EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

An energy-efficient random verification protocol for the detection of node clone attacks in wireless sensor networks

Yuping Zhou[1*], Zhenjie Huang[1], Juan Wang[1], Rufeng Huang[1] and Dongmei Yu[2]

Abstract

It is easy for adversaries to mount node replication attacks due to the unattended nature of wireless sensor networks. In several replica node detection schemes, witness nodes fail to work before replicas are detected due to the lack of effective random verification. This paper presents a novel distributed detection protocol to counteract node replication attacks. Our scheme distributes node location information to multiple randomly selected cells and then linear-multicasts the information for verification from the localized cells. Simulation results show that the proposed protocol improves detection efficiency compared with various existing protocols and prolongs the lifetime of the overall network.

Keywords: Wireless sensor networks; Linear multicast; Node replication attacks; Random verification

1 Introduction

Wireless sensor networks are known as one of the three high-tech industries in the new century due to their great promise and potential with their various applications, such as in military affairs, industrial production, and environmental monitoring. Now, more and more security requirements continue to arise due to the wide application and the popularization of wireless sensor networks. The ease of deploying sensor networks improves their appeal. One sensor node can be easily inserted into an arbitrary location in a wireless sensor network without triggering any intervention from the administrator and interaction with the base station. In fact, the intrusion is only realized by triggering a simple neighbor discovery protocol. On the other hand, sensor nodes deployed in an unattended environment lack prior knowledge and hardware shielding, which is advantageous for an adversary who wants to capture and comprise them. Due to the simple structure of the sensor node, once the attacker captures one or more of the sensor nodes in the network, the running program can be cracked through a reverse analysis technique. Furthermore, the private information of the nodes, such as the node ID and key, is extracted to be used to establish a secure channel with other nodes. If the adversary replicates the sensor node by utilizing its credentials and injecting them into strategic locations, then the destructiveness would spread throughout the network. This attack is called node replication attack. Node replication attacks leave wireless sensor networks vulnerable to various insidious attacks, e.g., the adversary can pour false data into the network to prevent the success of the data aggregation protocol or the node replication attack can revoke legitimate nodes and disconnect the network by triggering a correct execution of the node revocation protocols [1].

In an effort to detect node replication attacks, researchers first proposed a method called centralized detection. While located in the wireless sensor networks, the node produces its location claim and forwards it to several neighbors; then, one or more neighbors transfer this claim to a trusted third party, e.g., a base station, which is responsible for detecting conflicting location claims. The adversary can then attack the trusted third party to prevent the detection of the clone nodes, which creates a single-point failure [2]. As a result, the centralized monitor scheme fails. Another problem is that an undue data communication burden is placed on the nodes surrounding the trusted third party,

* Correspondence: yp_zhou@mnnu.edu.cn
[1]College of Computer Science, Minnan Normal University, Zhangzhou 363000, China
Full list of author information is available at the end of the article

which may shorten the lifespan of the network. Hence, an effective and efficient detection mechanism is highly desirable. Thus, the distributed approach is proposed. In 2005 Parno et al. [3] presented two distributed detection systems designed to address node replication attacks. Both algorithms randomly select detecting witness nodes from the entire wireless sensor networks. One protocol, named the randomized multicast (RM) algorithm, multicasts the location claims of a node to arbitrary $\sqrt{n}$ nodes, which act as witness nodes. Another protocol is the line-selected multicast (LSM), which explores the routing topology of the network to select witness nodes for the location of a node and utilizes geometric probability to detect replicated nodes. The complication is that there exists either a low replica detection success rate or a high communication cost; therefore, a balance between efficiency and security cannot be achieved. Thus, discovering an effective method for selecting the witness nodes is a serious dilemma.

In this paper, a preliminary distributed protocol is presented for its use in detecting node replication attacks, which is called the global deterministic linear propagation verification protocol (GDL). In a GDL scheme, the location information of the node is propagated and stored along the horizontal and the vertical directions. The collision of conflicting location claims, which refers to the nodes with the same location information or with the same ID but in different locations, appears in the intersection of both the horizontal and vertical lines. The GDL scheme is not resilient to a smart node replication attack due to its deterministic verification process. In order to increase its robustness against a smart attack, we also describe an extension of the GDL scheme, called the randomized parallel multiple cells linear propagation (RMC) verification protocol. The basis of the RMC scheme is the combination of the localized multicast and the linear multicast. In the RMC scheme, witness nodes are randomly selected from several geographically limited regions in the wireless sensor networks, which is named cell. Within a line-selected cell, witness nodes along certain x-axes and y-axes detect the clone nodes. The Birthday Paradox is applied to map the location claim of a node to arbitrary $\sqrt{n}$ cells. A collision will appear with high probability if clone nodes are inserted into the network. In other words, the location claims of the clone nodes with the same ID but in different locations will be mapped into the same cell belonging to the arbitrary $\sqrt{n}$ cells. One major advantage of the RMC protocol is that random verification is used to provide a much higher level of compromise-resilience; another advantage is the ability to increase both the resilience and the security of the protocol. Compared with the protocols of Parno et al., both are built on the principle of monitoring randomization versus deterministic monitoring; however, the detection rate is much higher and the communication overhead is much lower with our scheme.

The rest of this paper is organized as follows: In Section 2, some of the previous research related to the protocols used to detect clone attacks is summarized and their performances are analyzed. In Section 3, a preliminary approach is proposed, which utilizes global deterministic verification. In Section 4, the preliminary approach is extended and the novel distributed detection protocol, which is based on localized linear multicast random verification, is presented. Analysis of the security and efficiency of the novel protocol and the simulation results are shown in Sections 5 and 6, respectively. Finally, the conclusion is drawn in Section 7.

2 Related works

In terms of the category of detective techniques used for node replication attacks, there are two types of known detecting methods, including centralized techniques [4-7] and distributed techniques [3,8-13]. In centralized techniques, the base station is considered to be the center, which is responsible for information-collecting and decision-making. During the process, every node in the network sends its location claim to the base station through its neighboring nodes. Upon receiving all of the location claims, the base station checks the node IDs and their locations. If there are nodes with the same ID, but in different locations, then the base station raises a clone node alarm [1]. It is easy for this method to fall into a single-point fault. Brooks et al. [14] proposed an algorithm that would detect the node replication attacks by utilizing a statistical model based on the occurrence number of keys used to authenticate the nodes in wireless sensor networks, but the method can only be applied successfully with certain random key pre-distribution schemes. Choi et al. [4] proposed a SET protocol in which the whole network is divided into exclusive subsets. Each of the subsets has a subset leader and members are one hop away from the subset leader. Multiple roots are randomly determined to construct multiple subtrees. Each subset leader collects information from its members and forwards it to the root of the subtree. The intersection operation is performed on each root of the subtree to detect clone nodes. Yu et al. [5] proposed a centralized technique, called compressed sensing-based clone identification, for wireless sensor networks. Znaidi et al. [8] proposed a cluster head selection-based hierarchical distributed algorithm that detects clone nodes using a Bloom filter mechanism that includes the network reactions. Conti et al. [6] proposed another centralized protocol, called the randomized, efficient, and distributed (RED) protocol. In this protocol, the base station multicasts a random number to the global hash function in order to output the location of witness nodes in each round of detection.

The general concept and the main idea of the centralized solution were described for the first time in the paper by Parno et al. [3]. According to this paper, there are several drawbacks inherent to a centralized system. First, the trusted third party (e.g., base station) plays an important role in the clone node detection. The base station is more likely to be compromised and to fall into a single-point failure. Second, the nodes surrounding the base station bear large amounts of the routing load. Adversaries may block the tunnel of the communication, and thus circumvent detection. Meanwhile, the power of those nodes is used up, so the lifespan of the network is shortened. Finally, for many networks, there is no powerful base station due to its high cost, so it is necessary to apply a distribution solution.

Parno et al. [3] first proposed two distributed methods for detecting clone nodes: the randomized multicast and the line-selected multicast. In these two methods, a random verification mechanism with higher security is adopted. Unfortunately, the random multicast algorithm also requires a higher communication cost and the line-selected multicast has a low node replication attack detection success rate. Zhu et al. [15] presented a distributed approach, called the single deterministic cell (SDC). In their method, wireless sensor networks are divided into several cells, and the location claim of each node is mapped to a cell and broadcasted within the cell. Nodes in the cell store location claims with certain probabilities and detect the conflicts. Zhu et al. revised the method and proposed the parallel multiple probabilistic cells (P-MPC) method. The difference between the SDC and the P-MPC is that the latter method maps location claims to one or more cells with different probabilities. Compared with the method of Parno et al. [3], localized multicast is more efficient in terms of its communication and memory costs; but, the level of compromise-resilience is low because the method is a variant of deterministic verification. Different from random verification, the deterministic verification means that witness nodes can be predicted during the detection cycle. Adversaries escape detection by compromising or controlling witness nodes to protect their clone nodes, which is called a smart attack. Random verification is necessary for high resilience to smart attacks.

Zhang et al. [9] proposed four memory efficient multicast protocols to detect replicated nodes, namely, memory efficient multicast with Bloom filters (B-MEM), memory efficient multicast with Bloom filters and cell forwarding (BC-MEM), memory efficient multicast with cross-forwarding (C-MEM), and memory efficient multicast with cross and cell forwarding (CC-MEM). The B-MEM is an extension of the LSM, which generates more memory cost per node and lower detection rates. The CC-MEM and C-MEM work poorly. In 2010, Zeng et al. [10] proposed two detection protocols, namely, the Random Walk (RAWL) and the Table-Assisted Random Walk (TRAWL) to detect node replication attacks. Both of these protocols are an extension of the LSM and thus possess the same drawbacks. Although they can achieve much higher detection probabilities than the LSM, both the RAWL and TRAWL require more than twice the communication overhead of the LSM. It is important that random verification schemes improve the efficiency of the algorithm, including the communication and memory overhead required.

Node replica detection techniques for mobile WSNs have been developed in recent years [16-24]. Ho et al. [16,17] proposed a mobile replica detection scheme based on the sequential probability ratio test (SPRT). Deng and Xiong [18] presented a new protocol to detect the replicas in mobile WSNs using the theory of polynomials based on the pair-wise key pre-distribution and Bloom filters. Lou et al. [22] proposed a node clone attack detection protocol, namely, the single hop detection (SHP) for mobile wireless sensor networks. Zhu et al. [23] proposed two replica detection algorithms for mobile sensor networks. The first algorithm is a token-based authentication scheme; the second algorithm is a statistics-based detection scheme for detecting replicas that cooperates with one another.

3 The protocol framework

3.1 Protocol requirements

Wireless sensor networks are vulnerable to a wide variety of physical attacks. One of those attacks is known as the node replication attack, in which one or more nodes are added into the network with a legitimate ID stolen from a normal node. Detecting such an attempt by centralized monitoring is not preferred due to several inherent drawbacks. Utilizing distributed monitoring can avoid single-point failures effectively. In order to prevent the adversary from predicting the witness nodes and causing them to fail in advance, it is necessary for the protocol to utilize a random and distributed technique when selecting nodes to act as witnesses.

The revocation mechanism is also needed. As soon as the clone node is discovered, the subverted node and its clone nodes should be illegitimized. Normal nodes in the network stop communicate with the illegal nodes. Sensor nodes distributed in the network suffer from several inherent deficiencies, such as limited energy and a small amount of memory, which is in the order of a few kilobytes. The protocol must decrease the amount of communication and computation required to obtain the low communication and memory costs needed for satisfactory results. At the same time, we evaluate the efficiency of the protocol by analyzing its success rate in detecting node replication attacks.

3.2 The system and network model

Wireless sensor networks are composed of hundreds or thousands of small low-cost sensor nodes. These sensor nodes are uniformly spread across a wide area and function in an unsupervised fashion. During the life cycle of the wireless sensor network, new nodes are added into the network and other nodes die due to power loss or accidental damage and disappear. The base station in the wireless sensor network is assumed to be safe and trusted. In our protocol, each node knows its own location via GPS and the sensor network is considered a geographic grid, in which each unit is called a cell.

In our scheme, an identity-based public key system is applied, in which the private key is generated by signing its public key with a master secret held only by the trusted authority (TA). So, it is impossible for an adversary to create a new identity for intruding nodes [25]. In fact, some researchers have explored all kinds of techniques to prevent adversaries from deploying nodes with arbitrary IDs in the network. For example, Chan et al. [26] presented a key pre-distribution scheme, in which the ID of each node could correspond to the set of secret keys shared with its neighbors. In this case, the adversary cannot create a new ID without the appropriate keys. The only way for the attacker to compromise a legitimate node is to get a new ID.

The energy consumption of the sensor node includes three main parts: data sensing, data processing, and data transmission and reception, amongst which the energy consumed by communication is the most critical. We adopt the first-order radio model in the transmission and receiving modes. To transmit a l-bit message a distance d using the radio model, the radio expends

$$E\text{send} = \begin{cases} lE_{\text{elec}} + l\varepsilon_{\text{fs}}d^2 & d \prec d_0 \\ lE_{\text{elec}} + l\varepsilon_{\text{mp}}d^2 & d \succ d_0 \end{cases} \tag{1}$$

where E_{elec} represents the radio dissipates to run the transmitter or receiver circuitry; ε_{fs} represents the power amplification loss in the free space model, and ε_{mp} represents the power amplification loss in the multipath fading channel model. As shown in Equation 1, if the transmission distance is less than the threshold value d_0, then the power amplification loss in the free space model is used. If the transmission distance is greater than or equal to the threshold value d_0, then the power amplification loss in the multipath fading channel model is adopted. To receive this l-bit message, the radio expends

$$E_{\text{receive}} = lE_{\text{elec}} \tag{2}$$

3.3 Adversary model

In the system, a simple and powerful attacker can capture a limited number of legitimate nodes and subvert these nodes to get their private information, such as a pair of keys, credentials, and cryptograph information. With this secret information, a clone node can communicate with any of the nodes in the network. It is easy to insert a clone node into the network because the original design for the sensor networks was to facilitate *ad hoc* deployment. Once replicas are added into the networks, all kinds of attacks arise, including eavesdropping and modifying or replaying a message. We assume that there are only a small percentage of subverted sensor nodes because if most of the legitimate sensor nodes are compromised, then any protocol to detect the replicas will no longer be in force within the network. We also suppose that at least one neighbor of the replica is legitimate.

It is assumed that the adversary can remember the nodes that have been subverted and do not repeat their attempts to capture the same nodes. In order to avoid triggering an automated protocol to sweep the network and remove compromised nodes, we assume that the adversary operates in a stealthy manner.

4 The random multicell linear multicast approach for detecting node replication attacks

The GDL and its variant, the RMC, have been designed.

4.1 Global deterministic linear propagation verification

In the GDL scheme, the format of a location claim is expressed as

$$\{ID_L, l_L, \quad \text{SIG}_{\text{SKL}}(\text{H}(\text{ID}_\text{L}||l_L))\}$$

where ID_L is the identity of node L, and l_L is the location information of node L, which can be described by either the two-dimensional coordinate (x_L, y_L) or three-dimensional coordinate $(x_\text{L}, y_\text{L}, z_\text{L})$; ‖ denotes the concatenation operation, and $\text{SIG}_{\text{SKL}}(\text{H}(\text{ID}_\text{L}||l_L))$ denotes the encrypting hash code of the data, which holds a concatenation of the identity and the location information of node L using its private key in order to verify its identity.

When node L transmits its location claim along the horizontal x_L-axis and the vertical y_L-axis lines, the neighbors within a one-hop distance on the both axes first verify the plausibility of l_L, according to the location and the transmission range of the sensor, and then verify the validity of the signature in the location claim by applying an identity-based signature scheme. Only a signature generated using the private key corresponding to the claimed identity can be approved in the verification process, which means that the adversary cannot achieve a legitimate signature without the right private key.

The node will store the location claim and continue to forward the information to the neighbors within a one-hop distance on both the lines of the horizontal x_{L}-axis and the vertical y_{L}-axis as soon as verification is obtained. During this propagation procedure, every node on both the lines of the horizontal x_{L}-axis and the vertical y_{L}-axis becomes a witness node. Whenever any witness node receives a location claim, it judges whether there is another node with the same ID claiming a different location by comparing it with previously stored claims. If conflicting location claims appear, then the witness node would forward both location claims to the base station. The base station would then broadcast a message within the network to revoke the replicas and the subverted node. The propagation will not stop until a conflict is detected or the location claim packet reaches the border of the network.

If witness nodes are selected from the sensor nodes where the two lines intersect, then the probability of detecting replicas is relatively low. As shown in Figure 1, when the location claim of node L is forwarded from point a to point b and the location claim of replica L′ is forwarded from point c to point d, there is no sensor node deployed at the point where line ab intersects line cd to act as a witness node, so the detection of the replicas fails. In order to gain a high success rate, the scope of the area in which witnesses are selected should be extended. Surface-intersecting verification takes the place of line-intersecting verification, which can be interpreted as a verification surface that is composed of a circle area of witness nodes on the horizontal or vertical axes and its neighbors within a one-hop distance. Every node in the surface acts as a witness node. When a verification surface on the horizontal axis intersects with another verification surface on the vertical axis, there must be at least one intersecting point on the intersection of the two surfaces, which can detect the replica. This is not the case with line-intersecting verification. The success rate for detecting node replication attacks reaches 100%, which is a great improvement.

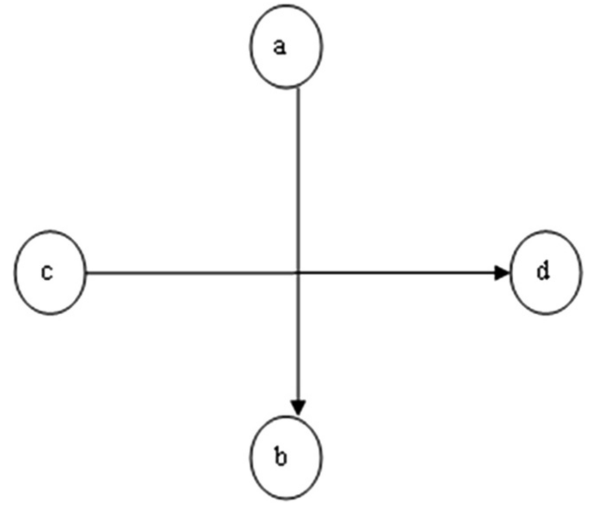

Figure 1 The invalidity of line intersection.

For example, in Figure 2, a node g with coordinate $(x_{\mathrm{g}},y_{\mathrm{g}})$ is supposed to be a replica of node a with coordinate (x_{α},y_{α}). Node a propagates its location claim along both the x_{α}-axis and the y_{α}-axis. At the same time, node g propagates its location claim along the x_{g}-axis and the y_{g}-axis. During the process of forwarding the packet, the witness nodes store received information and broadcast them to their one-hop distance neighbors, which also store information in order to detect replicas. As a result, the node m, which is selected from the intersection points where the verification surface centered on the x_{α}-axis meets the verification surface centered on the y_{g}-axis (or on y_{α}-axis and x_{g}-axis), discovers the replica. A major advantage of the GDL is that the protocol ensures a 100% success rate for the detection of node replication attacks. The communication cost and the memory cost are tightly related to the number of sensor nodes in the network. The communication cost is $O(\sqrt{n})$ and the memory cost is $O(\sqrt{n})$.

4.2 Randomized parallel multiple cells linear propagation

A potential risk comes from a smart attack against the GDL protocol. An adversary can predict the location of witness nodes and then capture and compromise them before the protocol starts to work by launching a smart attack. In GDL, when a clone node is deployed in the network, the adversary can block the forwarding of the replica's location claim on the horizontal or vertical axis. The propagation direction of the replica's location claim is deterministic because the deploying location is known. The attacker can compromise the one-hop neighbors of the clone node on the horizontal or vertical axis in order to prevent the propagation of the replica's location claim. The attacker can also subvert any node on the horizontal or vertical axis to prevent replicas from being detected.

In our RMC scheme, the wireless sensor networks are assumed to be composed of n sensor nodes and have a relative static detection cycle. The sensor node can be removed or added within the detection interval. Every node determines its geographic location information by using GPS or the positioning algorithm and acquires one two-dimensional coordinate (x,y), which creates a unique identity. The protocol includes three steps:

(1) Establishment stage for the geographic grid (cell)
In this stage, the whole network is divided into several exclusive clusters that are named cells. The LEACH routing algorithm [27] is adopted to determine the cluster headers in the first detection cycle. After the

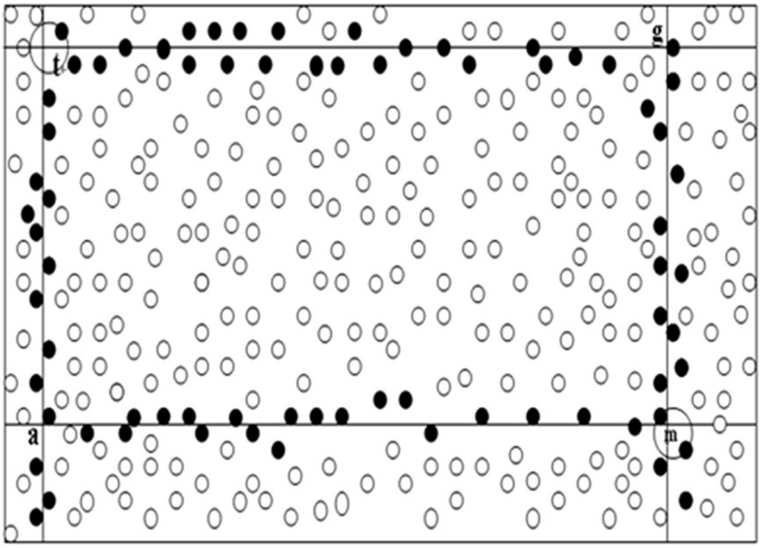

Figure 2 Protocol diagram. Empty circles denote the common sensor nodes, filled circles denote sensor nodes deployed in the forwarding path of the location information, letter a denotes one sensor node, letter g denotes the replica of node a, letter t and letter m denote sensor nodes deployed in the intersection points of the forwarding path.

first cycle, the leader in a grid cell is selected from the sensor nodes based on the principle of ‘more residual energy, more priority’ in the network. Nodes compete to be the cluster header according to the current energy/average energy ratio within a cluster, and multihop communication is used among the cluster header nodes. In fact, the cluster header node broadcasts its information as soon as it is elected as the cluster header. Normal nodes that apply to take part in a certain cluster are selected based on the principle of minimum communication cost (the communication costs are different between the node and different cluster headers). At the same time, the node records the information of the other cluster headers. The cluster header adds the node into its own routing table and the node identifies the cluster by using a geographic grid algorithm. As a result, a two-layer structure would be built into the network. Suppose the whole sensor network is divided into $m = u \times v$ cells, which indicates that there are a total of u rows and v columns in the network. A cell at the uth row and the vth column is uniquely identified as w (where $w = (u' - 1) \times v + v'$, $u' \in \{1, 2, \ldots, u\}, v' \in \{1, 2, \ldots, v\}$).

(2) Mapping stage of verification cells
At the beginning of the detection protocol, the location claim format of every node is expressed as

$$\{ID_L, l_L, \quad \mathrm{SIG_{SKL}}(\mathrm{H}(\mathrm{ID_L}||l_L))\}$$

where $\mathrm{ID_L}$ is the identity of node L, and l_L is the location information of node L, which can be described as either the two-dimensional coordinate $(x_\mathrm{L}, y_\mathrm{L})$ or the three-dimensional coordinate $(x_\mathrm{L}, y_\mathrm{L}, z\mathrm{L})$; || denotes the concatenation operation, and $\mathrm{SIG_{SKL}}(\mathrm{H}(\mathrm{ID_L}||l_\mathrm{L}))$ denotes the encrypting hash code of the data, which holds a concatenation of the identity and the location information of node L using the key of node L in order to verifying the identity of node L. When a node replication attack happens, node L randomly selects q cells in the wireless sensor networks and maps its own location information to the cells $C = \{C_1, C_2, \ldots, C_q\}$. According to the Birthday Paradox Theorem, in the sensor networks composed of m cells, every node maps its location information to $\sqrt{m}$ cells for verification, creating a very high probability of at least one collision appearing. Make q be the order of $O(\sqrt{m})$. At first, nodes receive a location claim within their respective cells; they authenticate its identity by unlocking the signature with the corresponding key and judge the rationality of the information according to its rough transmission radius. The location claim packets that cannot pass verification should be discarded. The location claim packets that pass verification will then begin the linear-selected nodes verification process within the cells.

(3) The linear-selected multicast verification stage inside the cell
Node L maps its own location claim to a certain cell. Suppose node a (x_α, y_α) is the first one that receives the location claim of node L. Node a is the first to authenticate the identity of the packet; if the authentication fails, then the packet is discarded. If

the authentication succeeds, then the time to live (TTL) in the location claim is set to $\sqrt{n/m}$. Node a stores the location claim of node L to act as a witness node; in addition, node a forwards the location claim to its neighbors within a one-hop distance, which then act as auxiliary witness nodes in the same way. In other words, the verification surface is centered in the witness node on or nearest to the x_α-axis or the y_α-axis and includes the neighbor nodes within a one-hop distance. Next, all of the nodes in the verification surface compare the location claim of node L with other stored location claims. If nodes appear that claim the same ID but have different geographic location, then the occurrence of a node replication attack would be determined; the witness node reports the ID of the clone nodes directly to the base station and the base station broadcasts a bulletin of the invalid ID within the network. Otherwise, node a propagates the location claim along both the horizontal x_α-axis and the vertical y_α-axis using a geographic routing protocol [28] and the TTL is decreased by 1. The neighbors nearest to the x_α-axis or the y_α-axis are selected to forward the packet until replicas are detected or the packet is discarded when the TTL equals to zero.

During the verification process, all the witness nodes are either on the same x_α-axis or the same y_α-axis within the cell. Thus, when the two conflicting location claim packets propagate along both the horizontal x_α-axis and the vertical y_α-axis in the same cell respectively, they must meet in the intersection of the verification surfaces and be detected with 100% probability. An example is shown in Figure 2. To reduce the memory cost, the complete information of the location claim is only stored in the witness nodes, not in the auxiliary witness nodes, because the auxiliary witness nodes no longer need to forward the location claim. A compressed location claim, including $\mathrm{ID_L}$ and l_L but not $\mathrm{SIG_{SKL}}(\mathrm{H}(\mathrm{ID_L}||l_L))$, will be stored in the auxiliary witness nodes.

5 Analysis of the RMC scheme

5.1 Security analysis

Suppose clone node L′ is deployed in l locations including $L = \{l_1, l_2, \ldots, l_l\}$. According to the Birthday Paradox Theorem, when the location claim of every node is mapped to $\sqrt{m}$ cells for verification within the wireless sensor network composed of m cells, there is a very high probability of collision. In this case, collision refers to when location claims with the same identity but in different locations are mapped to the same cell.

Every node maps its own location claim to q verification cells, which are selected randomly. According to the Birthday Paradox Theorem, the probability that q cells selected to map the location claim containing the position l_1 do not receive the q copies of the location claim containing the same identity in position l_2 is P_nc1:

$$P_\mathrm{nc1} = \left(1 - \frac{q}{m}\right)^q \tag{3}$$

In the same way, the probability that q cells selected to map the location claim containing position l_3 do not receive the $2q$ copies of the location claim containing the same identity respectively in position l_2 and position l_1 is P_nc2:

$$P_\mathrm{nc2} = \left(1 - \frac{2q}{m}\right)^q \tag{4}$$

So, the probability that location claims with the same identity but in different locations are not mapped to the same cell is P_nc:

$$P_\mathrm{nc} = \prod_{i=1}^{l-1} \left(1 - \frac{i \times q}{m}\right)^q \tag{5}$$

According to the standard approximation that $(1 + x) \le e^x$ make $x = -\frac{i \times q}{m}$ and then substitute the standard approximation into Equation 5 to obtain

$$P_\mathrm{nc} \le \prod_{i=1}^{l-1} e^{\frac{-i \times q^2}{m}} \tag{6}$$

According to $\prod_{i=1}^{l-1} e^{\frac{-i \times q^2}{m}} = e^{\frac{-1 \times q^2}{m}} \times e^{\frac{-2 \times q^2}{m}} \ldots \times e^{\frac{-(l-1) \times q^2}{m}} = e^{\frac{-q^2}{m} \times}(1 + 2 \ldots + l-1) = e^{\frac{-q^2}{m} \times \sum_{i=1}^{l-1} i}$ obtain

$$p_\mathrm{nc} \le e^{\frac{-q^2}{m} \times \sum_{i=1}^{l-1} i} \tag{7}$$

$$p_\mathrm{nc} \le e^{\frac{-q^2}{m} \times \frac{l(l-1)}{2}} \tag{8}$$

The probability of collision, in which the location claims with the same ID but in different positions are mapped to the same cell is p_c:

$$p_\mathrm{c} = 1 - p_\mathrm{nc} \tag{9}$$

Substitute Equation 8 into Equation 9 to obtain

$$p_\mathrm{c} \ge 1 - e^{\frac{-q^2}{m} \times \frac{l(l-1)}{2}} \tag{10}$$

Let $q = \sqrt{m}$. When l =1, there is only one clone node for node L, and the collision probability is 63%. When l =2, there are two clone nodes for node L, and the collision probability is over 96%, and so on. The greater the value of l, the greater the probability of collision is.

5.2 Analysis of energy consumption and efficiency

The metrics used to evaluate the energy consumption and efficiency of the RMC scheme are the following:

1. Communication cost: the average number of packets sent and received while running the replica detection algorithm in a wireless sensor network composed of n nodes, which is denoted as C_{com}
2. Memory cost: the average number of copies of the location claims stored on a sensor, which is denoted as C_{mem}
3. Percentage of the energy-exhausted nodes: the proportion of energy-exhausted nodes to all nodes

In the RMC scheme, the communication cost C_{com} is computed as follows:

$$C_{\mathrm{com}} = C_{\mathrm{f}} + C_{\mathrm{s}} \tag{11}$$

where C_{f} is the communication cost of mapping the location claim to the cell, and C_{s} is the communication cost of propagating the location claim along both the horizontal axis and the vertical axis for detection. Because nodes in the network are randomly deployed on the square unit and the average distance between any two randomly chosen nodes is approximately $0.521\sqrt{n}\approx\sqrt{n}/2$ [3], the communication cost C_{f} is in the order of $O(q \times \sqrt{n}/2)$ and the communication cost C_{s} is in the order of $O(q \times 2 \times \sqrt{{}^{n}/_{m}})$, where $\sqrt{{}^{n}/_{m}}$ is the average length of the side of a cell, and q is the number of the cells; q is in the order of $O(\sqrt{m})$, so according to Equation 11, the communication cost is $C_{\mathrm{com}} = O\left({}^{\sqrt{n}\times\sqrt{m}}/_{2}\right)$.

In terms of the memory cost C_{mem}, every location claim is stored and propagated along both the horizontal axis and the vertical axis, and during the process of verification, the node on the axis is taken as the center and works together with its one-hop distance neighbors to form a verification surface. In order to keep the collision probability over 63%, q is made to be in the order of $O(\sqrt{m})$, so the average memory cost for a node is in the order of $O\left(q \times \left(\sqrt{\frac{n}{m}} \times d + \sqrt{\frac{n}{m}}\right) \times 2\right)$, that is, $O(\sqrt{n})$, where d is the average number of neighbors for every node and $\sqrt{{}^{n}/_{m}}$ is the average length of the side of a cell. In terms of security, multiple cells are selected randomly in every detection cycle. In the cells, the sensor node receiving the newest information first forwards the location claim along both horizontal and vertical directions. The randomness reflected in the different stages of the RMC scheme is resistant to the smart attack of node replication. At the same time, the randomness can assist in avoiding a single-point failure and the phenomenon that energy consumption at the local area is so large that the nodes perish. The RMC scheme prolongs the lifespan of the network, while still achieving a high rate of detecting node replication attacks.

According to several random verification protocols, as is shown in Table 1, the average communication overhead and memory overhead per node of the RMC protocol is summarized, together with the LSM and RM protocols proposed by Parno et al. and the P-MPC protocol proposed by Zhu et al., where n is the number of sensor nodes in the network, d is the average number of neighbors of every node, g denotes the number of destinations to which a neighbor forwards the location claim, p_{f} is the probability that any neighbor of a node decides to forward the location claim from the node, w is the number of witness nodes, and m is the number of cells.

The general node-aging problem is examined by considering the percentage of the energy-exhausted nodes. In order to enhance the vitality of the network, methods for reducing the energy consumption should be considered in the design of a protocol for the detection of node replication attacks. The energy consumption of the communication module in the sensor node is the largest portion of the total energy consumption. The communication module is responsible for receiving and sending information packets, so a lower communication overhead makes the energy consumption lower. On the other hand, different verification mechanisms affect energy consumption. Random verification consumes the energy of the network evenly, which prolongs the lifespan of the network, whereas deterministic verification distributes the energy consumption unevenly. The undue data communication flaw may shorten the network's life expectancy.

6 Evaluation

A simulation experiment is performed to verify the accuracy of our scheme through OMNeT++, which is an extensible, modular, component-based C++ simulation library and framework, primarily used for building large-scale network simulators. In our simulation, n nodes are deployed uniformly at random within a 500 m × 500 m monitoring area. The communication between the different nodes follows the standard unit-disc bidirectional communication model. The number of sensor nodes varies from 1,000 to 10,000 at an increasing speed of 1,000. The communication range is adjusted to keep approximately 40 neighbors per node on average. Several physical parameters of the energy consumption model are set: the initial energy of each node E =0.5 J, the power amplification loss in the free space model $\varepsilon_{\mathrm{fs}}$ =10 pJ/bit/m^2, the power amplification loss in the multipath fading channel model $\varepsilon_{\mathrm{mp}}$ =0.0013 pJ/bit/m^4, the threshold distance d_0 = 88 m; the energy the radio dissipates to run the transmitter or receiver circuitry E_{elec} =50 nJ/bit.

Geographic routing protocol of greedy forwarding mechanism [27] is adopted to forward the information

Table 1 Comparisons of average communication overhead and memory overhead

	Communication overhead	Memory overhead
Randomized multicast (RM)	$O(n^2)$	$O(\sqrt{n})$
Line-selected multicast (LSM)	$O(g \times p_f \times d \times \sqrt{n})$	$O(g \times p_f \times d \times \sqrt{n})$
RMC	$O\left({}^{\sqrt{n}} \times \sqrt{m}\Big/_{2}\right)$	$O(\sqrt{n})$
P-MPC	$O(r \times \sqrt{n}) + O(s)$	$o(w)$

packets. We assume that there is only one compromised node and one clone node in our experiment, which are deployed randomly in the network monitoring area. The simulation experiment is run 100 times for each parameter and the average value is the final result reported.

We assume the number of cells in the network is m. Some assumptions in the formulation of the simulation that may affect the results are important. In order to reflect the fairness of the evaluation, the same assumptions will be made in this paper as are shown in the formulation used by Zhu et al. [15]. The specific configuration parameters are as follows:

$$m = k^2 \tag{12}$$

$$\begin{aligned} k &= \text{round}\left({}^{l}\Big/_{\sqrt{2}\times R}\right) \\ &= \text{round}\left({}^{l}\Big/\sqrt{2}\times\sqrt{dl^2/\pi n}\right) \\ &= \text{round}\left(\sqrt{\frac{\pi n}{2d}}\right) \end{aligned} \tag{13}$$

where l denotes the side length of the wireless sensor network, R is the communication range of a node, round ( ) is a function that rounds the input to the nearest integer; d is the average number of a node's neighbors; and n is the number of sensor nodes in the network. When there are some areas not covered in the broadcast range, the unicast mode is adopted.

Suppose l_{cell} is the side length of a cell when the location claim packet is propagated along the horizontal or vertical direction, the maximum forwarding distance does not exceed l_{cell} hop. The TTL contained in the packet is set as follows: $\text{TTL} = l_{cell} = \sqrt{{}^{n}/_{m}}$, and the collision probability of conflicting location claims in the same cell is 100%.

Every node randomly selects and maps its own location claim to q verification cells. If q is in the order of $O(\sqrt{m})$, the probability of collision when conflicting location claims are mapped to the same cell exceeds 63%. According to the computation of Equation 10, as is shown in Figure 3, when $q = 3\sqrt{m}$ holds, the probability of collision when conflicting location claims are mapped to the same cell is over 99%. When $q = 4\sqrt{m}$ holds, then the probability of collision when conflicting location claims are mapped to the same cell is 100%. Accordingly, the communication overhead and the storage overhead also increase. Simulation results show that when the size of cell s decreases to 1, then the number of mapping cells q increases to n; that is, $q = n$, s =1, then the RMC

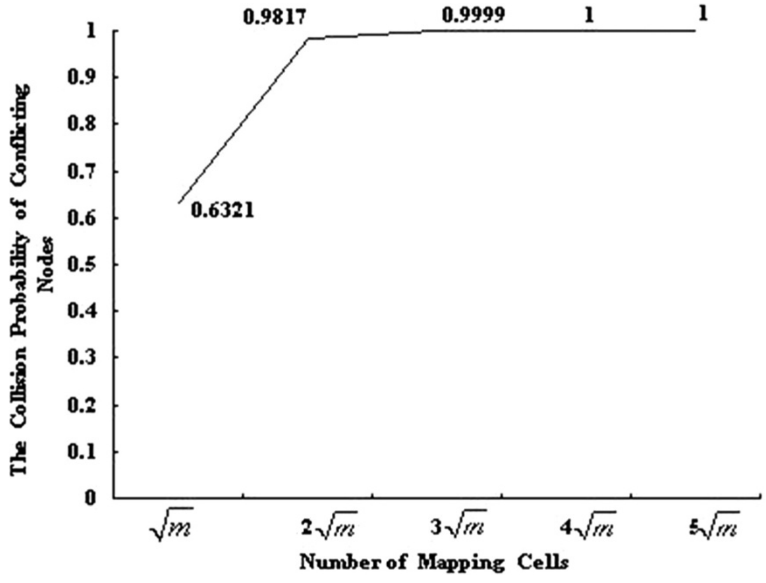

Figure 3 Collision probability of cell mapping.

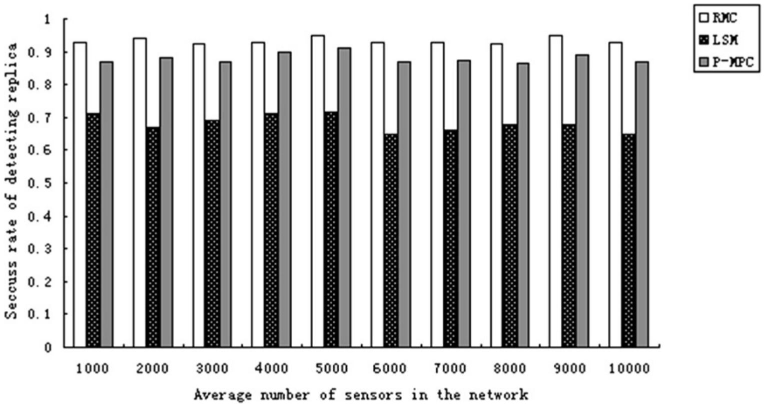

Figure 4 Success rate of detecting replicas in RMC, P-MPC, and LSM.

protocol becomes the RM protocol proposed by Parno et al., and the communication cost increases to the order of $O(n^2)$. When the size of cell s increases to n, then the number of mapping cells q decreases to 1; that is, q =1, $s = n$, then the RMC protocol becomes the GDL protocol. The difference between the RMC and the GDL is that the random verification is used in the RMC scheme and the deterministic verification is used in the GDL scheme.

The LSM and P-MPC schemes are also simulated in this experiment. The parameter settings of the two schemes are the same as the setting in papers of Parno [3] and Zhu et al. [15]. Every node has 40 neighbors in the LSM scheme; the probability that a neighbor of node L decides to forward L's location claim is $6/d$. In the P-MPC scheme, ps = 0.2, p_f =$3/d$, and v =3, which means that the probability that every node in the cell decides to store the location claim packet is 0.2 and the probability that a neighbor forwards the packet is $3/d$.

The probability of detecting node replication attacks when an adversary makes a single replica (there are two nodes with the same ID but in a different location) is the main measure to analyze the security of a wireless sensor network. Figure 4 shows the success rates of detecting replicas using the LSM protocol, the RMC protocol, and the P-MPC protocol within networks of different sizes. On average, the success rate of the RMC is 24.4% higher than that of the LSM. The main reason for this is that the LSM utilizes a line intersection detection mechanism, so the detection fails when there is no sensor node deployed on the point where the lines intersect, while the RMC utilizes a surface intersection detection mechanism, in which the node on both lines of the horizontal axis and the vertical axis is taken as the center and works with its one-hop distance neighbors to constitute the detection surface, which makes the detection rate 100%. Meanwhile, the success rate of the RMC is, on average, 5.1% higher than that of the P-MPC.

In order to evaluate the energy consumption of the RMC scheme, the communication overheads of the different schemes are compared. The communication overhead is the measure of the average number of packets sent and

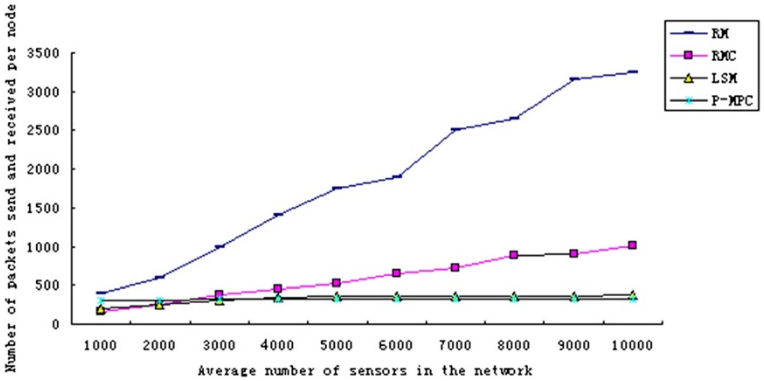

Figure 5 Communication overhead of RM, RMC, LSM, and P-MPC.

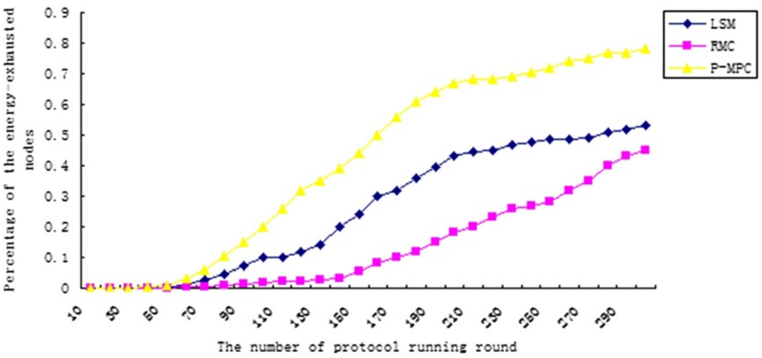

Figure 6 The percentage of energy depletion node varies with the number of protocol running rounds.

received while propagating the location claims. As is shown in Figure 5, when compared with the RM protocol that uses random verification, the communication overhead of the RMC protocol is significantly reduced. Compared with that of the other two protocols, the communication overhead of RMC is lower at the beginning. When the capacity of the network is more than 2,000 nodes, then the communication overhead of the RMC is larger than that of the LSM and the P-MPC protocols. The gap in the communication overhead is bigger as the capacity of network is larger. The main reason for this is that the RMC protocol applies the random verification mechanism. In order to maintain its high success rate, the number of mapping cells remains in the order of $O(\sqrt{m})$. The larger network size results in a larger number of mapping cells, so the communication overhead is larger. On the other hand, the P-MPC is a variant of the deterministic verification scheme. Its mapping cells are selected from deterministic cells and the number of cells is fixed. So, the number of mapping cells does not increase as the size of the network increases, thus its communication cost is relatively stable.

The percentage of the energy-exhausted nodes, which can be used to evaluate the lifetime of the overall network, is computed as the proportion of the energy-exhausted nodes to all of the nodes. Figure 6 shows that the percentage of the nodes that are exhausted of energy varies with the number of the protocol running rounds. As is shown in Figure 6, the energy consumption of the sensor nodes in the RMC protocol is apparently lower than that of the P-MPC and the LSM protocols. After running 300 rounds, 46% of the sensor nodes survive in the LSM protocol, 22% of the sensor nodes survive in the P-MPC protocol, and 54% of the sensor nodes survive in the RMC protocol. The reason for this is that the RMC protocol uses the random verification mechanism. The witness nodes are selected randomly and the energy consumption is approximately equal. At the same time, the communication overhead of the RMC is relative lower.

7 Conclusions

In this paper, two distributed detection schemes that are designed to detect node replication attacks in wireless sensor networks have been proposed. The preliminary scheme is the GDL approach. The improved version of the GDL scheme is the RMC approach. In our approach, randomly selected cells to which location claims are mapped and randomly linear-selected node verification within cells are combined to realize true randomization. This method is efficiently resilient to a smart attack of node replication. Our theoretical analysis and the empirical results show that when compared with Parno et al.'s schemes and Zhu et al.'s schemes, the success rate of detecting node replication attacks is higher in our approach. In terms of communication and memory costs, our scheme is more efficient than that of Parno et al. The simulation experiment is completed in a uniform topology environment. In our future work, a non-uniform topology environment should be used to create a simulation environment that is as close to the real application environment as possible.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by Fujian Province Education Department of China (No. JA11170, No. JA14202), the Natural Science Foundation of Fujian Province of China (No. 2012J01295) and the National Natural Science Foundation of China (No. 61170246).

Author details

[1]College of Computer Science, Minnan Normal University, Zhangzhou 363000, China. [2]College of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou 213001, China.

Received: 23 December 2013 Accepted: 11 September 2014
Published: 7 October 2014

References

1. W Khan, M Aalsalem, Detection and mitigation of node replication attacks in wireless sensor networks: a survey. Int. J. Distribute Sens. Netw. 2013, 1–22 (2013)

2. DQ Zhang, RB Zhu, SQ Men, V Raychoudhury, Query representation with global consistency on user click graph. J. Internet Technol. **14**(5), 759–769 (2013)
3. B Parno, A Perrig, V Gligpr, Distributed detection of node replication attacks sensor networks, in *Proceedings of IEEE Symposium on Security and Privacy (S & P)* (Washington D.C, 2005), pp. 49–63
4. H Choi, S Zhu, TF LA, SET Porta, Detecting node clones in sensor networks, in *Proceedings of the Third International Conference on Security and Privacy in Communication Networks* (Nice, 2007), pp. 341–350
5. CM Yu, CS Lu, SY Kuo, CSI: compressed sensing-based clone identification in sensor networks, in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops* (Lugano, 2012), pp. 290–295
6. M Conti, RD Pietro, LV Mancini, A Mei, Requirements and open issues in distributed detection of node identity replicas in WSN, in *Proceedings of the 2006 IEEE International Conference on Systems (Man and Cybernetics SMC 2006)* (Taipei Taiwan, 2006), pp. 1468–1473
7. K Xing, X Cheng, F Liu, DH Du, *Real-time detection of clone attacks in wireless sensor networks. Paper presented at the 28th international conference on distributed computing systems* (Beijing, 2008), pp. 3–10
8. W Znaidi, M Minier, S Ubeda, Hierarchical node replication attacks detection in wireless sensors networks, in *Proceedings of the 20th IEEE international symposium on personal, indoor and mobile radio communications* (PIMRC'09, Tokyo Japan, 2009), pp. 82–86
9. M Zhang, V Khanapure, S Chen, X Xiao, Memory efficient protocols for detecting node replication attacks in wireless sensor networks, in *Proceedings of the 17th IEEE International Conference on Network Protocols ICNP 2009* (Princeton, 2009), pp. 284–293
10. K Zeng, K Govindan, P Mohapatra, Non-cryptographic authentication and identification in wireless networks. IEEE Wirel. Commun. **17**(10), 56–62 (2010)
11. M Conti, R Di Pietro, LV Mancini, A Mei, A randomized, efficient, distributed protocol for the detection of node replication attacks in wireless sensor network, in *Proceedings of the 8th ACM International Symposium on Mobile (Ad Hoc Networking and Computing MobiHoc 2007)* (Montréal, 2007), pp. 80–89
12. B Zhu, VGK Addada, S Setia, S Jajodia, S Roy, Efficient distributed detection of node replication attacks in sensor networks, in *Proceedings of the 23rd Annual Computer Security Applications Conference ACSAC 2007* (Miami, 2007), pp. 257–266
13. WN Shu, L Cheng, A novel load balancing optimization algorithm based on peer-to-peer technology in streaming media. J. Convergence Inform. Technol. **7**(21), 189–196 (2012)
14. R Brooks, PY Govindaraju, M Pirretti, N Vijaykrishnan, MT Kandemir, On the detection of clones in sensor networks using random key predistribution. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(11), 1246–1258 (2007)
15. B Zhu, S Setia, S Jajodia, S Roy, L Wang, Localized multicast: efficient and distributed replica detection in large-scale sensor networks. IEEE Trans. Mob. Comput. **9**(7), 913–926 (2010)
16. JW Ho, M Wright, SK Das, Fast detection of mobile replica node attacks in wireless sensor networks using sequential hypothesis testing. IEEE Trans. Mob. Comput. **10**(6), 767–782 (2011)
17. JW Ho, M Wright, SK Das, Fast detection of replica node attacks in mobile sensor networks using sequential analysis, in *Proceedings of the IEEE INFOCOM* (Rio de Janeiro, 2009), pp. 1773–1781
18. XM Deng, Y Xiong, A new protocol for the detection of node replication attacks in mobile wireless sensor networks. J. Comput. Sci. Technol. **26**(4), 732–743 (2011)
19. CM Yu, CS Lu, SY Kuo, Mobile sensor network resilient against node replication attacks, in *Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor (Mesh and Ad Hoc Communications and Networks SECON 2008* (California, 2008), pp. 597–599
20. YJ Wang, TY Yang, YG Ma, GV Halade, JQ Zhang, ML Lindsey, YF Jin, Mathematical modeling and stability analysis of macrophage activation in left ventricular remodeling post-myocardial infarction. BMC Genomics **13**, 1–8 (2012)
21. LM Wang, Y Shi, Patrol detection for replica attacks on wireless sensor networks. Sensors **11**(3), 2496–2504 (2011)
22. Y Lou, Y Zhang, S Liu, Single hop detection of node clone attacks in mobile wireless sensor networks, in *Proceedings of the International Workshop on Information and Electronics Engineering (IWIEE)* (Harbin, 2012), pp. 2798–2803
23. WT Zhu, J Zhou, R Deng, F Bao, Detecting node replication attacks in mobile sensor networks: theory and approaches. Secur. Commun. Netw. **5**(5), 496–507 (2012)
24. RB Zhu, YY Qin, CF Lai, Adaptive packet scheduling scheme to support real-time traffic in WLAN mesh networks. KSII Transac. Internet Inform. Syst. **5**(9), 1492–1512 (2011)
25. F Hess, Efficient identity based signature schemes based on pairings, in *Proceedings of the Ninth Annual International Workshop on Selected Areas in Cryptography, SAC 2002* (Newfoundland, 2002), pp. 310–324
26. H Chan, A Perrig, D Song, Random key predistribution schemes for sensor networks, in *Proceedings of IEEE Symposium on Security and Privacy* (Berkeley, 2003), pp. 197–213
27. WR Heinzelman, A Chandrakasan, *Energy-efficient communication protocol for wireless microsensor networks. Paper presented at the 33rd Hawaii international conference on system sciences* (Hawaii, 2000), pp. 1–10
28. YJ Wang, P Chen, YF Jin, Trajectory planning for an unmanned ground vehicle group using augmented particle swarm optimization in a dynamic environment, in *Proceedings of IEEE International Conference on Systems (Man and Cybernetics)* (San Antonio, 2009), pp. 4341–4346

doi:10.1186/1687-1499-2014-163
Cite this article as: Zhou *et al.*: An energy-efficient random verification protocol for the detection of node clone attacks in wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking* 2014 **2014**:163.