Hindawi
Security and Communication Networks
Volume 2022, Article ID 2440313, 12 pages
https://doi.org/10.1155/2022/2440313

Research Article

Protecting Location Privacy in IoT Wireless Sensor Networks through Addresses Anonymity

Qiong Zhang [1,2] **and Kewang Zhang** [3]

[1]*School of Computer Science and Technology, Xi'an University of Posts and Tele-communications, Xi'an 710121, China*
[2]*Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an 710121, China*
[3]*School of Computer Science, Xi'an Jiaotong University, Xi'an 710049, China*

Correspondence should be addressed to Qiong Zhang; zhangqiong@xupt.edu.cn

Received 17 August 2021; Revised 1 December 2021; Accepted 27 March 2022; Published 18 April 2022

Academic Editor: Yuyu Yin

Location privacy is very important for event-triggered type of Wireless Sensor Networks (WSNs) applications such as tracking and monitoring of wild animals. Most of the security schemes for WSNs are designed to provide protection for content privacy. Contextual privacy such as node identity anonymity has received much less attention. The adversary can fully explore such contextual information to disclose the location of critical components such as source nodes or base station. Most existing schemes provide location privacy at network layer. As no measures are taken to provide node identity anonymity at data link layer, the adversary can launch traffic analysis attacks to jeopardize location privacy. In this paper, a scheme named HASHA is proposed to defend against traffic analysis attacks through hashed one-time addresses. Hashed results of payload are used to create dynamic one-time MAC addresses between the communication pairs. Because of inevitable wireless frame errors, it is impossible for adversaries to track dynamic addresses. Therefore, HASHA can provide strong node identity anonymity, which makes traffic analysis attacks much more difficult and provides better location privacy. Simulations and analysis results show that HASHA can provide better location privacy with limited communication overheads, which is particularly suitable for resource-limited WSNs.

1. Introduction

A typical Wireless Sensor Network (WSN) is composed of dozens to thousands of tiny, low-cost, and resource-constrained sensor nodes that are self-organized as an ad hoc network to monitor the physical world. One type of applications of WSNs is wildlife habitat monitoring, in which all sensor nodes are deployed randomly to monitor the target of interests [1]. Detection events are reported from the source node to the base station in a multihop fashion. Unattended operation and open wireless communication channel make WSNs vulnerable to attacks. However, as sensor node has limited memory, energy, and communication resources, traditional security techniques cannot be used in WSNs. Light-weighted schemes are required to achieve secure communication for WSNs [2].

Security for WSNs has focused on security services that provide authentication, confidentiality, integrity, and availability [3, 4]. Such techniques belong to content privacy. Now, however, there is a growing interest in contextual privacy, which focuses on hiding the contextual information of WSNs. Location information of key components is one of the most important contextual privacy parts that should be protected.

In the wildlife monitoring application, all sensor nodes detect occurrence of the target animal to the base station. In the case that one sensor node (source node) detects target, a packet is generated and sent to the base station hop by hop to report occurrence of the target. In such applications, geographic locations of the source node and base station are sensitive information that should be protected [5]. The base station is the only gateway to outside networks, and the

source node reveals physical location of wildlife. If the location of base station is disclosed by the adversary, the capture of the base station can make the entire network nonfunctional. And if the location of source node is disclosed, the adversary can find the animal easily because the geographic location of source node and the target must be very close. Therefore, providing location privacy of source node and base station is of great importance in such applications.

Existing techniques provide location privacy at network layer. Random Walk has packets that follow random route while forwarding the packets from the source node to base station [6, 7]. As it is difficult to the adversary to backtrack to the source node while random route is used, location privacy of source node is achieved. Dummy Data Source scheme invites some fake source nodes into the WSNs to confuse the adversary and provide location privacy [8, 9]. However, both schemes introduce additional communication overheads, which consume much more energy. For example, if the average hops count from the source node to base station is twice than that of shortest path, the energy consumption is twice too. For the same reason, if one more fake source node is added to the network, the power consumption doubled.

The adversary may launch traffic analysis attacks to find the geographic locations of the source node. As the content information is protected by the encryption techniques, the adversary cannot decrypt its contents without keys. However, as the contextual information is not well protected, it can be used to launch successful traffic analysis attacks.

The adversary first captures frames around the base station. The structure of frame at the data link layer data is <DA||SA||payload||FCS>. Payload is content from upper layer, FCS is the frame checksum, DA is receiver address, and SA is transmitter address. Supposing that the captured frame is <BS||B||payload||FCS>, the adversary cannot get any information from the payload because it is encrypted. However, two addresses indicate that the frame is from node B to the base station. To find the geographic location of node B, the adversary captures a series of data packets from node B at different locations and moves towards locations where stronger Received Signal Strength (RSS) presents. After finding the geographic location of node B, the adversary can find the next node by the same way. To find the source node, the adversary continues such process until no more next nodes are detected. Source location privacy was compromised.

Two steps are used repeatedly by the adversary to locate the source node. The first is forwarding relationship analysis. The adversary knows the address of base station by traffic analysis. Then, it knows the forwarding node closer to the source node by analyzing frames to the base station. The second step is to move closer to the forwarding node by analyzing RSS. Apparently, the addresses in data link layer frames are vital for successful traffic analysis attacks. It is much more difficult or impossible for the adversary to launch traffic analysis attacks if the addresses in the frame are well protected.

One way to hide node address is to break the relevance between physical node and the address of the node. For example, if node X and node X′ in WSN have the same address IDx, as two nodes have the same address but are deployed at different geographic location, the adversary cannot locate the node(s) by analyzing the RSS [9]. Thus, traffic analysis attacks can be eliminated. Of course, above simple scheme introduces great trouble to normal operation of networks. But breaking the relevance between physical node and the address is an effective way to defend against traffic analysis attack and provide location privacy [10–13].

Another way to break the relevance between physical node and the address is introducing more addresses to node that cannot be distinguished by the adversary. If node X communicates with base station using a serial of identities <X1, X2, X3, . . ., Xn>, and only node X and the base stations know that the addresses belongs to node X, the adversary cannot learn the communication relationship to track the source node by traffic analysis attacks [14].

Based on such observations, this paper proposes a novel scheme to provide location privacy at data link layer. The contributions of this paper are threefold: First, the proposed scheme protects location privacy at data link layer, which is more effective to defend against traffic analysis attacks. As compared to schemes at network layer, the proposed scheme introduces negligible communication overheads. Second, in tracking and monitoring applications, location privacy of the source node and that of base station are both very important. Exposure of base station will endanger the whole WSN, while source node location discloses the position of the target. Source location privacy and base station privacy are both provided in the proposed scheme. Existing schemes emphasize on either the source location privacy or base station location privacy. Third, the proposed scheme defends traffic analysis attacks through address anonymity, which can provide location privacy against both inner attackers and outside attackers. Protection against inner attackers is particularly important, because node compromise is fairly easy for unattended WSNs and the compromised node can be an inner attacker with some software modifications. Existing address anonymity schemes can only defend against outside attackers.

2. Related Works

Phantom Routing belongs to Random Walks type schemes that provide location privacy for WSNs [13–15]. To prevent being located by step-by-step tracing, the source node sends each packet to a randomly selected forward node. This forward node is called a Phantom node. On receiving the packet to be forwarded, the Phantom node routes the packet to the base station using broadcasting. Suppose that an adversary launches traffic analysis attacks to find the geographic location of the source node. As the Phantom node sends packets to base station via broadcasting instead of unicasting, it is fairly difficult for the adversary to trace to the Phantom node using traffic analysis. However, energy consumption in Phantom Routing is much greater than unicasting type of schemes, because broadcasting is used to forward packets from Phantom node to base station. To reduce power consumption of broadcasting, another scheme named Phantom Single-path Routing scheme (PSRS) is

proposed [16]. Different from original Phantom Routing, the Phantom node in PSRS routes packets to base station via unicast. As the source node selects different Phantom node for each packet, different paths are used for different packets. Therefore, it is still very difficult to locate source node via traffic analysis attacks. The PSRS can reduce power consumption because broadcasts are eliminated. But the randomly selected paths are much more power-consuming than the shortest ones.

Another type of schemes that provides location privacy is dummy source node [17, 18]. Fake source nodes are introduced to obfuscate real source node. The basic idea of such schemes is quite simple. There are many source nodes in the WSN, only one node is real source node, and other nodes are fake source nodes. The adversary can no longer see which one is real source node, even if success traffic analysis attacks are launched. Obviously, one additional fake source node introduces additional network traffic, which corresponds to additional energy consumption. The more fake source nodes introduced, the more power consumption.

Simple Anonymity Scheme (SAS) is the first scheme proposed to provide location privacy at data link layer by hiding the address. Each node communicates with neighbor using a pseudonym [19]. A large range of pseudonyms are used, and each node is assigned with a subspace of the pseudonym space. Both nodes of the communication pair at data link layer know each other's pseudonym spaces. Both nodes use different pseudonym within its pseudonym spaces. Therefore, the adversary cannot identify the physical node if the pseudonym space is unknown to it. The main drawback of SAS is that it cannot protect address anonymity if there is an internal attacker. For example, if the adversary has the full pseudonym space and the subspace allocation for each node, it can capture frame and compare each address with the pseudonym space and finally find out the physical node for each address. Another drawback of SAS is that each node must store pseudonym space for each neighbor, which introduces great storage overheads if many neighbors exist.

Cryptographic Anonymity Scheme (CAS) uses a keyed hash function to generate the pseudonym used for communication between the communication pairs at data link layer [20, 21]. Before deployment, the communication pairs are assigned a key k for pseudonyms generation. After deployment, the communication pairs create pseudonyms with a random number r and a sequence number seq. The ith pseudonym can be expressed as $\mathrm{ID}i = \mathrm{H}_k(r \oplus seq)$. Before frame transmission, a different sequence number seq is used, so each frame has different pseudonym. CAS reduces storage overhead at the expense of additional computation overheads. Apparently, CAS cannot prevent internal attackers from finding out that some pseudonyms belong to a physical node if the key k is stolen by the adversary via compromising.

The schemes mentioned above either cannot protect location privacy in the presence of inner attackers or consume too much energy resource because of communication overheads. And most of the schemes proposed focused on protecting location privacy of source node. In this paper, the proposed scheme protects location privacy at data link layer by address anonymity. The address anonymity can resist traffic analysis attackers launched by both outside attackers and inner attackers. With a modification to the network layer, the scheme can provide location privacy with much less energy consumption. Location privacy of both base station and source node can be protected with the proposed scheme.

3. Network and Adversary Models

3.1. Network Model. In this paper, it is assumed that many nodes are randomly deployed to monitor the geographic location of the target. Each node is capable of communication, computation, and sensing. All nodes in the network are powered by batteries and work in an unattended manner [21]. Therefore, power efficiency is the most important design consideration for both software and hardware. There is only one base station in the WSN, which is the gateway to outside networks.

All nodes in the WSN are working coordinately to detect the presence of a target. Any tracking approaches can be used to detect the target, provided that they are power efficient. The node that detects the target is called the source node. On detecting the target, the source node sends packets to the base station to report the information of the target. The source node reports to the base station for fixed time interval until the target moves outside the detection radius. Other nodes in the WSN sleep unless they are requested to forward the packets from the source node to the base station.

3.2. Adversary Model. Location privacy of the source node and that of base station are both important [21–23]. We consider two types of adversary. The first type of adversary is interested in catching the animals that are monitored by the WSN. Because the network traffic from the source node is an excellent guide to find the animals, the adversary attempts to find the node closer to the source node (and the animal) through traffic analysis attacks. The second type of adversary attempts to find the base station and damage it, which will make the entire WSN useless. With the same approach, the adversary can find the base station.

Only local adversary is considered in this paper. The reason is that global adversary requires much more expensive devices than the local adversary. Some researches suppose that the adversary is equipped with wireless devices that can cover the whole WSN. Such devices should be very expensive. Many nodes that are geographically separated in a WSN may transmit simultaneously without collision at the respective receivers. But as the wireless device of adversary can hear many simultaneous transmissions, collision may occur at the adversary. Expensive wireless device may not necessarily lead to better attack results. Therefore, we only consider local adversary.

Only passive adversary is considered in this paper. That means the adversary never transmits to avoid being detected by WSNs. To locate the source node and base station, the adversary captures and analyzes frames to get the communication relationship among nodes. To move closer to the source node or base station, the adversary may move closer to a node by comparing RSS from different locations.

The adversary may launch another traffic analysis attack named time correlation [24–27]. After detecting the target, the source node sends a packet to the next node closer to the base station to notify the event. The next node also relays the packet to a node closer to the base station. The adversary can observe the correlation in transmitting time between one node and the next node to find the route to the source node or base station. For a simple example, if the adversary notices that after node A transmits a packet, node B transmits a packet with the same size, it can learn that node A is closer to the source node, and node B is closer to the base station. The reason is that, in a typical tracking and monitoring application, only the source node generates packets and the base station is the only destination.

As nodes in a WSN are frequently deployed in unattended environment, node may be captured by the adversary. The adversary can analyze the software and hardware of the node. It is possible for the adversary to get the pairwise shared keys or other sensitive information [28, 29]. Even more, modification to the software is also possible if the adversary has enough skills [30–34]. The captured node then becomes an internal attacker. Protecting attacks launched by an internal attacker is much more difficult than that launched by outside attackers [35–37].

4. Proposed Location Privacy Scheme

4.1. Address Anonymity Scheme. A node may have different identity at different layers of the network protocol stack. Identity at network layer and upper layers can be protected by cryptographic system. However, identity at data link layer has not been well protected in popular wireless standards such as 802.15.4 and Lora [31, 32]. Without introducing confusion, identity and Media Access Control (MAC) address are used interchangeably in this paper. The frame structure at data link layer can be illustrated in Figure 1.

DA is destination address of the frame. SA is the address of the sender. Payload is data from upper layer. Upper layer of data link layer is network layer. Therefore, payload at data link layer is usually packet at network layer plus control information. FCS is frame checksum.

As wireless channel is error prone, Automatic Repeat request (ARQ) is used to provide reliable data transmission. On receiving DATA frame, the receiver responses an ACK frame to inform the sender that it has received the DATA frame successfully. Structure of ACK frame is illustrated in Figure 2.

As compared to DATA frame, the ACK frame is much shorter. But both destination address and source address are included in the ACK frame. Destination address is the address of DATA frame sender, and source address is the address of receiver.

As elaborated in the adversary model, SA and DA of each node are known to the adversary who captures frames through eavesdropping. By analyzing these addresses, the adversary knows how many nodes in the WSN and the MAC address of each node. Furthermore, based on such captured frames, the adversary can deduce the routing information of the network or even locate a certain node in the network.

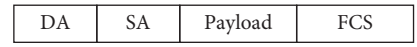

Figure 1: DATA frame at the data link layer is composed of destination address, source address, payload, and frame checksum.

Figure 2: ACK frame at the data link layer is used for reliable communication.

Therefore, unprotected addresses at data link layer are the root factor jeopardizing location privacy.

To protect the addresses in the DATA frame and ACK frame, a hash function Hash() and a keyed hash function HMAC() are used. For DATA frames from node a to node b, both nodes keep the following variables: Key[a- > b], IDS[a- > b], IDD[a- > b], where Key[a- > b] is a secret key to protect the addresses in MAC frames. IDS[a- > b] is the source address assigned to DATA frame and IDD[a- > b] is the destination address assigned to DATA frame.

Nodes in the WSN know each other by beacon broadcasting. For example, node a knows IDb after receiving beacons from node b. Node a and node b initialize these variables as follows:

(i) Key[a- > b] ←0xFFFFFFFF

(ii) IDS[a- > b] ←HMAC(Key[a- > b], IDa)

(iii) IDD[a- > b] ←HMAC(Key[a- > b], IDb)

For the first DATA frame from node a to node b, two hashed addresses IDS[a- > b] and IDD[a- > b] are used. After ACK frame from node b to node a, both nodes update key and addresses:

(i) Key[a- > b] ←Key[a- > b] ⊕ Hash(payload)

(ii) IDS[a- > b] ←HMAC(Key[a- > b], IDa)

(iii) IDD[a- > b] ←HMAC(Key[a- > b], IDb)

As payload of the first frame is received successfully by node b, it has the same Key[a- > b], IDS[a- > b], and IDD[a- > b] as node a. We call this a secret key update process.

As it is well known, wireless channel is error prone. Both DATA frame and ACK frame may be corrupted. On receiving corrupted DATA frame, node b will not acknowledge node a with ACK frame. Both nodes will not update the key. Node a retransmits the DATA frame using the old key as described above.

In another scenario, DATA frame is received correctly by node b, but the ACK frame to node a is corrupted or lost. Node a retransmits the DATA frame as it does not receive the ACK frame correctly. But node b has already updated the key. Key mismatch problem occurs.

To address key mismatch problem, two temporary addresses are used by node b to avoid key mismatching. Node b keeps a copy of old address on receiving DATA frame successfully. If the received address in next DATA frame does not match the **new** address, it will try to match the **OLD** temporary address. If the old one matches, that means this DATA frame is a retransmission. Just reply node a with the ACK frame that already transmitted.

Node *a* and node *b* repeat such process for all the frames from node *a* to node *b*. Such process creates one-time source address and destination address. We call it a dynamic address or hashed address (HASHA) (Algorithm 1).

HASHA updates key for the communication pairs after a successful data transmission. And the one-time secret is further used to update the addresses, which creates dynamic addresses. Such process can create great difficulty to the adversary.

Figure 3 illustrates a typical scenario that an adversary captures frames from node A to node B. Initially, as the adversary knows MAC addresses node A and node B through capturing beacons. The adversary knows the initial value of Key[a- > b]. Both node B and the adversary receive DATA1 and DATA2 successfully.

At time *t1*, ACK3 is corrupted and node B does not receive it correctly; node B receives the retransmission with backup key. This will not introduce trouble to the adversary.

At time *t3*, DATA4 is not received correctly by the adversary; as the adversary is passive attacker, it cannot ask node A for retransmission. Thereafter, the adversary cannot trace frame from node A to node B after time *t3*. The reason is that the addresses used by node A and node B are created by HMAC function with key5. Key5 is created by all previous payloads from node A to node B. The result is that the dynamic one-time addresses of the following frames from node A to node B are indistinguishable to the adversary. Address anonymity is achieved.

As wireless frames are error prone because of collision and interference, corrupted frames at the adversary will prevent it from identifying nodes in the WSN. Therefore, with HASHA, the eavesdropping adversary cannot identify number of nodes in the WSN. Therefore, it cannot retrieve the routing information. Without forward routing information, the adversary cannot trace the source node and base station.

4.2. Possible Attacks against HASHA and Countermeasures. Even though the addresses are hidden by address anonymity, the adversary can still launch two types of attacks to jeopardize the source node and base station location privacy. The first attack is time correlation attack. The adversary can deploy several attack nodes in the target WSN. These nodes are carefully deployed so as all communications in the WSN can be captured. The geographic coordinates of these nodes are recorded in a center control point. The attack nodes can communicate with each other to report captured frames to the center control point. Time synchronization algorithm can be used to distribute global time to these nodes. Therefore, the resulting attack network can be used to detect transmission all over the network.

In a typical event-triggered monitoring type of WSN, network traffic in the networks is triggered by event detected by the source node. The source node reports event to the base station with the help of the forwarding nodes. On forwarding the event to the base station, transmission time of the forwarding nodes may disclose the location of source node and the base station.

As illustrated in Figure 4, nodes a1, a2, and a3 are nodes of the attack network to monitor network traffic. Node S is source node and node D is base station. Node A and node B are relay nodes. To report event from node S to base station D, node S sends packet to node A, and node A sends packet to base station D with the help of node B. The transmission time is illustrated in Figure 5. By analyzing the transmissions time serial, the adversary can find that node S is the source node and node D is the base station, which are located near node a1 and node a3, respectively. The location privacy of source node and that of base station are jeopardized. Of course, with the help of address anonymity, the adversary cannot identify node S, node A, and node B. But it can still detect that the source node is close to a1 and the base station is close to node a3. As locations of node a1 and node a3 are known to the adversary, address anonymity cannot eliminate such time correlation attacks.

Time correlation attacks use the pattern of occurrence of transmissions along the forwarding path to find the source node and base station. For example, for each event, transmission of node S is always followed by transmission of node A, because node A is the next hop of the forwarding path. Transmission serial {S1, A1, B1}, {S2, A2, B2}, and {S3, A3, B3} disclose forwarding relationship among nodes, which can be used to jeopardize source node and base station location privacy. Breaking the transmission pattern is important to eliminate time correlation attacks. The solution is to introduce random delay while forwarding packet. As illustrated in Figure 6, node S and node A delay random time for packets. The resulting transmissions serial {S1, A1, S2, A2, S3, B1, B2, A3, B3} does not disclose any forwarding relationship anymore. With the help of address anonymity, it is more difficult for the adversary to locate the source node and base station via time correlation attacks.

The formal description of random delay can be expressed as follows.

Each node forwards packets with random delay, which is effective to prevent time correlation attacks. Of course, random delays may introduce delay to event reporting to base station. In some applications, timely delivery of important packet to base station is very important. To provide higher priority to such important data, a smaller random delay *rand* in Algorithm 2 can be selected.

Another traffic analysis attack is traffic outlining attack. As address anonymity and random delay are used to prevent traffic analysis attack and time correlation attack, respectively, it is much difficult for adversary to launch attacks based on node address and forwarding relationship. But the adversary can still attack the target network via traffic outlining attack. As mentioned above, the adversary can deploy many attack nodes in the network to launch a distributed attack. For example, in the network illustrated in Figure 7, source node S reports to base station B. The adversary can deploy many attack nodes to monitor network traffic. As network traffic of event-triggered WSN is characterized from source node to base station, it is impossible for the adversary to outline the traffic without the help of distributed attack nodes. All attack nodes report to the adversary only in the presence of traffic in a certain time period. As the geographic location of attack nodes is known to the adversary, the adversary knows geographic

```
if node is sender then
  Key[a- > b] ← 0xFFFFFFFF;
  if node has frame to be transmitted then
    IDS[a- > b] ← HMAC(Key[a- > b], IDa);
    IDD[a- > b] ← HMAC(Key[a- > b], IDb);
    Send DATA frame { IDS[a- > b], IDD[a- > b],payload, FCS};
    Create timer Stimer with a certain timeout;
  end
  if an ACK frame is received then
    for each neighbors do
      if FCS checksum OK and IDS[a- > b] from ACK frame = = IDS[a- > b] then
        //generate new key
        Key[a- > b] ←Key[a- > b] ⊕ Hash(payload);
        //generate new addresses
        IDS[a- > b] ←HMAC(Key[a- > b], IDa);
        IDD[a- > b] ←HMAC(Key[a- > b], IDb);
        kill timer Stimer;
      end
    end
  end
  if Stimer timeout then
    Send DATA frame { IDS[a- > b], IDD[a- > b],payload, FCS};
  end
end
if node is receiver then
  Key[a- > b] ← 0xFFFFFFFF;
  IDS[a- > b] ← HMAC(Key[a- > b], IDa);
  IDD[a- > b] ← HMAC(Key[a- > b], IDb);
  IDSo[a- > b] ← HMAC(Key[a- > b], IDa);
  IDDo[a- > b] ← HMAC(Key[a- > b], IDb);
  if a DATA frame is received then
    for each neighbor do
      if FCS checksum OK and IDD[a- > b] from frame = = IDD[a- > b] then
        //save old addresses
        IDSo[a- > b] ← IDS [a- > b];
        IDDo[a- > b] ← IDD [a- > b];
        //generate new key
        Key[a- > b] ←Key[a- > b] ⊕ Hash(payload);
        //generate new addresses
        IDS[a- > b]<-HMAC(Key[a- > b], IDa);
        IDD[a- > b]<-HMAC(Key[a- > b], IDb);
        deliver frame to upper layer;
        send ACK frame { IDDo[a- > b], IDSo[a- > b], FCS};
      else if FCS checksum OK and IDD[a- > b] from frame = = IDDo[a- > b] then
        //DATA frame already delivered
        send ACK frame { IDDo[a- > b], IDSo[a- > b], FCS};
      end
    end
  end
end
```

Algorithm 1: Address anonymity.

distribution of traffic in a certain time period. If the attack nodes are deployed dense enough, the network traffic outline can be drawn by the adversary. Figure 7 illustrates such attack. Obviously, traffic outlining attack cannot be eliminated by address anonymity and random delay.

The solution to traffic outlining attack is circular traffic, which is illustrated in Figure 8. Network traffic from the source node to the base station follows two semicircle paths. And the two semicircle paths form a circular path. Source node selects one of the two semicircles randomly to forward packet. As to the adversary, traffic outlining attack cannot find the source node and base station because traffic in the networks forms a circular path (Algorithm 3).

The proposed solution can eliminate traffic outlining attacks effectively, because the source node and base station are hidden in a circular traffic outline. Combined with

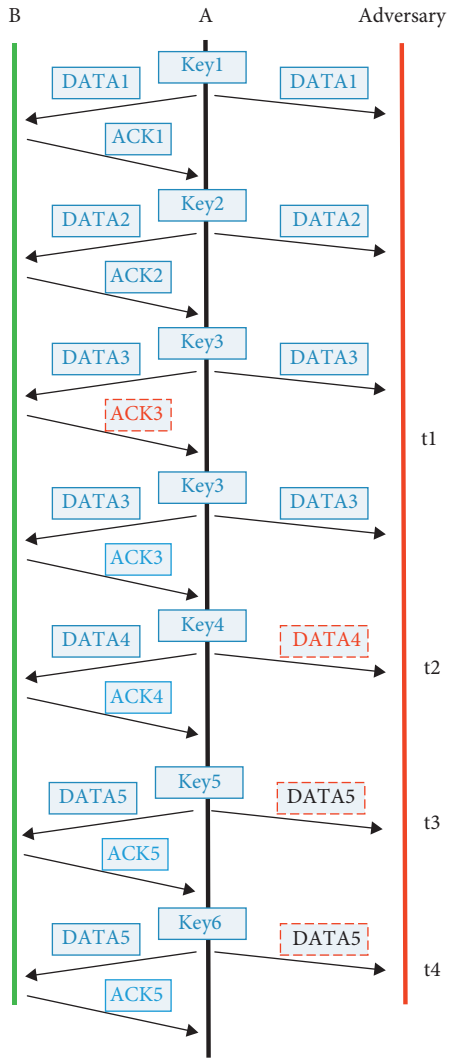

Figure 3: Frame error leads to address confusion.

address anonymity and random delays, traffic analysis attacks can be well addressed.

4.3. Performance Analysis. Efficiency is among the most important design considerations for data link layer schemes. Two different hash functions are used in HASHA, hash(), and HMAC(). hash() is used to generate hash value of payload, and the output is further used to create the key for HMAC(). According to the characteristics of HMAC() function, the computation complexity of brute-force attacks on hash key is 2^k, where k is the length of the key. Long key improves security strength. Therefore, the hashed results of hash() should be long enough to defend against brute-force attacks. Tiger/192 [38] is a good candidate for hash() because it is almost as fast as CRC32, but the width of the output is 192 bits. As HMAC() is used to create one-time address, performance is the top design consideration for HMAC() selection. Efficient MAC functions such as UMAC/32 [39] are a good candidate for HMAC().

Suppose that UMAC/32 is used for HMAC() and Tiger/192 is used for hash() in HASHA. According to the performance analysis of hash functions [38], the performance of UMAC/32 is 1 cycle per byte and Tiger/192 is 8.1 cycles per byte. From the illustrated HASHA process, hash() is called 1 time and HMAC() is called 2 times for both the sender and the receiver to transmit one frame. Supposing that the length of frame is *len* bytes and the MAC address is fixed to 6 byte, HASHA requires $len * 8.1 + 2 * 1$ cycles for one frame transmission and reception.

5. Performance Simulation

We use ns-2 to evaluate the energy consumption of HASHA. Several nodes are deployed over 200 m * 200 m network field, and the base station is located at the center. The nodes' radio transmission radius is 50 m.

We deploy only one source node to report event to the base station. The total number of nodes in WSN changes from 50 to 400 in 50 steps. We record the average power

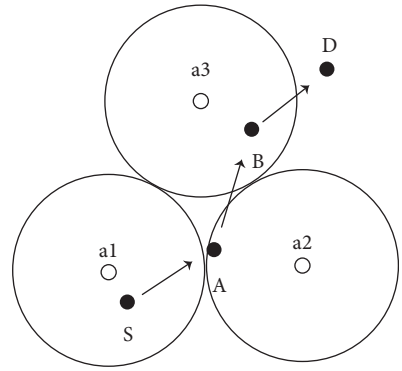

Figure 4: Example topology for time correlation attacks.

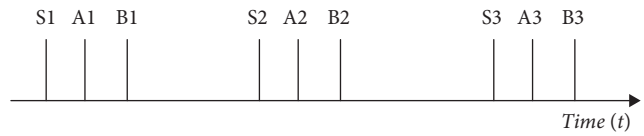

Figure 5: Time serial for reporting event from the source node to base station.

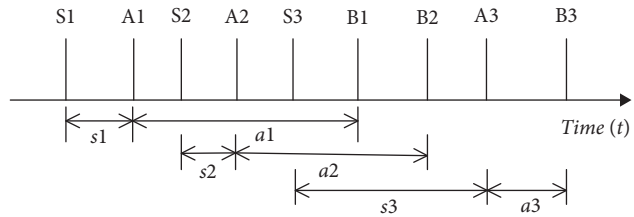

Figure 6: Nodes forwarding packets with random delays.

Node maintains a table with entry <*data*, *time_to_transmit*> to store data to be forwarded;
Node maintains clock timer, which is used for data transmission;
For data requested to be transmitted:
  Generate a random time *rand*;
  Insert into the table with entry <*data*, *timer* + *rand*>;
Node search the table to find data that could be transmitted:
  for each entry in the table **do**.
    if timer ≥ entry. time_to_transmit then
      Transmit the data;
    end
  end

Algorithm 2: Forwarding random delay.

consumption of HASHA and Phantom Routing [19], a well-known random location privacy preserve scheme. The power consumption of hash functions and wireless transmission and reception is listed in Table 1.

Figure 9 illustrates the overall power consumption of HASHA and Phantom Routing under different network size. While the size of the network is small (for example, 20 or 50 nodes), HASHA consumes more power than Phantom Routing. The reason is that hash operation is required for both transmitter and receiver, which introduce additional power consumption. Phantom Routing creates routing path longer than the shortest path. But as the network size is fairly small, the additional energy consumption for additional path is much less than hash operation. Therefore, the energy consumption of Phantom Routing is lower. As the size of network increased, the energy wasted on additional path increased dramatically. And that portion of energy cost is much greater than energy cost for hash operations.

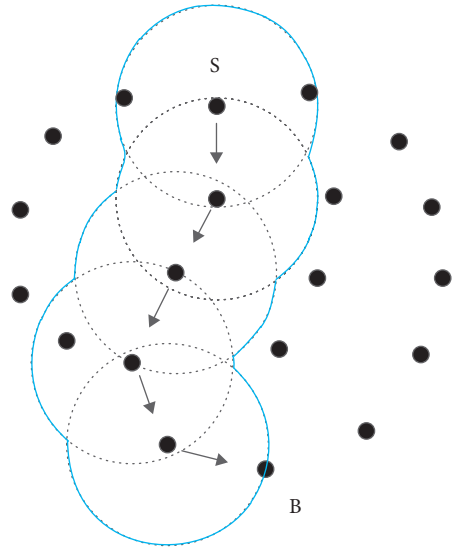

Figure 7: Traffic outlining attacks against event-triggered WSNs.

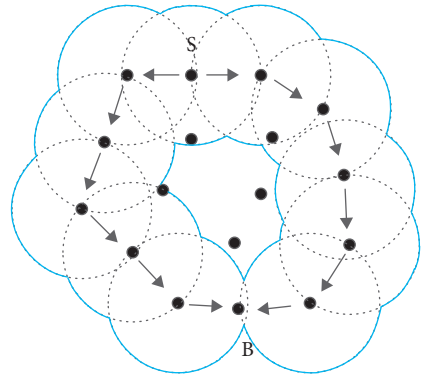

Figure 8: Circular traffic against traffic outlining attacks.

(1) Find the shortest path from the source node to base station according to routing protocol such as dijkstra.
(2) The base station calculates hops **n** from source node to base station and requests the node **n/2** hops away to initiate a circular forwarding path.
(3) The selected node broadcasts beacons which includes a counter with initial value **n/2**.
(4) All nodes that received the broadcasts decrease the value and forward it.
(5) All nodes that received the broadcasts with value 0 are candidates for circular forwarding.
(6) On having data to be sent, the source node selects one of the paths randomly to forward the data to the base station.

Algorithm 3: Circular forwarding against traffic outlining attacks.

Table 1: Power consumption of key operation.

Operation	Consumed energy
UMAC/32 hashing	0.143 uJ/byte
Tiger/192 hashing	0923 uJ/byte
Transmitting	5.623 uJ/byte
Receiving	6.39 uJ/byte

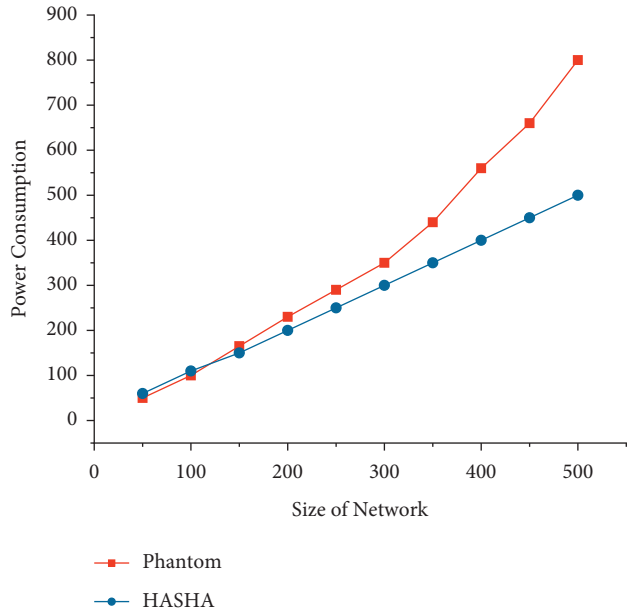

Figure 9: Power consumption of HASHA and Phantom Routing.

6. Conclusions

In this paper, we have identified that location privacy cannot be preserved efficiently at network layer, because address at data link layer is not protected well. The address at data link layer exposes node identity and packet routing information to the adversary. Traffic analysis attacks can be easily launched to jeopardize location privacy. HASHA scheme, which hides the addresses at data link layer, is proposed to protect location privacy. Analytical and simulation results show that HASHA is more energy efficient than traditional approaches [40, 41].

Data Availability

The simulation source file data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The initial version of this paper was published on IEEE International Conference on High Performance Computing and Communications. This is a substantial extension to the conference paper. The conference paper can be accessed at https://ieeexplore.ieee.org/document/8622908 [41].

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Qiong Zhang had the initial idea for this paper and rewrote the manuscript. Kewang Zhang conducted analysis and experiments.

References

[1] D. Kandris, C. Nakas, D. Vomvas, and G. Koulouras, "Applications of wireless sensor networks: an up-to-date survey," *Applied System Innovation*, vol. 3, no. 1, p. 14, 2020.

[2] S. Ali, T. Al Balushi, Z. Nadir, and O. K. Hussain, *WSN Security Mechanisms for CPS", Cyber Security for Cyber Physical Systems*, pp. 65–87, Springer, New York, NY, USA, 2018.

[3] W. Al Shehri, "A survey on security in wireless sensor networks," *International journal of Network Security & Its Applications*, vol. 9, no. 1, pp. 25–32, 2017.

[4] Q. Shafi, "Cyber physical systems security: a brief survey," in *Proceedings of the 12th IEEE International Conference on Computational Science and Its Applications*, pp. 146–150, Brazil, June 2012.

[5] Y. Li and J. Ren, "Preserving source-location privacy in wireless sensor networks," in *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pp. 1–9, IEEE, Italy, June 2009.

[6] L. Zhou and Y. Shan, "Multi-branch source location privacy protection scheme based on random walk in WSNs," *IEEE*, in *Proceedings of the 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis*, pp. 543–547, IEEE, China, April 2019.

[7] M. Conti, J. Willemsen, and B. Crispo, "Providing source location privacy in wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1238–1280, 2013.

[8] A. Tsitroulis, D. Lampoudis, and E. Tsekleves, "Exposing WPA2 security protocol vulnerabilities," *International Journal of Information and Computer Security*, vol. 6, no. 1, pp. 93–107, 2014.

[9] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *Proceedings of the 2nd ACM workshop on Security of Ad hoc and Sensor Networks, ser. SASN '04, ACM*, Washington, DC, USA, October 2004.

[10] D. Braginsky and D. Estrin, "Rumor routing algorithm for sensor networks," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, ser. WSNA '02, ACM*, vol. 9, pp. 22–31, ACM, New York, NY,USA, September 2002.

[11] E. Ekici, S. Vural, J. McNair, and D. Al-Abri, "Secure probabilistic location verification in randomly deployed wireless sensor networks," *Ad Hoc Networks*, vol. 6, no. 2, pp. 195–209, 2008.

[12] Y. Li, L. Lightfoot, and J. Ren, "Routing-based source-location privacy protection in wireless sensor networks," in *Proceedings of the IEEE International Conference on Electro/Information Technology*, vol. 6, pp. 29–34, IEEE, Canada, January 2009.

[13] I. Shaikh, H. Jameel, B. dAuriol, H. Lee, S. Lee, and Y.-J. Song, "Achieving network level privacy in wireless sensor networks," *Sensors*, vol. 10, no. 3, pp. 1447–1472, 2010.

[14] W. Yang and W. T. Zhu, "Protecting source location privacy in wireless sensor networks with data aggregation," *Ubiquitous Intelligence and Computing*, vol. 10, pp. 252–266, 2010.

[15] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "Towards a statistical framework for source anonymity in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 12, pp. 1–12, 2011.

[16] M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 10, pp. 1805–1818, 2012.

[17] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source-location privacy in sensor network routing," in *Proceedings of the 25th IEEE international conference on distributed computing systems*, pp. 599–608, IEEE, June 2005.

[18] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards statistically strong source anonymity for sensor networks," in *Proceedings of the The 27th Conference on Computer Communications, ser. INFOCOM 2008*, vol. 4, pp. 51–55, IEEE, Columbus, OH, USA, March 2008.

[19] H. Chen and W. Lou, "From nowhere to somewhere: protecting end-to-end location privacy in wireless sensor networks," in *Proceedings of the Performance Computing and Communications Conference, IEEE 29th International*, vol. 12, pp. 1–8, IEEE, Piscataway, USA, December 2010.

[20] S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensor network models," *ACM SIGMOBILE - Mobile Computing and Communications Review*, vol. 6, no. 2, pp. 28–36, 2002.

[21] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," *ACM* in *Proceedings of the first ACM conference on Wireless network security, ser. WiSec '08*, vol. 4, pp. 77–88, ACM, New York, NY, USA, March 2008.

[22] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing source-location privacy in sensor network routing," in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*, pp. 599–608, Columbus, OH, USA, June 2005.

[23] X. Hong, P. Wang, J. Kong, Q. Zheng, and J. Liu, "Effective probabilistic approach protecting sensor traffic," in *Proceedings of the IEEE Military Communication Conference*, pp. 169–175, Atlantic City, NJ, USA, 2005.

[24] S. Jiang and N. H. Vaidya, W. Zhao, "Routing in packet radio networks to prevent traffic analysis," in *Proceedings of the IEEE Information Assurance and Security Workshop*, West Point, NY, USA, February 2000.

[25] J. Deng, R. Han, and S. Mishra, "Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks," in *Proceedings of the IEEE International Conference on Dependable Systems and Networks*, Florence, Italy, July 2004.

[26] J. Deng, R. Han, and S. Mishra, "INSENS: intrusion-tolerant routing for wireless sensor networks," *Computer Communications*, vol. 29, no. 2, pp. 216–230, 2006.

[27] J. Deng, R. Han, and S. Mishra, "Decorrelating Wireless sensor network traffic to inhibit traffic analysis attacks," *Pervasive and Mobile Computing*, vol. 2, no. 2, pp. 159–186, 2006.

[28] H. Gao, C. Liu, Y. Yin, Y. Xu, and Y. Li, "A hybrid approach to trust node assessment and management for VANETs cooperative data communication: historical interaction perspective," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.

[29] H. Gao, L. Zhou, J. Y. Kim, Y. Li, and W. Huang, "The behavior guidance and abnormality detection for A-MCI patients under wireless sensor network," *ACM Transactions on Sensor Networks*, 2021.

[30] S. Olariu, M. Eltoweissy, and M. Younis, "ANSWER: autonomous wireless sensor network," in *Proceedings of the 1st ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks (Q2SWinet'05)*, pp. 88–95, Montreal, Quebec, Canada, October 2005.

[31] J. Kong, X. Hong, and M. Gerla, "An identity-free and on-demand routing scheme against anonymity threats in mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 8, pp. 888–902, 2007.

[32] Y. Yanchao Zhang, W. Wei Liu, W. Wenjing Lou, and Y. Fang, "MASK: anonymous on-demand routing in mobile ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2376–2385, 2006.

[33] H. Gao, Y. Zhang, H. Miao, R. J. D. Barroso, and X. Yang, "SDTIOA: modeling the timed privacy requirements of IoT service composition: a user interaction perspective for automatic transformation from bpel to timed automata," *Mobile Networks and Applications*, vol. 26, no. 6, pp. 2272–2297, 2021.

[34] H. Gao, X. Qin, R. J. D. Barroso et al., "Collaborative learning-based industrial IoT API recommendation for software-defined devices: the implicit knowledge discovery perspective," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 1, pp. 66–76, 2022.

[35] J. Al-Muhtadi, R. Campbell, A. Kapadia, and M. Dennis, "Routing through the mist: privacy preserving communication in ubiquitous computing environments," in *Proceedings of the 22nd International Conference on Distributed Computing Systems*, p. 74, Vienna, Austria, July 2002.

[36] Y. Huang, H. Xu, H. Gao, X. Ma, and W. Hussain, "SSUR: an approach to optimizing virtual machine allocation strategy based on user requirements for cloud data center," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 670–681, 2021.

[37] X. Ma, H. Xu, H. Gao, and M. Bian, "Real-time multiple-workflow scheduling in cloud environments," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4002–4018, 2021.

[38] J.-H. Park, Y. Jung, H. Ko, J. Kim, and M. Jun, "A privacy technique for providing anonymity to sensor nodes in a sensor network," in *Proceedings of the International*

Conference on Ubiquitous Computing and Multimedia Applications, Springer, Berlin, Heidelberg, 2011.

[39] J. Black, S. Halevi, H. Krawczyk, T. Krovetz, and P. Rogaway, "UMAC: fast and secure message authentication," in *Advances in Cryptology - CRYPTO' 99*, M. Wiener, Ed., vol. 1666, pp. 216–233, Springer, Berlin, Germany, 1999.

[40] S. Misra and G. Xue, "Efficient anonymity schemes for clustered wireless sensor networks," *International Journal of Sensor Networks*, vol. 1, no. 1/2, pp. 50–63, 2006.

[41] K. Zhang and Q. Zhang, "Preserve location privacy for cyber-physical systems with addresses hashing at data link layer," in *Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications*, pp. 1028–1032, Exeter, UK, June 2018, https://ieeexplore.ieee.org/document/8622908.