Hindawi
Journal of Sensors
Volume 2018, Article ID 4396381, 16 pages
https://doi.org/10.1155/2018/4396381

Research Article

A Clone Detection Algorithm with Low Resource Expenditure for Wireless Sensor Networks

Zhihua Zhang, Shoushan Luo, Hongliang Zhu, and Yang Xin

Beijing University of Posts and Telecommunications, Beijing, China

Correspondence should be addressed to Zhihua Zhang; zhangzhihua@bupt.edu.cn

Received 23 August 2017; Accepted 11 January 2018; Published 29 March 2018

Academic Editor: Jaime Lloret

Wireless sensor networks (WSNs) are facing the threats of clone attacks which can launch a variety of other attacks to control or damage the networks. In this paper, a novel distributed clone detection protocol with low resource expenditure is proposed for randomly deployed networks. The method consisting of witness chain establishment and clone detection route generation is implemented in the nonhotspot area of the network organized in a ring structure, which balances the resource consumption in the whole network. The witness chains and detection routes are in the centrifugal direction and circumferential direction, respectively, which can ensure the encounter of witnesses and detection routes of nodes with the same ID but different positions to detect clone attacks. Theoretical analysis demonstrates that the detection probability can be up to 1 with reliable witnesses. Moreover, both theoretical analysis and simulation results manifest that the proposed method can achieve better network lifetime and storage requirements with low resource expenditure and outperforms most methods in the literature.

1. Introduction

Wireless sensor networks (WSNs) usually consist of a large number of randomly distributed low-cost sensor nodes with limited resources in the target area. The networks have been widely used in various fields for the purpose of event monitoring and data gathering, including environment monitoring, forest fire monitoring, traffic data collection, and battlefield data gathering [1–3]. However, many WSNs are deployed in harsh or hostile environment which is a challenge for their secure operation. Due to the openness of wireless communication and lack of physical protection, sensor nodes are often compromised or attacked by attackers, making WSNs suffer from various attacks [4]. One of the most challenging attacks is clone attack or node replica attack, which refers to multiple nodes with the same ID. Since the sensor nodes are often unattended and lack of tamper-resistance devices, an attacker could capture a few nodes to obtain all the information materials in them including code and cryptographic mechanism. Hereafter, adversaries could duplicate the captured nodes. The cloned nodes seem to be legal ones for the network because they are exactly the same as the original ones, thus they could join the network freely without being recognized. It is much easier and cheaper to replicate a compromised node than to capture another normal node. Once the node is captured by an adversary, it could be replicated in large numbers and deployed in different areas of the network to jam or manipulate the network under the control of the adversary. Meanwhile, cloned nodes could initiate other inner attacks [5, 6], including selective forwarding attacks, black-hole attacks, energy exhaustion attacks, and data tampering attacks. With a certain number of cloned nodes that occupy strategic positions, the adversary may take over the whole network [7]. Therefore, it is essential to detect clone attacks effectively to avoid the serious harm to the network. Fortunately, the cloned nodes are often deployed in different positions, because it is not helpful to deploy them in the same location as the original one. It means that the cloned nodes and the original node have the same ID but different positions, which provides favorable conditions for clone detection.

Various methods for clone detection or node replica detection have been proposed up to now. According to different features, we could classify them into different

categories: witness-based or not, position dependent or not, centralized or distributed, the witnesses are deterministic or random, and the scheme is for randomly deployed or group deployed, which are detailed in Section 2. The algorithm we proposed is witness-based, location-dependent, distributed, with randomly selected witnesses, and it is for randomly deployed networks. There are some typical methods in the literature similar to our work, such as line-select multicast (LSM) protocol [8], randomized, efficient, and distributed (RED) protocol [9], energy-efficient ring-based clone detection (ERCD) protocol [10], and low-storage clone detection (LSCD) protocol [11]. In these schemes, some witnesses are selected randomly from the network to verify the legitimacy of nodes or detect the cloned nodes according to the private information (ID and location) reported to them. Hence, a clone is identified when at least two nodes possess the same ID but different locations.

However, these solutions have some drawbacks in two aspects. First, the clone detection probability is not high enough. Because all of these methods are distributed, that is, the witness and detection routes (legitimacy verification paths) are distributed, the clone is detected only if the witnesses and the detection routes encounter. Due to the randomness of witness selection and detection routes, the methods LSM, RED and ERCD could not ensure the encounter mentioned above. Whereas the LSCD could ensure the detection probability equal to 1 theoretically, because it adopts ring structure and forms witness arcs with a certain length, meanwhile, the detection routes are perpendicular to the arc (centrifugal) and the distance of each two adjacent detection routes is less than the arc length. Second, the resource consumptions of these methods are relatively high. The resource expenditure of LSM and RED is depending on the number of nodes in the network, and the resource expenditure increases significantly with the increase in the scale of the network. Although the resource expenditure of ERCD and LSCD has been improved to some extent, the resource consumptions are still at a high level which shortens the network lifetime.

In this paper, we propose an effective clone detection algorithm (referred to as CDLR) with a higher detection probability (equal to 1 theoretically) and a lower resource expenditure; meanwhile, the method CDLR avoids consuming the energy of nodes in the hotspot area, and all these measures prolong the network lifetime. Similar to the protocols ERCD and LSCD, a ring structure with the BS as the center is used in our work to ensure the encounter of the witnesses and the detection routes. Different from the two methods, the algorithm CDLR adopts random witness chains in the centrifugal direction (just as the radius of a circle) and detection routes in the circumferential detection, which ensures the encounter of witnesses and detection routes. Comparing with the ERCD with witness rings and circumferential verification paths and LSCD with witness arcs and multiple centrifugal detection routes, the CDLR has a lower resource expenditure and longer lifetime.

The major contributions of this work are as follows:

(1) The CDLR algorithm provides a high detection probability against clone attacks. In the ring structure, the random witness chains run through the entire nonhotspot area in the centrifugal direction, and the random detection routes or verification paths are formed along the circumference. Hence, the witness chains must encounter the detection routes, which ensures the detection probability equal to 1 theoretically.

(2) The resource expenditure of the CDLR algorithm is at a low level. The storage requirements of nodes with CDLR algorithm are not related to the density of nodes, and it almost does not increase with the increment in network scale. Furthermore, the communication load is lower than the similar method ERCD.

(3) The CDLR method makes full use of the energy of nodes in the nonhotspot area and prolongs the network lifetime. The CDLR fully used the residual energy of nodes in outer rings, that is, the nonhotspot area, because all the witness chains and detection routes are formed in outer rings, which effectively prolongs the network lifetime.

The rest of this work is organized as follows. Section 2 reviews the previous related works. Section 3 presents the network model and assumptions. The method CDLR for clone detection is proposed in Section 4. Then in Section 5, the theoretical performance analysis is conducted. Experiments and simulations are given in Section 6. Finally, Section 7 concludes this paper.

2. Related Works

Clone attacks have attracted the attention of researchers, and there has been much effort on clone detection up to now [5–18]. According to different features, we could classify them into different categories: centralized [12–15] or distributed [5–11, 16–18], witness-based [5, 8–11, 16–18] or not, position dependent [5, 8–11, 17, 18] or not, and the scheme is for randomly deployed [5–16] or group-deployed networks [17, 18].

The most common classification in the literature is based on centralized or distributed. For the centralized methods, the BS or sink is responsible for clone detection according to the information reported by nodes [12–15]. The advantages of these methods are that they have low overhead and high detection probability because of the comprehensive information. However, the shortcomings are also explicit: the BS easily suffers from a single point of failure and the nodes around the BS consume much more energy than others due to forwarding packets. In order to solve these problems, distributed schemes are proposed [5–11, 16–18], which assign the detection tasks to different areas and nodes, yet the resource consumptions of nodes are increasing sharply. Most of the works are conducted to balance the detection probability and the resource expenditure.

There are also another categories: according to witness requirements, the schemes are divided into witness-based and no witness-based; based on the location requirements,

Table 1: Categories of different clone detection schemes.

Scheme	Detection mechanism		Witness requirements		Location requirements		Network deployment	
	Centralized	Distributed	Witness-based	No witness	Location dependent	Location independent	Randomly deployed	Group deployed
[5]		√	√		√		√	
[7]		√		√		√	√	
[8]		√	√		√		√	
[9]		√	√		√		√	
[10]		√	√		√		√	
[11]		√	√		√		√	
[12]	√			√		√	√	
[13]	√			√		√	√	
[14]	√			√		√	√	
[15]	√			√		√	√	
[16]		√	√			√	√	
[17]		√		√	√			√
[18]		√	√		√			√

the methods are classified into location dependent and location independent; on the basis of network deployment requirements, the protocols are classified into for randomly deployed and for group deployed. All of the categories mentioned above are listed in Table 1.

From Table 1, we could see that most of the distributed methods are witness-based, location-dependent, and for randomly deployed networks. According to the practical application and our work, we focus on the methods with these conditions [5, 8–11].

Randomized multicast (RM) and line-selected multicast (LSM) were proposed in [8]. Both methods are witness-based, and LSM is an improvement of RM. In both methods, the neighbors of each node randomly select a fraction of nodes as its witnesses. Differences are that the clone detection is conducted according to the birthday paradox problem in RM, that is, at least one witness will discover the conflict of nodes with the same ID but different locations; whereas the nodes along the routes from the node to its random witnesses are also selected as witnesses, thus the intersections of different routes could improve the detection probability. However, the detection probability is still in a low level, and the resource consumptions are closely related to the number of nodes, which is not suitable for large-scale networks.

Randomized, efficient, and distributed protocol (RED) was put forward in [9], which is another improvement of RM. It is also witness-based, and some witnesses are randomly selected by the neighbors of the node according to its ID, that is, the witness set of nodes with the same ID will be identical, thus the detection probability is improved. However, the "randomly" selected witnesses of a node based on node ID are always the same ones, which means the witness selection is deterministic in fact, and it is easy to be exploited by adversaries. Besides, the storage overhead is still related to the number of nodes, which is also not suitable for the networks with a large number of nodes.

Similar to LSM, a random walk (RAWL) protocol was proposed in [5], which improves the detection probability by a random walk of all randomly selected witnesses expanding the scope of witnesses. Both the detection probability and the resource consumptions are improved, but the storage overhead is still related to the scale of network.

In order to balance the detection probability and resource consumptions, some other protocols are proposed. To the best of our knowledge, the best two methods are energy-efficient ring-based clone detection (ERCD) protocol and low-storage clone detection (LSCD) protocol described in [10, 11], respectively. Both methods are random witness-based and suitable for large-scale network, which adopt ring structure with the BS (or sink) as the center. The clone detection process in both protocols consists of two phases: witness selection and legitimacy verification (or detection route generation).

In ERCD [10], the witnesses of each node form a witness ring in a randomly selected network ring, and a witness header which is responsible for clone detection is selected randomly from all the witnesses. The legitimacy of each node has to be verified by ERCD before communicating with others. The verification message is transmitted to its corresponding witnesses, and it is broadcasted to witness header in the witness ring and neighbor rings. The clone attack will be detected if the witness header discovers a conflict of nodes with the same ID but different locations in reported verification messages. Theoretical analysis shows that the ERCD has a high detection probability and a constant level of storage overhead. However, the clone can be detected only if the witnesses of two replica nodes are in the same ring or in the neighbor rings. Besides, in the process of witness selection and legitimacy verification, each message is forwarded along the ring, which causes a high communication overhead.

In the process of witness selection of LSCD [11], the witnesses of each node form a witness arc with a certain length in a randomly selected network ring. In detection process, some centrifugal detection routes that are perpendicular to the witness arc are generated from the second ring. The

distance between two adjacent routes is less than the witness arc length to ensure the encounter of witnesses and the detection routes. During the establishment of detection routes, it is necessary to check the relationship between the arc length and the adjacent route spacing at any time to increase the detection routes. Thus, the area far from the center has more detection routes. The dynamic mechanism in detection route establishment ensures the high detection probability, and the storage overhead of nodes is relatively low. However, the messages for detection route establishment of all nodes need to be forwarded to the second ring to initiate the detection route generation, which will consume much energy of nodes in the second ring. Hence, the energy consumptions of these nodes in the hotspot area will become the bottleneck of the network lifetime.

Unlike prior works, the algorithm we proposed has a high detection probability (equal to 1 theoretically) and low resource consumptions, which avoids consuming the energy of nodes in the hotspot area and makes full use of the resource of nodes in the nonhotspot area. The random witness chains are formed along the radius of network to run through the whole nonhotspot area. Meanwhile, the circumferential detection routes are generated in randomly selected rings to ensure the encounter of witnesses and detection routes. Note that all these witness chains and detection routes are established in the nonhotspot area, which prolongs the network lifetime effectively. The implementation of CDLR does not introduce a new bottleneck of network lifetime. Besides, different from the most previous works, the storage requirements using CDLR are not relative to the density of nodes in the network.

3. System Model and Assumptions

In this section, the network model, the adversary model, and assumptions are introduced.

3.1. Network Model. Based on ERCD and LSCD, we propose a clone detection algorithm with low resource expenditure (CDLR). Similar to ERCD and LSCD, the ring structure of network with the BS as the center is adopted in our work. In this model, the BS is at the center of the network region, and all nodes are densely and randomly distributed around the BS. The communication radius of each node is r, and the radius of the whole network denoted as $R = hr$. The density of nodes is ρ. Each node has its own relative location to the BS (the hops between a node and the BS) and its 1-hop neighbors; moreover, each node in the network knows its own geographic position by any mature mechanism introduced in [19, 20]. Take the BS as the center, the whole network is virtually divided into concentric circles (or rings), and the width of each ring is r, equal to the communication radius of each node. The identification of rings from inside to outside is from 1 to h. The rings near the BS are considered as hotspot due to heavy traffic load. The nonhotspot area is set as k rings from $h-k+1$ to h.

Besides, each node in the network is stationary and has a unique ID. Nodes with new ID are not permitted to join the network after the deployment finished. The communication between BS and nodes is through the other intermediate nodes. The nodes report their observing data to the BS periodically. The model of the network is shown in Figure 1.

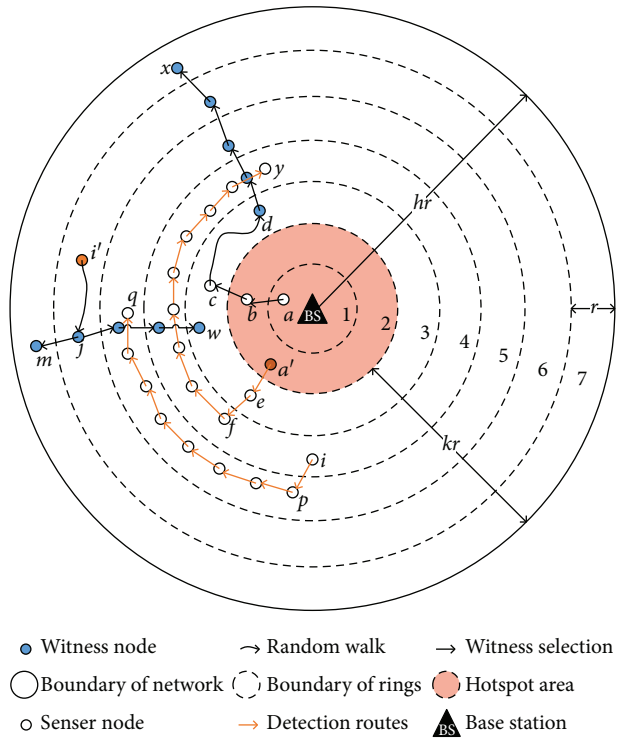


Figure 1: The model of a ring structure WSN.

The information transmitted between nodes are encrypted through a conventional bootstrapping cryptography mechanism. Before deployment, a key pair (ID, secret key) is allocated to each node. All nodes share their ID in the whole network, and the BS is assumed to be secure enough.

3.2. The Adversary Model. The purpose of the adversary is to damage or control the network at a lower price. It is cheaper to duplicate some nodes than to compromise the same amount of nodes. Therefore, the adversary usually compromises only a few sensor nodes, obtains all the information from captured nodes, and then replicates them to a large amount. Henceforth, the adversary could deploy the replica nodes into the network at strategic positions to acquire much more information or damage the network. The replica nodes deployed in different areas of network seem to be normal ones because they have legitimate IDs and encryption authentication mechanism.

We assume that the adversary could capture a limited number of normal sensor nodes at any position and the adversary could not create new IDs to join the network. Meanwhile, the adversary tries to avoid being detected [9].

3.3. Assumptions. We summarize the assumptions used in this work as follows:

(i) For the network

(A1) The nodes in the network are stationary and distributed densely and randomly.

(A2) Each node has a unique ID, and new IDs are forbidden after the network deployed.

(A3) Each node knows its own and neighbors' relative locations and geographic positions.

(A4) The communication between nodes is encrypted through a conventional bootstrapping cryptography mechanism.

(A5) The BS cannot be compromised.

(ii) For the adversary

(A1) The adversary could capture a limited amount of normal nodes.

(A2) The adversary deploys the replica nodes into different positions.

(A3) The adversary could not create new IDs for replica nodes.

The aim of our work is to discover a scheme with higher detection probability and lower resource expenditure; thus, the network lifetime will be maximized.

4. The Proposed Clone Detection Algorithm

In this section, the proposed method, a clone detection algorithm with low resource expenditure (CDLR) is introduced, which has a high detection probability (equal to 1 theoretically) and low resource consumptions and prolongs the network lifetime efficiently.

4.1. Overview of the Proposed CDLR Algorithm. The proposed CDLR approach is a distributed and random witness-based method for WSNs, which consists of two phases: the random witness chain establishment and the clone detection route generation. In order to avoid consuming energy of nodes in the hotspot area, the nonhotspot area should be predefined before deployment. The CDLR is implemented in the nonhotspot area to prolong the network lifetime. Each node in nonhotspot is responsible for both data collection and clone detection. The node should be checked by CDLR procedure before transmitting information to others.

In the first phase, a random witness chain for a node is generated by the node sending its encrypted private information including ID and location to the witness nodes. The witness chain of the node runs through the whole nonhotspot area in the direction of network radius, that is, in the centrifugal direction, and each witness node stores the private information of the source node.

In the second phase, a detection route for the node is established in a ring that is randomly selected from the nonhotspot area, and the route is formulated along the ring, that is, in the circumferential direction. The detection information is broadcast in the ring, and the clone is detected when the witness node encountering the detection route discovers the conflict of nodes with the same ID but different positions.

The process of CDLR is just shown in Figure 1. It is implemented in the nonhotspot area, that is, the outer k rings. The black arrows demonstrate the witness selection and the witness chain generation, and the blue solid circles represent the witnesses of a node. Similarly, the red arrows indicate the clone detection route establishment, and the red solid circles represent the cloned nodes. The intersection of the witness chain and detection route of the nodes with the same ID demonstrates the clone detection.

In order to facilitate understanding our approach, see Notations for the list of symbols used in this work.

4.2. Witness Chain Establishment. The witness chain establishment is slightly different according to different locations of nodes. Suppose that the whole network is divided into h virtual rings and the nonhotspot area is defined as the outer k rings, that is, the hotspot area is $h-k$ rings near around the BS, just as shown in Figure 1. The witness chain formulation for nodes in hotspot area and nonhotspot area is described in this section.

For nodes in the hotspot area, the witness chain establishment consists of three steps. *Step 1*. The node transmits the encrypted witness selection message (ID, position) to any node in the nearest ring of nonhotspot area, that is, the $(h\text{-}k+1)$th ring. In this transmission process, the nodes on the forwarding path do not need to store the forwarded message, because they are not the witnesses of the source node. *Step 2*. In order to confuse the adversary, the node in the $(h\text{-}k+1)$th ring does not generate a witness chain directly, instead, it forwards the received message randomly to another node in the same ring, that is, the node randomly walks a few hops denoted as ξ. This node is selected as the first witness and stores the received message. *Step 3*. The first selected witness node initiates the witness chain generation by forwarding the message to the node in its communication range and in the next outer ring. The node that received the message forwards it to the node in the next outer ring until the outermost ring, that is, the hth ring. The nodes on the path of message transmitting formulate the witness chain of the source node. Take node a in Figure 1 as an example, node a locates in the hotspot area of the network, firstly, it should sent its message to node c in the nearest ring of nonhotspot area through node b; then, node c randomly walks ξ hops to node d in the same ring; at last, node d stores and forwards the message to nodes in the outer ring until reaching the node x in the hth ring. The nodes from d to x formulate the witness chain of node a, and the nodes on the chain store the message from node a.

For nodes in the nonhotspot area, the witness chain formulation consists of two steps. They do not need to send messages out of hotspot because they are already in the nonhotspot area. *Step 1*. The node randomly walks ξ hops to another node in the same ring, just like step 2 described in the last paragraph. *Step 2*. The first selected witness node forwards the message in the centrifugal and centripetal directions until reaching the outermost and innermost ring of the nonhotspot area, respectively. Take node i' in Figure 1

as an example, node i' randomly walks ξ hops to node j in the same ring, then node j stores and forwards the message in the centrifugal direction to node m and in the centripetal direction to node w, respectively. The nodes from m to w formulate the witness chain of node i'.

Through the process described above, the random witness chains for nodes are generated. The length of the witness chain is equal to the width of nonhotspot area, which is in a small constant level for the network.

4.3. Detection Route Generation. The detection route generation is also implemented in the nonhotspot area, because the witness chains formulated by the first phase are in the nonhotspot area. Only if the witness chains and the detection routes of nodes with the same ID are encountering, will the clone attacks be detected.

The detection route generation comprises two steps for all nodes in the network. *Step 1*. The node randomly selects a ring in the nonhotspot area to generate detection route and sends the encrypted detection message (ID, position) to any node in the selected ring. *Step 2*. The node received the detection message broadcasts it in the same ring. The witnesses of the source node compare the message they have stored with the received detection message from the same ID, if a conflict of nodes with the same ID but different positions occurs, the clone attack will be detected. Therefore, the revocation procedure for cloned nodes is triggered, and the ID and positions are the evidence.

Take nodes a and a' in Figure 1 as examples. Suppose node a' is a clone of node a, and they are located at different positions. Node a and a' have their own witness chains separately. Node a' first selects a random ring from nonhotspot (the 4^{th} ring in Figure 1) and sends its detection message to node f in the 4th ring. Then, node f broadcasts the received message to the nodes in the same ring. The clone is detected when the detection message of node a' encounters the witness chain of node a, because the witness discovers the conflict that node a has different positions in the network.

From the description above, the random witness chains in the direction of network radius and the detection routes in the circumferential direction must encounter, which ensures the high detection probability. Furthermore, the process of witness selection and clone detection is implemented only in the nonhotspot area, which avoids consuming energy of nodes in hotspot and prolongs the network lifetime.

4.4. The CDLR Algorithm Description. The procedures of CDLR algorithm depicted above are summarized in Algorithm 1.

5. Theoretical Performance Analysis

The performance of CDLR is evaluated and analyzed from the aspects of detection probability, communication load, and storage requirements theoretically.

5.1. Detection Probability Analysis. The clone detection probability refers to whether any witness of a node can discover at least two nodes with the same ID but different positions (if exists) or not. If there exist multiple replica nodes for a source node, the clone attack will be detected successfully when one of the replica nodes is discovered, because all of these replica nodes have the same ID.

Theorem 1. *Given that the randomly selected witnesses of a source node are not compromised, if there exist replica nodes of this source node, the cloned nodes could always be detected successfully.*

Proof. To make the cloned nodes be detected, one of the following conditions should be met:

(1) At least one of the witnesses of the source node encounters one of the detection routes of cloned nodes.

(2) At least one of the witnesses of all cloned nodes encounters the detection route of the source node.

The CDLR algorithm takes some measures to ensure the encounter of the witness chains and the detection routes. First of all, the witness chain of each node (including cloned nodes) runs through the whole nonhotspot area, that is, all the witness chains are in the direction of network radius or in the centrifugal direction and extend to the network boundary. Secondly, the detection route of each node (also including cloned nodes) is generated along one of the virtual rings randomly selected from nonhotspot area by broadcasting the detection message, that is, all the detection routes are in the circumferential detection which are perpendicular to the witness chains. Thus, the detection routes must encounter the witness chains, that is, the witness chain of the source node must encounter the detection routes of cloned nodes and vice versa. Hence, one of the two conditions above must be met during the detection. When the witness chains and detection routes encounter, the conflict of nodes with the same ID but different locations will occur, and the cloned node could always be detected.

5.2. Communication Load and Network Lifetime Analysis. Due to the limited resources of sensor nodes, especially the limited energy supplied by batteries, the network is sensitive to resource consumptions. The energy consumption is related to the communication load of nodes. Moreover, the resource expenditure has a significant impact on the network lifetime. It is necessary to decrease the resource expenditure of nodes to prolong the network lifetime. Here, the network lifetime is defined as the duration from the network deployment to the moment that any node exhausts its energy [10].

In this section, the performance of CDLR is evaluated in terms of communication load of nodes and the network lifetime. The communication load of each node is analyzed at first, and then the network lifetime can be calculated from

```
Initialization:
Preset encryption mechanism
Obtain the relative locations to the BS and the identification of each ring
Predefine the nonhotspot area width k
Exchange the relative information with neighbors
PHASE 1: Witness chain establishment
X_a=Encrypt (ID_a, l_a)
i = R_a
while i < h − k + 1 do
    Forward X_a to the i + 1 ring
    i + 1
end while
if i ≥ h − k + 1 do
    Forward X_a ξ hops randomly to node b in the same ring
    node b records (ID_a, l_a) in X_a
    W_a ← b
    for j = i; j > h − k + 1; j– do
        Forward X_a to node x in j-1 ring
        node x records (ID_a, l_a) in X_a
        W_a ← x
    end for
    for u = i; u < h; u++ do
        Forward X_a to node x in u + 1 ring
        node x records (ID_a, l_a) in X_a
        W_a ← x
    end for
end if
PHASE 2: Detection route generation
X_a=Encrypt (ID_a, l_a)
i = R_a
if i < h − k + 1 do
    Select a ring R_w in nonhotspot randomly
    while i < R_w do
        Forward X_a to node x in i + 1 ring
        i + 1
    end while
end if
node x broadcast X_a in the i^th ring
for each node w in W_a that hears X_a do
    if (ID_a, l_a) stored in w ≠ (ID_a, l_a) in X_a do
        Trigger revocation procedure
    end if
end for
```

Algorithm 1: A clone detection algorithm with low resource expenditure.

it by the given maximum communication load of each node in its whole lifecycle. In the evaluation, we assume that the communication load of each node in the same ring is the same. The outer (inner) rings of ring i refer to the rings whose identification is larger (smaller) than i.

The communication load of each node consists of three parts: witness selection, clone detection, and observing data collection, which is expressed by Theorems 2, 3, and 4, respectively.

Theorem 2. *Let ε_w, f_w, and ξ represent the size of the request message for witness selection, the frequency of witness selection, and the number of random walk hops, respectively. The communication load for witness selection of each node in ring i, $i \in [1, h]$, can be expressed as*

$$C_i^{\mathrm{w}} = \begin{cases} \dfrac{i^2 \varepsilon_{\mathrm{w}} f_{\mathrm{w}}}{2i-1}, & i < h-k+1, \\ (\xi+1)\dfrac{i^2 \varepsilon_{\mathrm{w}} f_{\mathrm{w}}}{2i-1}, & i = h-k+1, \\ \left(\xi + 1 + \dfrac{h^2}{2i-1}\right)\varepsilon_{\mathrm{w}} f_{\mathrm{w}}, & i > h-k+1. \end{cases} \quad (1)$$

Proof. The communication load of each node for witness selection is different according to its location in the network,

that is, the node is located in the hotspot area or nonhotspot area. We assume that the width of nonhotspot area is k in hop counts. If the nodes are located in the hotspot area, they will be responsible for forwarding the request messages from nodes in the inner ring and themselves to nodes in the outer ring. The number of nodes in the ring i and its inner rings is $\pi i^2 r^2 \rho$, and the communication volume is $\pi i^2 r^2 \rho \times \varepsilon_w f_w$. Since the number of nodes in ring i can be expressed as

$$\pi i^2 r^2 \rho - \pi (i-1)^2 r^2 \rho = \pi (2i-1) r^2 \rho. \quad (2)$$

The communication load for witness selection of each node in ring i is

$$C_i^w = \frac{\pi i^2 r^2 \rho \times \varepsilon_w f_w}{\pi (2i-1) r^2 \rho} = \frac{i^2 \varepsilon_w f_w}{2i-1}, \quad i < h-k+1. \quad (3)$$

If a node is located in the nonhotspot area, there are two cases based on the different locations and responsibilities, that is, the node is located in the innermost ring or other rings of the nonhotspot area.

For the nodes in the innermost ring of the nonhotspot area, they should first forward the request messages from nodes in the inner ring and themselves ξ times in the same ring because of the random walk mechanism and then forward these request messages to nodes in the outer ring. Thus, the communication volume of nodes in ring i is $\pi i^2 r^2 \rho \times \varepsilon_w f_w \xi + \pi i^2 r^2 \rho \times \varepsilon_w f_w$, and the number of nodes in ring i is obtained according to (2); therefore, the communication load for witness selection of each node in ring i is

$$C_i^w = \frac{(\xi+1)\pi i^2 r^2 \rho \times \varepsilon_w f_w}{\pi (2i-1) r^2 \rho} = (\xi+1)\frac{i^2 \varepsilon_w f_w}{2i-1}, \quad i = h-k+1. \quad (4)$$

For the nodes in other rings except the innermost ring of the nonhotspot area, they have to randomly walk ξ hops and select witnesses in the centrifugal and centripetal directions. Therefore, they have three responsibilities: (1) forward their own messages ξ times in the same ring because of the random walk mechanism; (2) forward the messages from nodes in the inner ring and themselves to nodes in the outer ring; and (3) forward the messages from nodes in the outer ring and themselves to nodes in the inner ring. The communication load of each node for responsibility 1 is

$$\frac{\pi (2i-1) r^2 \rho \varepsilon_w f_w \xi}{\pi (2i-1) r^2 \rho} = \varepsilon_w f_w \xi. \quad (5)$$

The communication load of each node for responsibility 2 is the same as (3), and that for responsibility 3 is

$$\frac{\pi r^2 \left[h^2 - (i-1)^2\right] \rho \varepsilon_w f_w}{\pi (2i-1) r^2 \rho} = \frac{\left[h^2 - (i-1)^2\right] \varepsilon_w f_w}{2i-1}. \quad (6)$$

Thus, the communication load for each node in other rings except the innermost ring of the nonhotspot area is obtained by (3), (5), and (6) as

$$C_i^w = \varepsilon_w f_w \xi + \frac{i^2 \varepsilon_w f_w}{2i-1} + \frac{\left[h^2 - (i-1)^2\right] \varepsilon_w f_w}{2i-1} = \left(\xi + 1 + \frac{h^2}{2i-1}\right) \varepsilon_w f_w, \quad i > h-k+1. \quad (7)$$

Overall, the communication load for witness selection of each node in ring i can be expressed in (1).

The communication load of each node for clone detection is described as follows.

Theorem 3. *Let ε_c and f_c denote the size of the request message for clone detection and the frequency of clone detection, respectively. The communication load for clone detection of each node in ring i, i∈[1, h], can be expressed as*

$$C_i^c = \begin{cases} \dfrac{i^2 \varepsilon_c f_c}{2i-1}, & i < h-k+1, \\ \dfrac{\pi i h^2 \varepsilon_c f_c}{k(2i-1)}, & i \geq h-k+1. \end{cases} \quad (8)$$

Proof. The communication load of each node for clone detection is different according to different locations, that is, the node is located in the hotspot area or nonhotspot area. For nodes in the hotspot area, they only need to forward the detection messages from nodes in the inner ring and themselves to nodes in the outer ring, and this communication load is similar to (3), which can be expressed as

$$C_i^c = \frac{\pi i^2 r^2 \rho \times \varepsilon_c f_c}{\pi (2i-1) r^2 \rho} = \frac{i^2 \varepsilon_c f_c}{2i-1}, \quad i < h-k+1. \quad (9)$$

For nodes in the nonhotspot area, we consider the following: because the clone detection is implemented in the nonhotspot area, the clone detection messages generated by all nodes in the network should be transmitted to the nonhotspot area, that is, all the k rings, thus each ring i in the nonhotspot area takes $1/k$ of the whole detection messages. Then, the messages are broadcasted in ring i to encounter the witness chains. Thus, the number of detection messages that ring i should take is $\pi h^2 r^2 \rho \varepsilon_c f_c / k$, and these messages are broadcasted in ring i for πi times in an average. Therefore, the communication load of each node in nonhotspot area for clone detection is

$$C_i^c = \frac{\pi h^2 r^2 \rho \varepsilon_c f_c}{k\pi (2i-1) r^2 \rho} \times \pi i = \frac{\pi i h^2 \varepsilon_c f_c}{k(2i-1)}, \quad i \geq h-k+1. \quad (10)$$

Overall, the communication load of each node for clone detection can be expressed as (8).

The original mission of the network is to collect the observing data periodically, and the communication load of each node in ring i for collecting data is calculated as follows.

Theorem 4. *Let ε_d and f_d denote the size of the message for observing data and the frequency of observing data collection, respectively. The communication load for observing data collection of each node in ring i, i ∈ [1, h], can be expressed as*

$$C_i^{\mathrm{d}} = \frac{\left[h^2 - (i-1)^2\right]\varepsilon_{\mathrm{d}} f_{\mathrm{d}}}{2i-1}. \quad (11)$$

Proof. In the process of normal data collection, the observing data is transmitted to the BS through the intermediate nodes; thus, the nodes in ring i are responsible for transmitting the messages of their own and from the outer rings to the nodes in the inner ring. The number of nodes in ring i and its outer rings is $\pi r^2[h^2 - (i-1)^2]\rho$, and the number of nodes in ring i can be obtained by (2). Therefore, the communication load of each node for observing data collection is

$$C_i^{\mathrm{d}} = \frac{\pi r^2\left[h^2 - (i-1)^2\right]\rho\varepsilon_{\mathrm{d}} f_{\mathrm{d}}}{\pi(2i-1)r^2\rho} = \frac{\left[h^2 - (i-1)^2\right]\varepsilon_{\mathrm{d}} f_{\mathrm{d}}}{2i-1}. \quad (12)$$

According to Theorem 2, 3, and 4, the overall communication load for each node in ring i can be obtained by

$$C_i^{\mathrm{o}} = C_i^{\mathrm{w}} + C_i^{\mathrm{c}} + C_i^{\mathrm{d}}. \quad (13)$$

If the parameters of ε_{w}, f_{w}, ε_{c}, f_{c}, ε_{d}, f_{d}, and ξ are known for the network, the optimal k will be acquired based on (13) to maximize the network lifetime. Figure 2 demonstrates the communication load of each node in ring i with different width of nonhotspot area k, from which we can see that the parameter k has a significant impact on the communication load of nodes; thus, it can also affect the energy consumption and network lifetime. Here in Figure 2, under the condition that $h = 20$, $f_{\mathrm{c}} = 10$, and other parameters are set to 1, when k is 19, 18, and 17, the communication load of nodes in rings 2, 4, and 5 is the highest, that is, consumes the most energy, respectively. From the evaluation of the communication load, we can obtain the optimal k.

Under the same conditions, a comparison of the communication load of nodes in ring i with different k using CDLR and ERCD is conducted according to (13) and the evaluation described in [10]. As shown in Figure 3, no matter what value is assigned to k, the corresponding communication load of each node in ring i using CDLR is lower than that using ERCD. The most straightforward reason is that the communication load of nodes in witness section using CDLR is much lower than that using ERCD, because the witness selection of ERCD is along the circumferential direction whereas that of CDLR is along the radius of the network; thus, the number of witnesses of ERCD is much more than that of CDLR.

Based on the communication load evaluation above, the network lifetime using CDLR is analyzed and compared with the lifetime using LSM [8] and ERCD [10] under the same conditions. According to the definition of network lifetime, the energy exhaustion of any node means the end of network lifetime in the evaluation. To simplify the evaluation, we assume that the communication capability of each node in network is the same and the total size of messages a node can transmit in its life cycle denoted as T_z is 1 million. In the evaluation, the

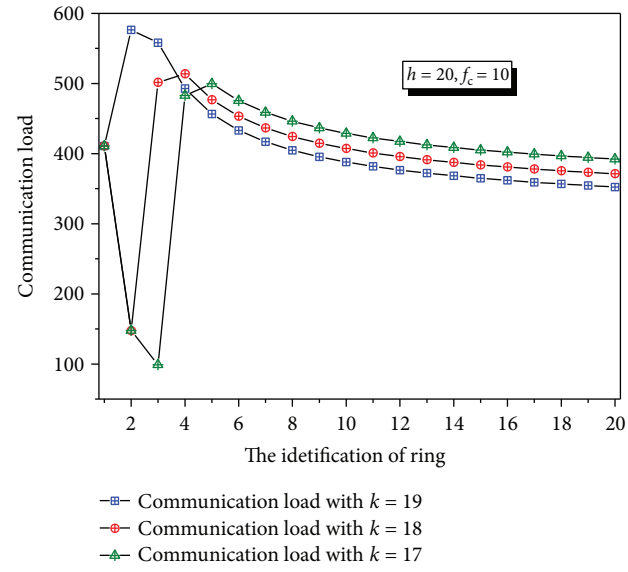


Figure 2: The communication load of each node in ring i with different k.

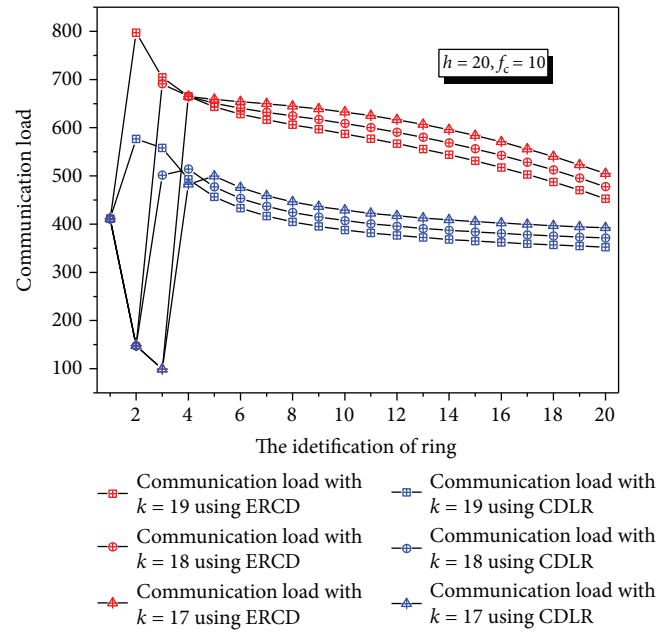


Figure 3: The comparison of communication load of each node using ERCD and CDLR.

parameters used are assumed as $\varepsilon_{\mathrm{w}} = \varepsilon_{\mathrm{c}} = \varepsilon_{\mathrm{d}} = 1$, $f_{\mathrm{w}} = f_{\mathrm{d}} = 1$, and $\xi = 1$. Meanwhile, in LSM, the variable g denotes the number of witnesses selected by each neighbor of a source node, p represents the probability that the neighbor transmits the message from the source node, and δ is the average degree of each node or the average number of neighbors of each node.

Theorem 5. *If $\varepsilon_w = \varepsilon_c = \varepsilon_d = 1$, $f_w = f_d = 1$, and $\xi = 1$, the ratio of network lifetime using CDLR algorithm and LSM algorithm is*

$$\frac{h^2 + gp\delta h f_c\sqrt{\delta+1}}{\max\left(\left(1+f_c+h^2\right), \max\left(\left(\left(\pi i h^2 f_c/3k\right)+\left(h^2+i^2/2i-1\right)+1\right), \max\left(\left(\pi i h^2 f_c/3k\right)+\left(2h^2-i^2/2i-1\right)+3\right)\right)\right)}. \quad (14)$$

Proof. In LSM, the communication load consists of observing data collection and clone detection, because the witness selection and clone detection are the same process. In observing data collection, the communication load of each node in ring i can be obtained by (11), from which we can see that the communication load of nodes in ring 1 is the maximum, that is, h^2. Thus, the communication load of nodes in ring 1 determines the network lifetime. The communication load for clone detection in ring 1 is $gp\delta\sqrt{n}$, where n is the number of nodes in the whole network and $n = \pi h^2 r^2 \rho$. All nodes in ring 1 are neighbors to each other; thus, the number of nodes in ring 1 is equal to the number of neighbors of a node plus itself, that is, $\pi r^2\rho = \delta + 1$. Therefore, the overall communication load of each node in ring 1 is $h^2 + gp\delta h f_c\sqrt{\delta+1}$, and the network lifetime using LSM is expressed as

$$\mathrm{LT}_{\mathrm{LSM}} = \frac{T_z}{h^2 + gp\delta h f_c\sqrt{\delta+1}}. \quad (15)$$

In CDLR algorithm, suppose the node in ring i has the maximum overall communication load. Due to the difference of nodes in the hotspot area and nonhotspot area, the maximum communication load is different according to the locations of nodes. If the node is located in the hotspot area, that is, $i < h-k+1$, the maximum communication load is the node in ring 1 calculated as $1+f_c+h^2$, where 1, f_c, and h^2 are the communication load for witness selection, clone detection, and observing data collection, respectively. If the node is in the nonhotspot area, there are two cases for communication load calculation according to whether the node is in the innermost ring of nonhotspot area or not. For the node in the innermost ring of nonhotspot area, that is, the node in ring $i = h-k+1$, the overall communication load is calculated according to (13) as

$$C_i^{\mathrm{o}} = \frac{2i^2}{2i-1} + \frac{\pi i h^2 f_c}{3k} + \frac{h^2-(i-1)^2}{2i-1}, \quad i = h-k+1. \quad (16)$$

For the node in other rings except the innermost ring of nonhotspot area, that is, $i > h-k+1$, the overall communication load is calculated according to (13) as

$$C_i^{\mathrm{o}} = 2 + \frac{h^2}{2i-1} + \frac{\pi i h^2 f_c}{3k} + \frac{h^2-(i-1)^2}{2i-1}, \quad i > h-k+1. \quad (17)$$

The network lifetime is determined by the lifetime of the node in ring i which has the maximum communication load; therefore, the network lifetime using CDLR is expressed as

$$\mathrm{LT}_{\mathrm{CDLR}} = \frac{T_z}{\max\left(\left(1+f_c+h^2\right), \left(\left(2i^2/2i-1\right)+\left(\pi i h^2 f_c/3k\right)+\left(h^2-(i-1)^2/2i-1\right)\right), \max\left(2+\left(h^2/2i-1\right)+\left(\pi i h^2 f_c/3k\right)+\left(h^2-(i-1)^2/2i-1\right)\right)\right)}. \quad (18)$$

Based on (18) and (15), the ratio of network lifetime using CDLR algorithm and LSM algorithm is obtained as shown in (14).

Theorem 6. *If $\varepsilon_w = \varepsilon_c = \varepsilon_d = 1$, $f_w = f_d = 1$, and $\xi = 1$, the ratio of network lifetime using CDLR algorithm and ERCD algorithm is*

$$\frac{\max\left(\left(1+f_c+h^2\right), \max\left(\left(h^2-(i-1)^2/2i-1\right)+hf_c+\left(2\pi i h^2/k(2i-1)\right)\right)\right)}{\max\left(\left(1+f_c+h^2\right), \max\left(\left(\left(\pi i h^2 f_c/3k\right)+\left(h^2+i^2/2i-1\right)+1\right), \max\left(\left(\pi i h^2 f_c/3k\right)+\left(2h^2-i^2/2i-1\right)+3\right)\right)\right)}. \quad (19)$$

Proof. The network lifetime using CDLR is proved in Theorem 5. For ERCD protocol, the communication load for nodes is also different based on the locations. For nodes in the hotspot area, the communication load of nodes is the same as CDLR, that is, when $i < h-k+1$, the maximum communication load is the node in ring 1, and the load is $1+f_c+h^2$. For nodes in the nonhotspot area, that is, when $i \geq h-k+1$, the communication load of each node for witness selection, legitimacy verification, and data collection in ring i is $2\pi i h^2/(k(2i-1))$, hf_c, and $(h^2-(i-1)^2)/(2i-1)$, respectively [10]. Thus the network lifetime using ERCD is expressed as

$$\mathrm{LT}_{\mathrm{ERCD}} = \frac{T_z}{\max\left(\left(1 + f_c + h^2\right), \max\left(\left(h^2 - (i-1)^2/2i - 1\right) + hf_c + \left(2\pi i h^2/k(2i-1)\right)\right)\right)}. \quad (20)$$

Therefore, based on (18) and (20), the ratio of network lifetime using CDLR algorithm and ERCD algorithm is obtained as shown in (19).

The comparison of network lifetime using CDLR, LSM, and ERCD and their ratio under the same conditions with different parameters is shown in Figures 4–6 according to (18), (15), (20), (14), and (19), respectively.

Figures 4–6 manifest that the network lifetime using CDLR obviously outperforms other methods including LSM and ERCD under the same conditions. Figures 4 and 5 demonstrate that the network lifetime using CDLR is about average 2.4 times and 1.4 times that of using LSM and ERCD, respectively. The main reasons are as follows: first of all, the CDLR implemented in the nonhotspot area makes full use of the energy of nodes in the nonhotspot area and avoids introducing new bottleneck by clone detection. Secondly, the communication load of nodes using CDLR is lower than other methods due to the random witness chains in the centrifugal direction and clone detection routes in the circumferential direction. Furthermore, the network lifetime is not related to the density of nodes in the network, which can be obtained by Figure 6(a). The network lifetime using CDLR and ERCD is constant with the change of node degree, whereas that using LSM is sensitive to the density of nodes. Therefore, the theoretical analysis declares that the performance of CDLR is better than other approaches LSM and ERCD under the same conditions.

5.3. Storage Requirement Analysis. Due to the constraints of the storage of sensor nodes, it is necessary to decrease the storage requirements during the designing of a protocol. Here, the storage requirements of sensor nodes using CDLR are evaluated.

Theorem 7. *The storage requirements of each node using CDLR algorithm are O(h).*

Proof. In CDLR, the number of witnesses of each source node in its witness chain is k, and the witness selection message from the source node is stored by these k witnesses. There are $n = \pi h^2 r^2 \rho$ nodes in the whole network, and all the witness selection messages generated by all nodes should be stored by nodes in the nonhotspot area, the number of which is $\pi h^2 r^2 \rho - \pi(h-k)^2 r^2 \rho$. Therefore, the storage requirements of each node implementing the clone detection are calculated as

$$\frac{\pi h^2 r^2 \rho k}{\pi h^2 r^2 \rho - \pi(h-k)^2 r^2 \rho} = \frac{h^2 k}{h^2 - (h-k)^2} = \frac{h^2}{2h-k} = O(h). \quad (21)$$

From (21), we can see that the storage requirements of each sensor node are not related to the number of nodes in the network, it is only related to the radius in hops of the network. Therefore, the storage requirements are not related to the density of node in network, which is similar to ERCD, but different from many previous works.

6. Experiments and Simulations

In this section, the performance of CDLR is evaluated by experiments, and the performance comparison of different methods including LSM [8], ERCD [10], and LSCD [11] is also conducted on OMNET++ [21].

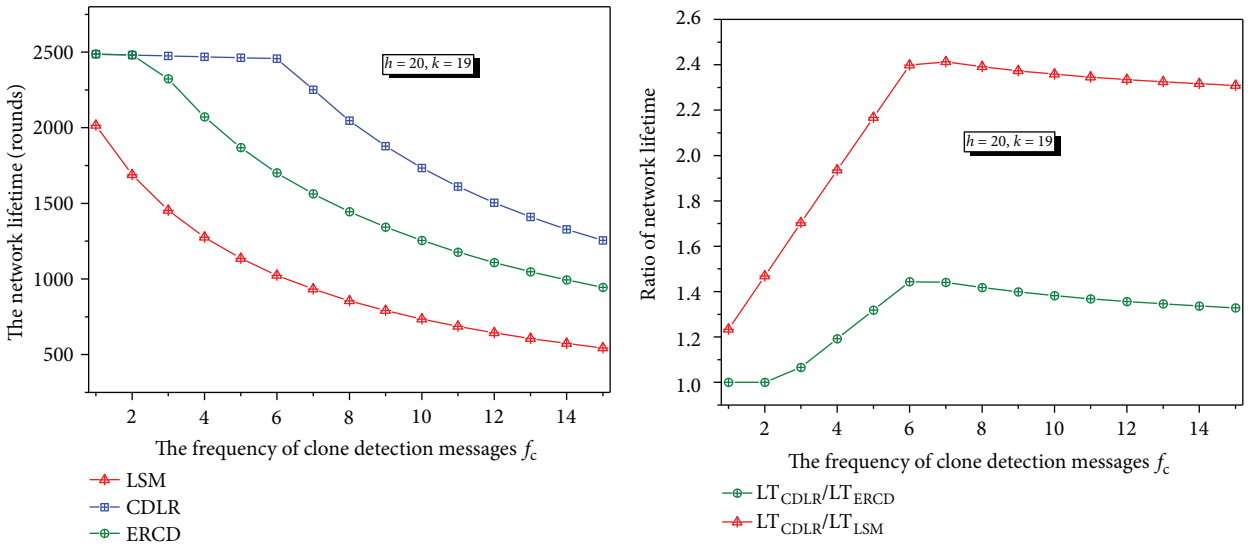
6.1. Experiments Description. In experiments, the network consisting of 2000 nodes is deployed in a circular shape whose radius is 600 m, and the BS is at the position near the center. The communication range of each node is 40 m, and ring 1 is set as the hotspot area in our experiments. In the simulation, the frequency of witness selection and observing data collection is set the same, that is, $f_w = f_d = 1$, and the frequency of clone detection is set as $f_c = 10$. To simplify the experiments, the different messages transmitted among nodes are in the same size, that is, $\varepsilon_w = \varepsilon_c = \varepsilon_d = 100$ bytes. The widely accepted energy consumption model in WSN detailed in [2] is adopted in our experiments. The simulation parameters and the range of values used in the experiments are listed in Table 2.

6.2. Experiment Results and Analysis. The experiment results in terms of clone detection probability and the network lifetime are displayed and analyzed in this section.

The detection probability of CDLR is equal to 1 theoretically under the condition that the witnesses selected are normal and trusted. However, a compromised node or a cloned node may be selected in practice situation, the behavior of which cannot be trusted. Therefore, the witness nodes controlled by adversary will cause the failure of the clone detection. Suppose that there are 10% cloned nodes in the network, the actual detection probability of CDLR is shown in Figure 7, from which we can see that the detection probability decreases with the increment in malicious nodes, but it is still in a high detection rate reaching more than 98%. Because in the set of random witnesses, the probability of a malicious node being selected as witness is very low, which is also confirmed by experiment results.

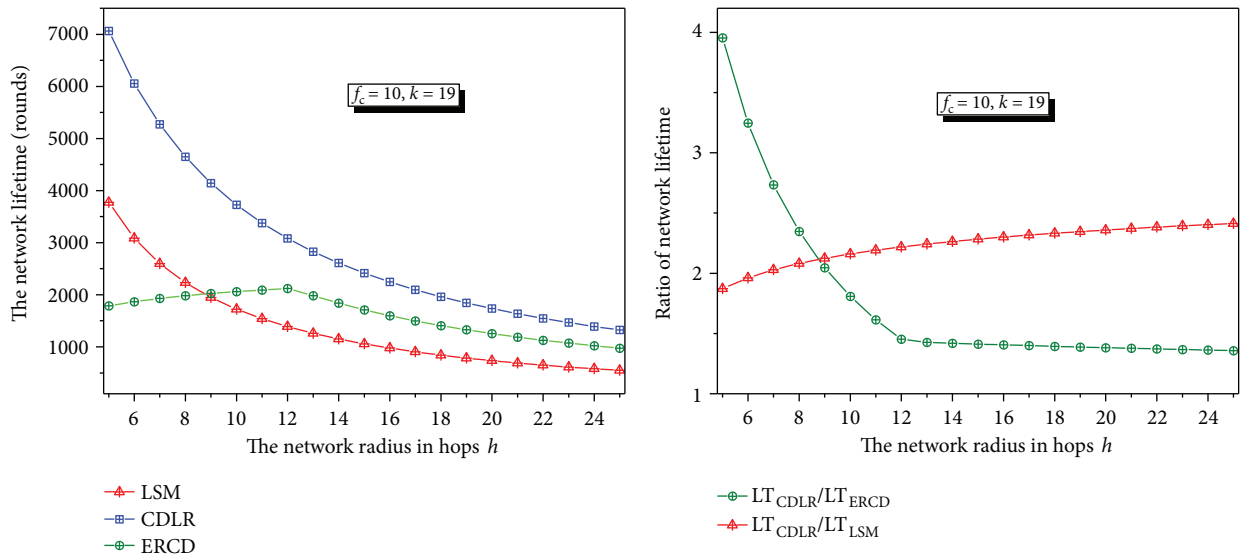
The clone detection probability of four methods including LSM, ERCD, LSCD, and CDLR is compared under the same conditions with different sensor node density (the node degree or the number of neighbors) and network scale (the radius of network). Figures 8 and 9 display the experiment results, from which we can draw a conclusion that the clone detection probability using CDLR is higher than that using other three methods under the same conditions.

Figure 8 demonstrates that the clone detection probability using CDLR, LSCD, and ERCD increases from 89% to



(a) The network lifetime (b) Ratio of network lifetime

FIGURE 4: The comparison of network lifetime of three methods with different f_c.

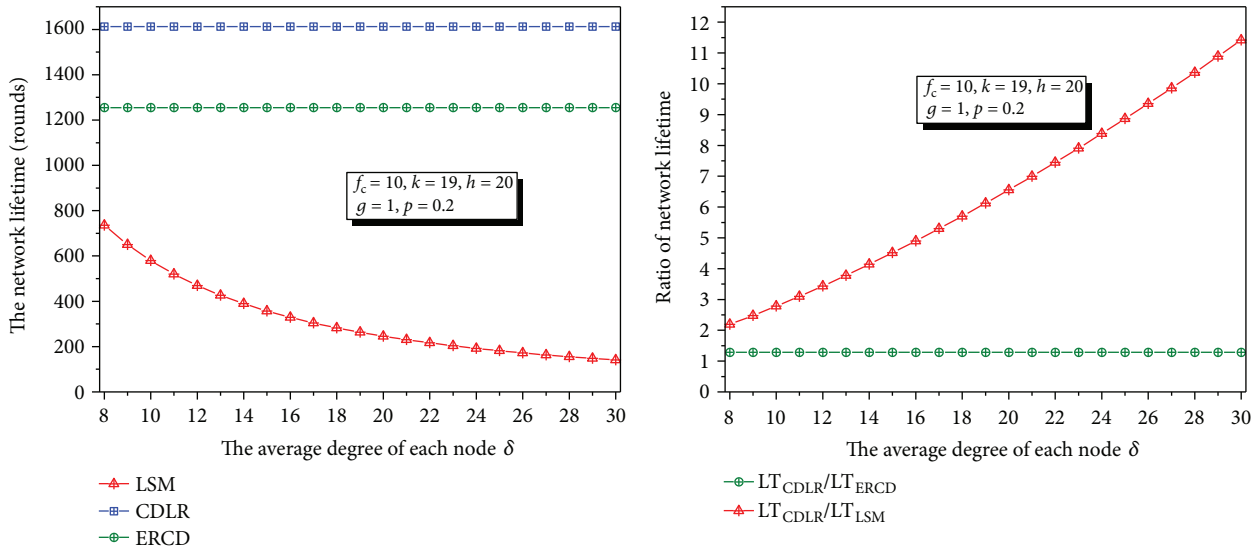


(a) The network lifetime (b) Ratio of network lifetime

FIGURE 5: The comparison of network lifetime of three methods with different h.

98%, from 86% to 96%, and from 85% to 95% with the increment in node degree, respectively. Whereas the detection probability using LSM decreases from 70% to 60% with the increment in node degree. The reason for better performance of CDLR is that the probability of success in witness selection and clone detection message broadcasting is higher when node density is larger. Moreover, the encounter probability of the witnesses and detection routes of nodes with the same ID under CDLR is higher than that under ERCD and LSCD because of the mechanism of witness selection. As for LSM, the reason for detection probability decreasing is that the LSM uses at least two-path intersection at the same node to detect clone attacks. If the node density is large, there will be more options for the next hop and the probability of two-path intersection at the same node will decrease.

Figure 9 manifests that the detection probability of four methods with the increment in network scale, that is, the network radius in hops, from which we can see that the detection probability using CDLR, LSCD, and ERCD is around 97%, 96%, and 95%, respectively, and there is not much fluctuation. It indicates that the detection probability of CDLR and LSCD is not depending on the network scale. The detection probability under ERCD has declined a little with the increase in the network radius, because both the witnesses and verification paths in ERCD are in ring structure, the probability of selecting the same ring or neighbor rings is lower with the increase in the number of rings. Overall, the



(a) The network lifetime

(b) Ratio of network lifetime

Figure 6: The comparison of network lifetime of three methods with different δ.

Table 2: The simulation parameters and the range of values in experiments.

Items	Values or ranges
The radius of WSNs deployed (m)	600
The number of members in WSNs	2000
Transmission range of nodes (m)	40
The frequency of witness selection	1
The frequency of observing data collection	1
The frequency of clone detection	10
The size of the messages transmitted in WSNs (bytes)	100
The number of random hops in witness selection	1
The proportion of replica nodes	1%–10%
Initial energy of a node (J)	0.5

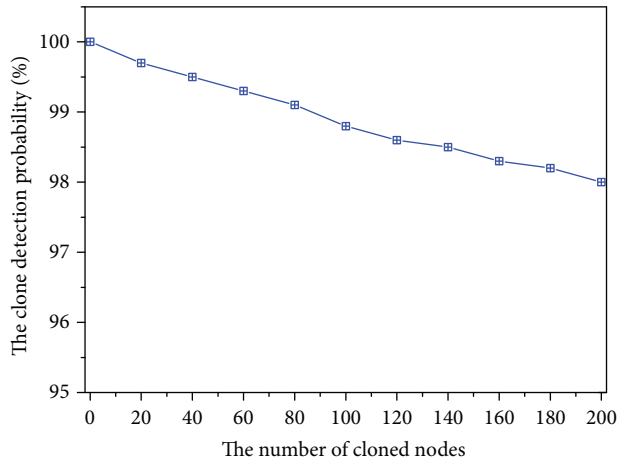


Figure 7: The clone detection probability with the number of cloned nodes.

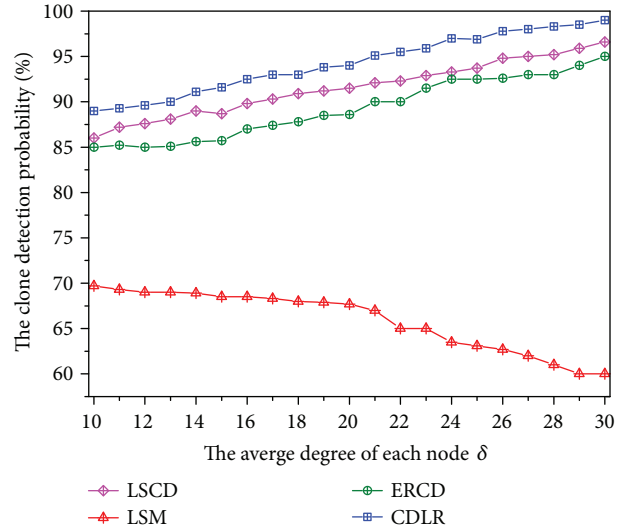


Figure 8: Comparison of detection probability of four methods with different node density.

performance of CDLR in detection probability is better than that of ERCD, LSCD, and LSM.

The network lifetime is compared under four different protocols, and the results are displayed in Figures 10–12. As the theoretical analysis results, the network lifetime using different methods is affected by clone detection frequency, network scale (network radius in hops), and node density. The experiments are conducted in these three aspects to compare the performance of four methods.

During the comparison, the differences between theoretical and experimental data are also considered. The differences are mainly from the evaluation of energy consumption. In theoretical analysis, we only consider the energy consumption caused by message sending for simplicity, because message sending consumes the most energy of nodes. In fact,

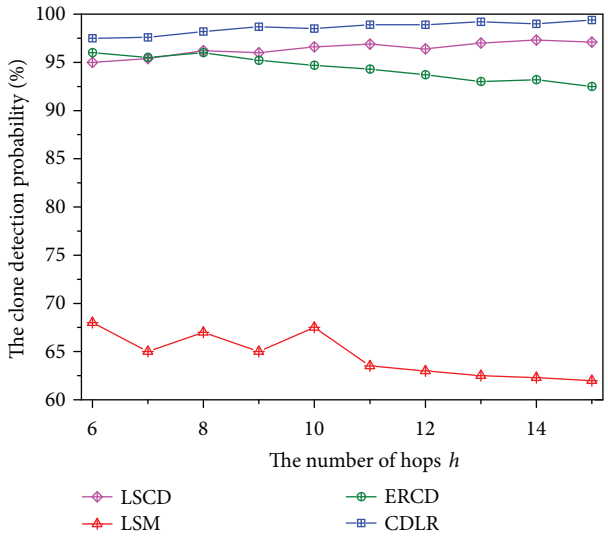


Figure 9: Comparison of detection probability of four methods with different network radius in hops.

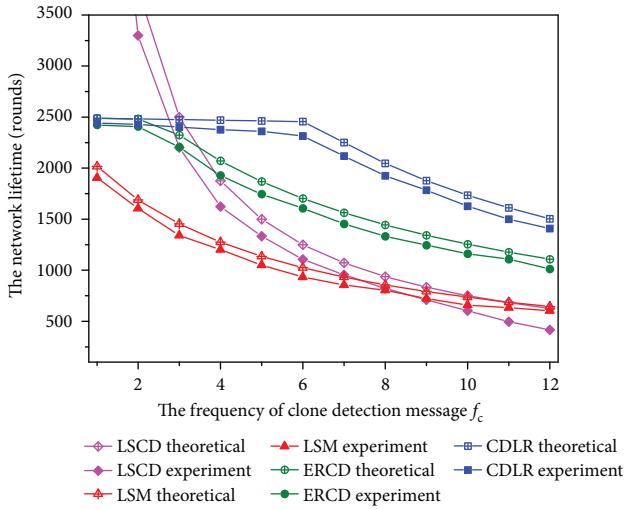


Figure 10: Comparison of network lifetime using four methods with different frequency of detection.

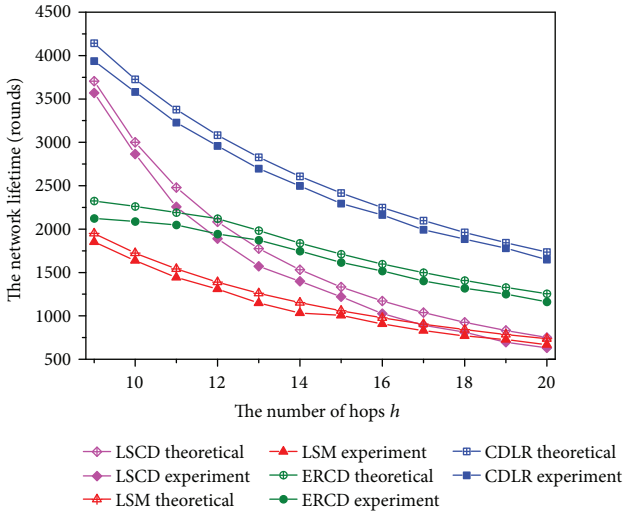


Figure 11: Comparison of network lifetime using four methods with different network radius in hops.

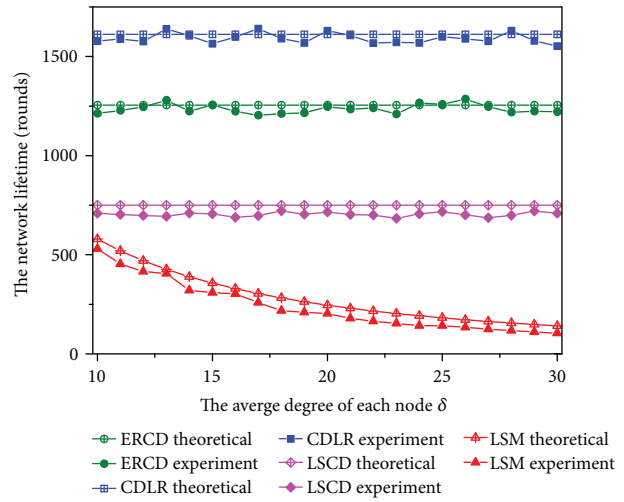


Figure 12: Comparison of network lifetime using four methods with different node density.

the energy of nodes is depleted by many factors, such as message receiving, sending, and data computing.

Figure 10 depicts the network lifetime with the increase in the frequency of detection messages using four methods, from which we can see that the network lifetime using all these methods decreases with the increase in the frequency of clone detection. However, the network lifetime using CDLR begins decreasing significantly when the frequency of clone detection is more than 6, because most of the energy consumption occurs in ring 1 mainly caused by observing data collection when the frequency of clone detection messages is less than or equal to 6. That means, in this situation, the clone detection does not affect the network lifetime. The same phenomenon occurs for ERCD when the frequency of clone detection messages is less than or equal to 2. However, the network lifetime using LSM and LSCD decreases sharply with the increase in the frequency of clone detection messages, especially for LSCD, because each clone detection message of all the nodes should start from the nodes in the second ring, whose lifetime is the bottleneck of the whole network. Thus, the network lifetime using CDLR is an average of 1.4 times, 1.7 times, and 2 times of that using ERCD, LSCD, and LSM, respectively.

Figure 11 demonstrates the comparison of network lifetime under four methods with different network scales, that is, different network radius in hops. We can get a conclusion that the network lifetime decreases with the increase in the network scale, because both the number of witnesses and the length of clone detection routes are increasing as the expansion of the network scale; thus, the communication load is also increasing. However, the network lifetime using CDLR is longer than that using other three methods, because under the same conditions, the communication load of network using CDLR is the minimal in these methods. In general, the network lifetime using CDLR is improved by

an average of 50%, 75%, and 120% compared to that using ERCD, LSCD, and LSM, respectively.

The comparison of network lifetime under four methods with different node densities is shown in Figure 12, from which we can see that the network lifetime using CDLR, LSCD, and ERCD is not related to the node density and the lifetime fluctuates in a small range, whereas the network lifetime using LSM decreases significantly with the increase in the node density. Because there is no obvious change for both the number of witnesses and the length of clone detection routes, thus the communication load is also stable. As for LSM, the number of witnesses and routes increases with the increase in node density; hence, the communication load is also increasing. The number of witnesses and routes using CDLR is the minimal in these methods, and the communication load of CDLR is the lowest of all; therefore, the network lifetime using CDLR is the longest in these methods.

7. Conclusion

In this paper, we have proposed a distributed clone detection algorithm with low resource expenditure for randomly deployed WSNs with ring structure, which consists of two phases: random witness chain establishment and detection route generation. In the proposed method, the witness chains are in the direction of network radius or in the centrifugal direction, and the detection routes are in the circumferential direction in each ring, which could ensure the encounter of the witness chains and the detection routes of nodes with the same ID but different positions. The detection probability is equal to 1 according to the theoretical analysis under the conditions that the witness nodes are not compromised. Furthermore, the performance of proposed method in terms of network lifetime and storage requirements is better than most other existing methods, such as LSM, LSCD, and ERCD. The communication load of proposed method is low, and the detection process is implemented in the nonhotspot area, which makes full use of the resource of nodes far from the BS and avoids consuming the energy of nodes in the hotspot area. Experiments and simulations have demonstrated that the proposed method outperforms most other methods in detection probability, network lifetime, and storage requirements, at the same time, the method is suitable for large-scale or densely deployed networks.

Notations

r: The communication range of sensor nodes
h: The radius of the network in hop counts
ρ: The density of deployed sensor nodes
k: The width of the nonhotspot area in hop counts
ξ: The number of random walk hops
ID_a: The ID of node a
l_a: The location/position of node a
X_a: The encrypted message of node a
R_a: The identification of the ring that node a locates in
W_a: The witness set of node a
ε_{d}: The size of the message for observing data
f_{d}: The frequency of observing data collection
ε_{w}: The size of the request message for witness selection
f_{w}: The frequency of witness selection
ε_{c}: The size of the request message for clone detection
f_{c}: The frequency of clone detection
C_i^{d}: The communication load of each node in ring i during data collection
C_i^{w}: The communication load of each node in ring i during witness selection
C_i^{c}: The communication load of each node in ring i during clone detection
C_i^{o}: The overall communication load of each node in ring i
T_z: The total size of messages a node can transmit in its life cycle
$\mathrm{LT_A}$: The network lifetime using algorithm A.

Conflicts of Interest

Zhihua Zhang, Shoushan Luo, Hongliang Zhu, and Yang Xin are with the Information Security Center in School of Cyberspace Security, Beijing University of Posts and Telecommunications, and National Engineering Laboratory for Disaster Backup and Recovery, Beijing 100876, China.

Acknowledgments

This research is supported in part by the National Key R&D Program of China under Grant 2017YFB0802300, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2015AA017201, in part by the National Natural Science Foundation of China under Grant U1536119, and in part by Applied Sci-Tech R&D Special Fund Program of Guangdong Province of China under Grant 2015B010131007.

References

[1] Y. Zhang, S. He, and J. Chen, "Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks," in *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*, vol. 24no. 3, pp. 273–281, New Orleans, LA, USA, June 2013.

[2] Y. Hu and A. Liu, "An efficient heuristic subtraction deployment strategy to guarantee quality of event detection for WSNs," *The Computer Journal*, vol. 58, no. 8, pp. 1747–1762, 2015.

[3] L. Jiang, A. Liu, Y. Hu, and Z. Chen, "Lifetime maximization through dynamic ring-based routing scheme for correlated data collecting in WSNs," *Computers and Electrical Engineering*, vol. 41, pp. 191–215, 2015.

[4] I. Butun, S. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," in *2015 6th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO)*, vol. 16no. 1, pp. 266–282, Istanbul, Turkey, May 2014.

[5] Y. Zeng, J. Cao, S. Zhang, S. Guo, and L. Xie, "Random-walk based approach to detect clone attacks in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 5, pp. 677–691, 2010.

[6] J. Ho, M. Wright, and S. K. Das, "Fast detection of mobile replica node attacks in wireless sensor networks using sequential hypothesis testing," *IEEE Transactions on Mobile Computing*, vol. 10, no. 6, pp. 767–782, 2011.

[7] Z. Li and G. Gong, "On the node clone detection in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1799–1811, 2013.

[8] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," in *2005 IEEE Symposium on Security and Privacy (S&P'05)*, pp. 49–63, Oakland, CA, USA, May 2005.

[9] M. Conti, R. D. Pietro, L. Mancini, and A. Mei, "Distributed detection of clone attacks in wireless sensor networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 5, pp. 685–698, 2011.

[10] Z. Zheng, A. Liu, L. Cai, Z. Chen, and X. Shen, "Energy and memory efficient clone detection in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 5, pp. 1130–1143, 2016.

[11] M. Dong, K. OTA, L. Yang, A. Liu, and M. Guo, "LSCD: a low-storage clone detection protocol for cyber-physical systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 5, pp. 712–723, 2016.

[12] H. Choi, S. Zhu, and P. TFL, "SET: detecting node clones in sensor networks," in *2007 Third International Conference on Security and Privacy in Communications Networks and the Workshops - SecureComm 2007*, pp. 341–350, Nice, France, France, September 2007.

[13] R. Brooks, P. Y. Govindaraju, M. Pirretti, N. Vijaykrishnan, and M. T. Kandemir, "On the detection of clones in sensor networks using random key predistribution," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1246–1258, 2007.

[14] W. Naruephiphat, Y. Ji, and C. Charnsripinyo, "An area-based approach for node replica detection in wireless sensor networks," in *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 745–750, Liverpool, UK, June 2012.

[15] C. Yu, C. Lu, and S. Kuo, "Compressed sensing-based clone identification in sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 3071–3084, 2016.

[16] A. KumarMishra and A. Turuk, "A zone-based node replica detection scheme for wireless sensor networks," *Wireless Personal Communications*, vol. 69, no. 2, pp. 601–621, 2012.

[17] N. Shashidhar, C. Kari, and R. Verma, "The efficacy of epidemic algorithms on detecting node replicas in wireless sensor networks," *Journal of Sensor and Actuator Networks*, vol. 4, no. 4, pp. 378–409, 2015.

[18] C. Ding, L. Yang, and M. Wu, "Localization-free detection of replica node attacks in wireless sensor networks using similarity estimation with group deployment knowledge," *Sensors*, vol. 17, no. 1, 2017.

[19] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor localization for very small devices," *IEEE Personal Communications*, vol. 7, no. 5, pp. 28–34, 2000.

[20] J. Newsome and D. Song, "GEM: graph embedding for routing and data-centric storage in sensor networks with-out geographic information," in *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, November 2003.

[21] "OMNet++ network simulation framework," http://www.omnetpp.org.