i$  = proxy source;
Initialize_neighbor_node_list( $T_i$ );
Initialize_packetInfo( $pI$ ) = ( $Q_{\text{proxy}}$ ,  $E_{K_{s,b}}(M)$ );
while(cur_node  $i$  != sink) do
   $u$  = first_neighbor(node  $i$ );
  while( $E_u \geq E_{\text{sender}}(n, l) + E_{\text{receiver}}(n)$ ) do
    save  $ID_u$  in the array  $A[]$ ;
     $u$  = next_neighbor(node  $i$ );
  end while
  hops-min =  $N$ ;
  for(node  $u$  = first of ( $A[]$ ); node  $u$  in array  $A[]$ ;  $u$  = next of ( $A[]$ ))
    if(Minhops $_{u,b}$  < hops-min)
    {
      hops-min = Minhops $_{u,b}$ ;
       $ID_{\text{hops-min}}$  =  $ID_u$ ;
    }
  end for
   $i$  forwards the  $pI$  to  $ID_{\text{hops-min}}$ ;
  cur_node  $i$  =  $ID_{\text{hops-min}}$ ;
end while

```

ALGORITHM 1: Shortest path routing algorithm based on residual energy.

candidate region in the network. When node u receives the beacon message BM from node v , it performs the following operations in sequence:

First, u stores hops h in its own basic information.

Secondly, we search ID_v in the neighbor node list T_u . If the search is successful, i.e., there already exists node v in the list, we compare the newly received hop $_b$ with the original minimum hop value Minhops $_{v,b}$, node u retains the smaller value and updates it as the minimum hop value Minhops $_{v,b}$ from u to the base station. Otherwise, we add a new record to T_u , assign the ID_v , and use the number of hops hop $_b$ as the

Minhops $_{v,b}$, i.e., the minimum number of the hops from v to the base station.

Thirdly, we add 1 to the number of hops hop $_u$ received, compare it with Minhops $_{u,b}$ in the quadruple, and select the smaller value as the new Minhops $_{u,b}$.

Finally, we replace ID_v in the message with its own ID_u , together with the latest hop $_b$ (at this time hop $_b$ = hop $_b$ + 1) and the received hop count h , construct a new beacon message $BM = \{ID_u, \text{hop}_b, h\}$ and forward it to the next node.

After flooding, each node i in the network is associated with h in the limited flooding, the minimum number of the

FIGURE 3: An anonymous path to protect the privacy of the source location.

node is located at Q_{proxy} , and then encrypts the message M to be sent with the DH key $K_{s,b}$ shared by the real source node and the base station, forming a data packet $(Q_{\text{proxy}}, E_{K_{s,b}}(M))$ with the queue Q_{proxy} . Send the data packet randomly to a certain number of neighbor nodes, and the selected neighbor nodes must include the first hop node in Q_{proxy} . For the neighbor nodes selected for each hop, there are two situations. First, if the node is not in the queue, it will randomly send the data packet to the next node. Second, if the node is in the queue, then it will select the next node based on the ID stored in Q_{proxy} .

Repeat this process till the proxy source node receives the data packet and stops the transmission. The process of Stage 1 is shown in Figure 4.

Stage 2: Ps-sink message forwarding

The proxy source node uses the shortest path routing strategy based on the residual energy mentioned in Section 3.3 to send the data packet to the base station. First, the proxy source node looks up the residual energy in the neighbor node list and finds the neighbor nodes that meet the energy requirements. We search the neighbor node list to find the one neighbor node that has the smallest minimum number of hops from it to the base station in these neighbor nodes that meets the energy requirements. The proxy source node sends the data packet to the neighbor node. Then the neighbor node forwards the data packet to the next node in the same way, until the base station receives the data packet and stops the transmission. After the sink receives the data packet, it can get the

message by decrypting M with the operation of $D_{K_{s,b}}(M)$, where M is the message to be transmitted to the sink.

The process of Stage 2 is shown in Figure 5.

3.5. Enhanced Anonymous Communication. In order to further improve the anonymity of the scheme in this paper, an enhanced anonymous communication scheme is proposed in this section by dividing the candidate region into several sectors.

According to the scheme proposed above, if the real source node selects the proxy source node before data transmission, it selects the node on the same or similar anonymous proxy path several times in a row. The network will be affected by the node receiving and forwarding the data. The generated traffic will be concentrated in a certain area for a period of time, which will make it easy for the adversary to guess the location of the real source node through traffic analysis. Therefore, in order to make the real source node evenly select a node on each anonymous proxy path and further resist traffic analysis attacks, we propose an enhanced anonymous communication strategy based on sector division.

For the candidate region established in Section 3.2, we define $R_{\min}$ to represent the distance between the selected lower limit and the real source node, and define $R_{\max}$ to represent the distance between the selected upper limit and the real source node. Then, we divide the candidate region into several equal sectors, each sector spans an angle μ , the total number of the sectors is $s = (2\pi/\mu)$, and we define these

FIGURE 4: Forward the data packet from RS to PS.

sectors as $area_1, area_2, \dots, area_\mu$. We can use the following equation to calculate μ .

$$\mu = \frac{\arcsin(R_{\min}/R_{\max}) + \arcsin(R_{\min}/H)}{\pi}, \quad (4)$$

which is used to calculate the value of the angle μ [28]. Among them, $R_{\min}$ is the radius of the visible region, $R_{\max}$ is the radius of the candidate region, and H is the number of the hops from the real source node to the sink. Then we can

FIGURE 5: Forward the data packet from PS to the sink.

calculate the total number of sectors s via μ . The candidate region divided into sectors is shown in Figure 6.

When selecting the proxy source node, we first randomly select an area $area_i$, where i falls in $[1, \mu]$. Then, we generate a random angle β , which is in the range of $[(i-1)\mu, i\mu]$. Finally,

we generate a random distance d , which satisfies the range $[R_{\min}, R_{\max}]$. The relative position of the selected proxy source node is $(x + d \cos(\beta), y + d \sin(\beta))$, where (x, y) are the coordinates of the real source node. Because the location of the proxy source node is randomly selected, we may not

FIGURE 6: Sector division in the candidate area. The angle of each sector is μ , and there are (s) sectors in total.

see any node in the desired region. If there is no node in the desired region, the last hop node routed to the selected location path becomes the proxy source node.

It is important to note that there may be duplicate nodes in these paths, but this will not affect our operations. When we select the proxy source node through the candidate region before each data transmission, it should alternately select the proxy source node from different sectors instead of the same sector. That is, when the real source node selects the node in the area $area_i$ ($i = 1, 2, \dots, \mu$) as the proxy source node in a packet transmission, it will not select nodes in the adjacent area of $area_i$ used as the proxy source node in the next packet transmission, and the node in the $area_i$ will not be selected as the proxy source node in the subsequent k ($k \leq (\mu/4)$) data packet transmissions.

In this way, the traffic in the network is evenly distributed in different areas within a period of time, instead of being concentrated in the same area, which makes it difficult for the adversary to track and guess the location of the real source node.

4. Results and Discussion

In this paper, we have proposed an anonymous communication scheme based on the proxy source node and the shortest path routing algorithm to protect source location privacy in the IoT. The scheme has the location privacy of the source node, anonymity, and can also prevent adversaries from collecting and analyzing communication messages in the whole network, and monitoring of network traffic in a certain region, such as impersonation and backtracking attacks. Table 3 compares the security performance of our scheme with Random Walk, GROW [35, 36], and ARPLP scheme.

4.1. Source Node Privacy. The scheme in this paper selects the proxy source node to replace the real source node to send data packets, and there is no direct connection between the proxy source node and the real source node. And the proxy source node is randomly selected from the candidate region, that is to say, there is also no direct

connection between the selected proxy source nodes each time. Therefore, for the adversary, they do not know any strategy about how the proxy source node is selected, and will not obtain the location information of the real source node through the proxy source node. Specifically, assume that two consecutive data transmissions use x and y as the proxy source nodes; however, because they are randomly selected, there is no connection between x and y . The adversary cannot judge the selection rule of the proxy source node through the two selections of the proxy source node.

At the same time, this scheme ensures that the selected proxy source node is far enough from the real source node through the selection mechanism of the proxy source node within the candidate region, thereby increasing the distance between the real source node and the proxy source node, and making the real source node have better privacy. In the process of establishing anonymous proxy paths, since the detection data packet is sent to the nodes that meet the energy requirements and are in the optional set, the number of hops from the next node to the real source node is greater than the number of hops from the current node to the real source node. In this way, after h hops, the probability that the proxy source node is less than $h/5$ hops from the real source node will be greatly reduced from the original $P = 1 - e^{-(h/25)}$.

In conclusion, this scheme can well hide the location of the real source node and protect the location privacy of the source node.

4.2. Anonymity. The scheme in this paper realizes the anonymity of the transmitted messages and the anonymity of the nodes in the wireless sensor networks.

Before the data packet transmission starts, the real source node and the base station jointly negotiate a DH shared key $K_{s,b}$, and only the real source node and the base station can know this key. The data to be transmitted are encrypted with this shared key, which forms the data packet with Q_i and is forwarded to the next hop. Passing through the proxy source node, the data arrive at the base station, and the base station uses the shared key to decrypt the data packet. When the data packet is transmitted from the real

TABLE 3: Comparison of security performance.

Scheme	Source location privacy	Anonymity	Anti-Impersonation	Backtracking attack
Random walk	No	No	No	No
GROW	Yes	No	No	Yes
ARPLP	Yes	Yes	No	Yes
Proposed	Yes	Yes	Yes	Yes

source node to the base station through the anonymous path we have established, each intermediate node does not know the shared key; therefore, others cannot decrypt the data packet and tamper with the content of the data packet, and the packet does not contain any information about the identity of the node.

At the same time, anonymity also includes the anonymity of the source node in the network and the anonymity of the communication relationship. The source node sends the information to the proxy source node through the anonymous proxy path, and then the proxy source node sends the information to the base station through the shortest path routing based on the residual energy. There is no identity information involved in the process. Each node only knows who its previous hop and next hop are, and does not know the source and destination of the information.

4.3. Anti-Impersonation Attack. Impersonation attacks refer to malicious nodes pretending to be legitimate nodes to forward messages, causing messages to be tampered with or interrupted in forwarding. In our scheme, it is not feasible for a node in the network to pretend to be a proxy source node. The reason is as follows:

When the real source node sends the data packet to the next node, the selected neighbor nodes must include the first hop after the source node ID recorded in the selected anonymous proxy path, and then the current node also selects the next node according to the ID recorded in the path. If the malicious node is not in the selected anonymous proxy path, then its ID will not appear in the selected anonymous proxy path; if the malicious node is in the selected initial path, but the real source node knows the ID of the selected proxy source node, it will not select the nodes with other hops on the path.

Specifically, if the real source node selects node p in the candidate region as the proxy source node before a message transmission starts. The node p is a node on the anonymous proxy path Q_r . Then, in the first stage of the anonymous message forwarding, the real source node forwards the data packet ($Q_r, E_{K_{sb}}(M)$) to a certain number of the neighbor nodes, containing the node represented by the first ID other than the real source node ID recorded in Q_r . If the malicious node is not in the selected Q_r , then its ID will not appear in Q_r and will not affect the transmission of the data packets; if the malicious node is in Q_r , since the real source node knows

ID_p on Q_r , it will not select the nodes with other hops on Q_r , but will stop until the data packet is transmitted to p .

4.4. Backtracking Attack. In wireless sensor networks, backtracking attack means that the adversary located near the base station will observe that the destination node receives the data information, and then start from the destination node and trace back hop-by-hop along the path until the source sensor node is found, which is the sender of the information. Our anonymous communication scheme can resist adversary backtracking attacks. The reason is as follows:

In our scheme, a candidate region formed by a flooding mechanism is set up, in which all nodes may become proxy source nodes and send messages instead of the real source nodes. In the process of message forwarding from the real source node to the proxy source node, there will be a lot of branch traffic to confuse the adversary, so the adversary can only trace back to the proxy source node, but cannot continue to trace back to find the location of the real source node.

We used PyCharm [37] to simulate the proposed scheme. For the energy consumption model presented in Section 3.1, the simulation results are as follows: Figure 7 shows the change trend of the energy consumed by the sender as the distance between the sender and the receiver changes when the number of the message bits transmitted are 50, 100, 150, and 200, respectively. We can see that as the distance between the sender and the receiver increases, the energy consumed by the sender, i.e., E_{sender} , is also increasing. The larger the number of bits, the more energy is consumed. Figure 8 shows the change trend of the energy consumed by the sender as the number of the message bits transmitted changes when the distance between the sender and the receiver is 50, 100, 150, and 200, respectively. It can be seen that as the number of message bits continues to increase, the energy consumed by the sender is also increasing. The greater the distance, the more energy is consumed.

Figure 9 shows the change trend of the energy consumed by the sender when the number of transmitted bits and the distance between the sender and the receiver simultaneously vary from 0 to 200.

At the same time, we simulated the relationship between the number of the transmitted message bits and the energy

FIGURE 7: The energy consumption of the sender varies with the transmission distance.

FIGURE 8: The energy consumption of the sender varies with the number of transmission bits.

FIGURE 9: The energy consumption of the sender varies with the transmission distance and the number of transmission bits.

consumed by the receiver in a segmented form. As shown in Figure 10, the number of the transmitted message bits is divided into four closed intervals, which are $[0, 50]$, $[51,$

$100]$, $[101, 150]$, and $[151, 200]$, and the energy consumed by the receiver in the four intervals is obtained. In each interval, as the number of message bits increases, the energy

FIGURE 10: Energy consumption diagram of received data.

consumed by the receiver will increase geometrically. At the same time, as the number of bits in each interval increases, the energy consumed by the receiver will also increase linearly.

5. Conclusion

In the context of the Internet of Things, while wireless sensor networks are widely used in various fields, they also face many security problems. Among them, the privacy protection of the source location is a very important security issue. In response to this problem, we proposed an anonymous communication scheme that protects the privacy of the source location in the IoT. By establishing a candidate region, the proxy source node is randomly selected in the candidate region to replace the real source node to send data packets, thereby achieving the purpose of protecting the location of the real source node. In the process of data packet transmission from the proxy source node to the sink, we used the shortest path routing algorithm based on the residual energy, so as to achieve the goal of saving energy and improving efficiency. But the work of this article does not involve the part of protecting the base station, i.e., all the nodes in the network know the location of the base station, and there is no specific application part. Therefore, our future work will focus on how to protect the location privacy of the base station while ensuring the privacy of the overhead and source location, and actively integrate it with practical applications.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was partly funded by EU Horizon 2020 DOMINOES Project (Grant Number: 771066).

References

- [1] N. Papernot, P. McDaniel, and A. Sinha, "Towards the science of security and privacy in machine learning," 2016, <https://arxiv.org/abs/1611.03814>.
- [2] Y. Wang, G. Yang, and T. Li, "Belief and fairness: a secure two-party protocol toward the view of entropy for IoT devices," *Journal of Network and Computer Applications*, vol. 161, Article ID 102641, 2020.
- [3] M. Lu, Z. Zhao, and X. Tang, "Research on source location privacy protection based on phantom routing," *Information Security and Technology*, vol. 3, no. 10, pp. 72–76, 2012.
- [4] J. Chen, B.-X. Fang, L.-H. Yin, and S. Su, "A source-location privacy preservation protocol in wireless sensor networks using source-based restricted flooding," *Chinese Journal of Computers*, vol. 33, no. 9, pp. 1736–1747, 2010.
- [5] P. Mohassel and Y. Zhang, "SecureML: a system for scalable privacy-preserving machine learning," in *Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP)*, pp. 19–38, IEEE, San Jose, CA, USA, May 2017.
- [6] K. Bonawitz, V. Ivanov, and B. Kreuter, "Practical secure aggregation for privacy-preserving machine learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, IEEE, Dallas, TX, USA, October 2017.
- [7] F. Li, D. Wang, and Yi. Wang, "Blockchain-based trust management in distributed internet of things," *Wireless Communications and Mobile Computing*, vol. 10, no. 1155, Article ID 8864533, 2020.
- [8] Y. Wang, Y. Wang, and Z. Wang, "Research cooperations of blockchain: toward the view of complexity network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 92, pp. 1–14, 2020.
- [9] Y. Wang, G. Yang, A. Bracciali et al., "Incentive compatible and anti-compounding of wealth in proof-of-stake," *Information Sciences*, vol. 530, pp. 85–94, 2020.

- [10] T. Li, Y. Chen, and Y. Wang, "Rational protocols and attacks in blockchain system," *Security and Communication Networks*, vol. 10, no. 1155, Article ID 8839047, 2020.
- [11] F. Li, R. Ge, and Y. Wang, "Tesia: a trusted efficient service evaluation model in IoT based on improved aggregation signature," *Concurrency and Computation: Practice and Experience*, vol. 10, no. 1002, 2020.
- [12] C. Zhao, S. Zhao, and M. Zhao, "Secure multi-party computation: theory, practice and applications," *Information Sciences*, vol. 476, pp. 357–372, 2019.
- [13] H. Wang, G. Han, and C. Zhu, "TCSLP: a trace cost based source location privacy protection scheme in wsns for smart cities," *Future Generation Computer Systems*, vol. 101, pp. 965–974, 2020.
- [14] X. Zhu, "Research on privacy protection strategy based on source node location in wireless sensor network," *Communication Power Technology*, vol. 37, no. 2, pp. 42–43, 2020.
- [15] G. Han, M. Xu, and Y. He, "A dynamic ring-based routing scheme for source location privacy in wireless sensor networks," *Information Sciences*, vol. 504, 2019.
- [16] S. D. Muruganathan, D. C. F. Ma, and R. I. Bhasin, "A centralized energy-efficient routing protocol for wireless sensor networks," *IEEE Communications Magazine*, vol. 43, no. 3, pp. S8–S13, 2005.
- [17] L. Mutalemwa and S. Shin, "Achieving source location privacy protection in monitoring wireless sensor networks through proxy node routing," *Sensors (Basel, Switzerland)*, vol. 19, no. 5, 2019.
- [18] B. Chakraborty and S. Verma, *Differentially Private Location Privacy Preservation in Wireless Sensor Networks*, Vol. 104, Springer US, New York, NY, USA, 2019.
- [19] J. Wang, R. Zhu, and S. Liu, "Node location privacy protection based on differentially private grids in industrial wireless sensor networks," *Sensors*, vol. 18, no. 2, 2018.
- [20] P. Spachos and D. Toumpakaris, "Angle-based dynamic routing scheme for source location privacy in wireless sensor networks," in *Proceedings of the 79th Vehicular Technology Conference*, IEEE, Seoul, Korea, May 2014.
- [21] Y. Liu and Y. Xu, "Source location privacy protection scheme based on variable included angle dynamic routing in WSNs," *Computer Application Research*, vol. 35, no. 1, pp. 257–260, 2018.
- [22] L. Bai, L. Li, and S. Qian, "Privacy protection algorithm based on expected phantom source node in wireless sensor network," in *2016 7th IEEE International Conference on Software Engineering and Service Science*, pp. 1006–1009, IEEE, Beijing, China, August 2016.
- [23] S. Li, Y. Xiao, L. Q. and Z. Qi, "A novel routing strategy to provide source location privacy in wireless sensor networks," *Wuhan University Journal of Natural Sciences*, vol. 21, no. 4, pp. 298–306, 2016.
- [24] L. Lin, *Research on Privacy Protection of Node Location and Data in Wireless Sensor Network*, Beijing University of Posts and Telecommunications, Beijing, China, 2015.
- [25] F. Li, C. Cui, and D. Wang, "Privacy-aware secure anonymous communication protocol in CPSS cloud computing," *IEEE Access*, vol. 10, no. 1109, Article ID 2982961, 2020.
- [26] S. Sharma and K. Ghosh, "Providing Privacy and security of Wireless sensor network using ACTOR nodes," *Control Theory and Informatics*, vol. 4, no. 4, 2014.
- [27] G. Tan, *Research on Privacy Protection of Source Node Location in Wireless Sensor Network*, Anhui University, Hefei, China, 2014.
- [28] Z. Zhao, Y. Liu, and F. Zhang, "Research on WSN source location privacy preserving routing based on angle and probability," *Journal of Shandong University (Science Edition)*, vol. 48, no. 9, pp. 1–9, 2013.
- [29] J. P. Sheu and J. R. Jiang, "Anonymous path routing in wireless sensor networks," in *Proceedings of the IEEE International Conference on Communications*, pp. 634–640, IEEE, Beijing, China, May 2008.
- [30] Y. Li and J. Ren, "Source-location privacy through dynamic routing in wireless sensor networks," in *2010 Proceedings IEEE INFOCOM*, IEEE, San Diego, CA, USA, March 2010.
- [31] J. Ren, Y. Zhang, and K. Liu, "An energy-efficient cyclic diversionary routing strategy against global eavesdroppers in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 151, 2013.
- [32] L. Zhou and Y. Shan, "An anonymous routing scheme for preserving location privacy in wireless sensor networks," in *Proceedings of the IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 10, no. 1109, pp. 262–265, IEEE, Chengdu, China, March 2019, Article ID 8728980.
- [33] F. Li, Z. Liu, and L. Li, "Privacy-aware PKI model with strong forward security," *International Journal of Intelligent Systems*, vol. 10, no. 1002, p. 22283, 2020.
- [34] C. Cui, F. Li, and T. Li, *Research on Direct Anonymous Attestation Mechanism in Enterprise Information Management Enterprise Information Systems*, vol. 55, pp. 1–17, 2019.
- [35] X. Yong and L. Schwiebert, "Preserving source location privacy in monitoring-based wireless sensor networks," in *Proceeding of 20th International Conference on Parallel and Distributed Processing Symposium (IPDPS'06)*, Island of Rhodes, Greece, January 2006.
- [36] L. Zhou and Q. Wen, "Providing location privacy against a global adversary in wireless sensor networks," *Journal of Information and Computational Science*, vol. 10, no. 15, pp. 5043–5053, 2013.
- [37] Q. Hu and L. Ma, "Deepgraph: a pycharm tool for visualizing and understanding deep learning models," in *Proceedings of the 2018 25th Asia-Pacific Software Engineering Conference (APSEC)*, IEEE, Nara, Japan, December 2018.