Location Privacy Issues in Wireless Sensor Networks

Jiří Kůr and Andriy Stetsko

Faculty of Informatics
Masaryk University
{xkur,xstetsko}@fi.muni.cz

Abstract. We discuss location privacy issues in wireless sensor networks. We consider sensor nodes with more responsible roles and the need to protect locations of such nodes. Available countermeasures against various types of traffic analysis attacks are examined and their problems are identified. We do not propose new traffic analysis resistance technique. Instead, we draw attention to blanks in current situation and identify several open questions, which should be answered in order to ensure location privacy of nodes.

1 Introduction

A wireless sensor network (WSN) is a heterogenous network composed of a large number of tiny low-cost devices, denoted as nodes, and a few general-purpose computing devices referred to as base stations. A general purpose of the WSN is to monitor some physical phenomenons (e.g. temperature, barometric pressure, light) inside an area of deployment.

Nodes are equipped with a communication unit (e.g., radio transceiver), processing unit, battery and sensor(s). General nodes are constrained in processing power and energy, whereas base stations (also called sinks) have laptop capabilities and unlimited energy resources. The base stations act as gateways between the WSN and other networks (e.g., Internet).

There is a wide variety of applications for WSNs [1], ranging from military applications (e.g., battlefield surveillance) through environmental (e.g., forest fire detection) to health applications (e.g., patient health monitoring). However, security of the WSNs has to be examined prior to their massive deployment.

In this paper we consider large WSNs with thousands of nodes, which have static geographical locations. We primarily aim to defend nodes with additional responsibilities (e.g., base stations) against traffic analysis which helps to reveal the locations. We do not propose new solutions for the problem, but we recapitulate and analyse the state-of-the-art countermeasures. Furthermore, we put down design considerations which have to be taken into account while designing new traffic decorrelation techniques. This paper is a starting point for the future research in this area and identifies several open questions.

V. Matyáš et al. (Eds.): The Future of Identity, IFIP AICT 298, pp. 160–169, 2009.

2 Identity and Location Privacy in WSN

In the context of WSNs we propose to use the following definition of node identity [8][1]: "An identity is any subset of attributes, which sufficiently identifies the node within any set of nodes. So, usually there is no such thing as 'the identity', but several of them". So far, we have identified several attributes, which can constitute an identity in itself or in combination with other ones:

- *Unique ID.* An application dependent identifier. An artificial identity used for the purposes of a particular application. It is typically represented by a bitstring. This identity can be abused not only on the particular application level, but also on other levels, for example, for tracking messages.
- *Global network address.* Typical identity in conventional networks. This identity can be forged by the adversary to attack, for example, routing algorithm, e.g. Sybil attack. However, global addressing scheme is not always implemented in WSNs due to resource limitations.
- *Local network address.* Local network address is not unique within the whole network and therefore it cannot in itself constitute the identity. However, it is unique within a neighborhood and thus can be used in combination with other attributes.
- *Sensed data.* Sensed data are often used to address nodes. For example, base station requests data from the nodes, which experience temperature above specified value. This data-centric approach substitutes classic network addressing schemes. Thus, sensed data can be considered as an attribute of node's identity. By these data, an adversary can identify a particular node or at least significantly reduce an anonymity set. For example, she can be interested in the location of nodes, which have experienced specific temperature during last hour. So, in this case the specific temperature represents a node identity.
- *Geographic location.* Geographic location of sensed data is very important (e.g. forest fire detection). We consider this location as an attribute, which comprise a special case of identity. We call this *location identity.* Protecting the location identity means ensuring a location privacy.

In order to ensure node location privacy the following requirements should be fulfilled [10]: (a) no one knows the exact location of the node, except itself; (b) other nodes, typically intermediate nodes on route, have no information about their distance, i.e. the number of hops, from that node.

Together with a node identity we can also define its *role.* The role is an expected behavior (i.e., sequence of actions) in a given context[8][2]. There is a number of roles in WSNs (e.g., a "base station" role, a "cluster head" role or

[1] In the original definition a term "person" is used instead of "node".

[2] "In sociology, a "role" or "social role" is a set of connected actions, as conceptualized by actors in a social situation (i.e., situation-dependent identity attributes). It is mostly defined as an expected behavior (i.e., sequences of actions) in a given individual social context."[8].

"sensing node" role). Some of them involve more responsibilities and hence nodes in such roles are more tempting for an adversary than other ones (e.g., a base station, a cluster head). In the rest of the paper, we call these nodes *important nodes*.

In order to protect the important node we should provide an unlinkability of the node's identity and its role. For example, if an adversary links a node's location identity to the "base station" role, she can either isolate the node (base station) or physically destroy it and hence ruin the whole network. Therefore, we primarily aim to ensure the location privacy of the important nodes.

3 Traffic Analysis

Traffic analysis attacks help to reveal communication patterns, which allow an adversary to deduce a location of important nodes and then to compromise or to destroy them. Three classes of the traffic analysis attack are identified in WSNs: the rate monitoring attack, the time correlation attack and the content analysis attack.

In the rate monitoring attack, an adversary observes nodes sending packet rate and moves closer to the node, which has the highest sending packet rate. In the time correlation attack, an adversary monitors a correlation in sending times between a node and its neighbors. The adversary tries to detect which node forwards the current packet and traces the path directly to a base station. In the content analysis attack, an adversary tries to obtain a valuable information (e.g., a base station location) from packet headers and payloads.

In [3] the authors employ two metrics to evaluate effectiveness of proposed mechanisms against the rate monitoring attack. First, an entropy metric measures a randomness of network traffic. Second, a heuristic-based algorithm combines a hill-climbing search algorithm with a random restart mechanism. In this algorithm, an adversary starts at some location and monitors network traffic within his/her range. The adversary moves to the node with the highest sending rate. In case he/she reaches a local sending rate maximum he/she selects another location at random and repeats this algorithm. This algorithm is based on the rate monitoring analysis and can be used by an adversary to locate a base station.

4 State-of-the-Art

Many solutions that provide location privacy and anonymity have been proposed for MANETs [6,9,10]. However, these schemes are not suitable for highly resource-constrained WSNs where the predominant traffic pattern is many-to-one.

Only a few satisfying solutions have been proposed for WSNs. Encryption techniques can be used to hide a destination address, a packet type and a packet payload. However, the end-to-end encryption does not solve a problem with a packet appearance, which remains the same along the path. In order to make it harder for an attacker to trace packets to a base station, a hop-by-hop re-encryption scheme should be used. However, this technique introduces extra delays in forwarding packets. In [2] the authors propose a mechanism which

defends against the rate monitoring attack. The packet to send is repeatedly transmitted by a child node until the packet is accepted by a parent node. If the child node has no packet to send it injects a dummy packet. This mechanism ensures a uniform sending rate across the entire network but significantly decreases the lifetime of WSN. In order to defend against the time correlation attack, the authors of [5] propose to add random delays to packet retransmission at each forwarding node. Tolerance against base station isolation might be increased by usage of several base stations [2]. In this approach a node sends packets to the different base stations.

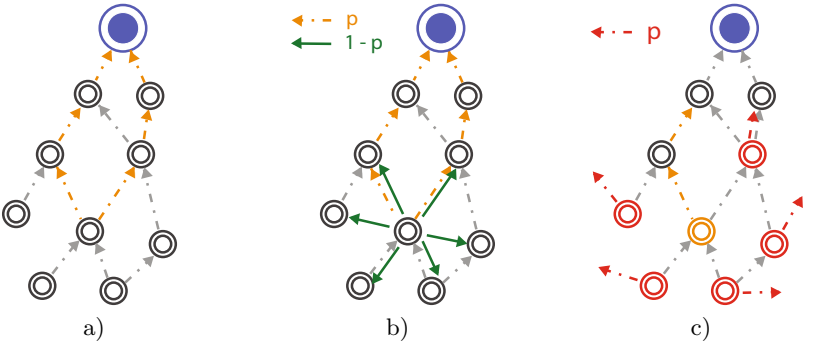

Fig. 1. Multi-parent routing schemes

More advanced techniques are presented in [4]. The rate monitoring attack can be partially prevented by a multiple parent routing scheme since traffic spreads along multiple paths. Each node has multiple parent nodes, which route packets to a base station, see Fig.1 a). In order to forward a packet, a node randomly selects one of its parent nodes. This scheme can be extended by a controlled random walk. A node forwards a packet to one of its parent nodes with probability p. With probability $1 - p$ the node forwards the packet to one of its neighbors – this does not eliminate the fact that the node selects a parent node, see Fig.1 b). This technique introduces delivery time delays, which are proportional to extra hops used for forwarding the packets. This technique is still vulnerable to the time correlation attack. Therefore, the authors propose a new technique called the multi-parent routing scheme with fractal propagation. When a node hears that a neighbor forwards a packet to a base station, the node generates a fake packet with probability p and forwards it to one of its neighbors, see Fig.1 c). The main problem with this technique is that it generates a large amount of traffic near a base station. This can be solved by a differential fractal propagation technique. When a node forwards packets more frequently it sets a lower probability for creating new fake packets. In order to make traffic analysis more difficult, the authors propose to generate artificial areas (called hot-spots) of high communication activity.

In order to minimize the damage of nodes compromise, the authors propose to use a directional pairwise identification mechanism [4], such that each node

uses a different identification for communication with its child nodes and parent nodes. That means that a compromised child node does not know a parent node identification used to send data. This idea is extended in [7]. A base station is responsible for routing process and it assigns incoming and outgoing labels to nodes in both uplink and downlink directions. Each node only forwards packets with labels which match either its downlink or uplink label.

5 Suggested Improvements

In this section we try to identify shortcomings of some techniques presented in the previous section.

In [4] authors propose to use a cluster key both to protect a packet content and to change a packet appearance. Each node possesses the neighbors' cluster keys and uses them to decrypt packets and determine whether these packets are fake or original ones. Thus, by capturing a single node, an adversary can decrypt traffic within its neighborhood and reveal the neighbors' IDs. By observing a traffic pattern in the neighborhood, the adversary may estimate the potential direction of the packet's path and determine a next node to capture. The more nodes the adversary controls, the more accurate is his/her estimation. Consequently he/she can easily track the packets to a base station by subsequent capture of the nodes.

We propose to use double encryption to mitigate an impact of the node capture. First, the packet is encrypted by a pairwise key, then an information whether the packet is fake or original is concatenated and finally the result is encrypted with the cluster key. This reduces area compromised by a single captured node from the node's neighborhood to the node itself. By capturing a node, the adversary can still distinguish fake packets from original ones in the whole neighborhood, but he/she can completely decrypt only packets passing right through the captured node instead of packets passing through its neighbors. Thus, an adversary has to capture significantly larger number of nodes in order to trace packets to a base station. We assume that the amount of original packets passing through the nodes is large enough to prevent simple time correlation attacks and a decryption is needed to link two distinct packets.

We have also encountered a problem with the differential fractal propagation scheme, which had been proposed in [4]. In order to prevent a creation of a large amount of traffic near a base station, authors propose a mechanisms, which ensures that nodes with higher sending rate (nodes closer to a base station) generate fake packets with lower probabilities. This rule applies to the nodes with a sending rate higher than some threshold h. Thus, there is an area around the base station, in which these probabilities indicate a distance of the node to the base station. Therefore, by capturing the nodes from this area an adversary can easily find out whether he/she moves closer to the base station or not. The size of the area depends on the threshold value h. The area has to be large enough to prevent a creation of a significant amount of traffic near the base station. On the other hand the larger the area is the higher is a probability that the adversary captures a node within this area. Hence, the area size should be small enough to prevent an adversary from capturing a node within the area.

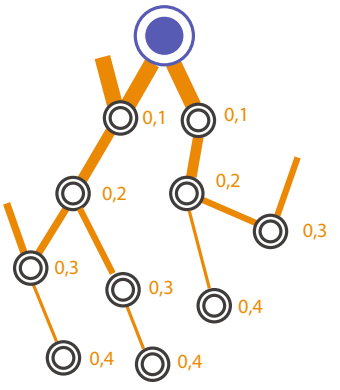

Fig. 2. Differential fractal propagation scheme

There is a question whether it is possible to find a solution that would satisfy both requirements. Obviously, we can find the tradeoff that would minimize the probability of capturing a node within the area while still keeping the amount of traffic near base station at reasonably low level. While being an optimal tradeoff, the size of area can be still too large to apply this technique in practice.

6 Open Questions

In this section we want to introduce several open questions, which we have encountered during the study of identity and location privacy issues in WSNs.

6.1 Attacker Model

Some authors assume an attacker model with a global eavesdropper, which monitors all possible communication links between all nodes all the time. We think that this model is not adequate for the most WSN applications and we should assume an adversary, which observes only a limited part (but variable for different applications) of the network. Someone might argue that the global eavesdropper model is stronger and hence by taking into account this model we gain a better security in WSN. It is not necessary true. In order to defend against a stronger attacker we design a stronger countermeasures, which typically consume more energy resources and hence decrease a WSN lifetime.

The global eavesdropper model is a passive attacker model. However, WSNs are not physically protected and it is necessary to assume an active attacker, which in addition to the traffic analysis attack can launch a variety of other attacks (e.g., attacks on network layer). However, authors of countermeasure techniques described in the section 4 do not consider the presence of internal attacks. In some cases it turns into an increase of damage done by these internal attacks. For example, in the sinkhole attack, where an adversary tries to attract all traffic destined to the base station, the traffic analysis countermeasures presented in the section 4 create useless traffic and waste network energy resources. By performing the sinkhole attack, the adversary can drain batteries of the nodes in its neighborhood.

Also it might happen that traffic analysis countermeasures hinder in detection of malicious nodes. On one hand, by employing traffic analysis countermeasures we want to hide some behavior patterns (e.g., nodes sending rates). On other hand we want to keep these patterns in order to detect attacks such as selective forwarding attack.

The countermeasures presented in section 4 are based on the multi-parent routing scheme. However, the multi-parent routing scheme has not to be necessary multi-parent in the presence of Sybil nodes, each of which owns several identities. Therefore, there is a need to analyze the influence of the attack on the proposed techniques.

6.2 Location Privacy of Other Important Nodes

A base station is often considered as the only important node, whose location privacy has to be protected. However, WSNs might include other important nodes, for example cluster heads or nodes, which run intrusion detection system (IDS). There is a need to search for new sources of information, which might reveal a location of these nodes. There are two types of IDS: cooperative and non-cooperative. The cooperative IDSs may have either peer-to-peer or hierarchical architecture. In both architectures, IDS nodes share a detection state information, which might be used to deduce their location. Since there is no sense to employ an IDS without a response system, we assume that a detection of an attack is always followed by actions, which try either to stop or prevent the attack. Therefore, an adversary can capture a node and generate internal attacks in order to force the IDS to react and provide the adversary with traffic, which is sufficient to locate IDS nodes. Some attacks might be detected locally (e.g., blackhole attack) but others require a cooperation from IDS nodes (e.g., wormhole attack). This fact can be exploited by an adversary to force IDS nodes to cooperate and hence produce additional traffic.

In some scenarios a base station is involved in the response system. It might flood a WSN with certain actions, which should be taken by nodes in order to stop or prevent an attack. We are not aware of decorrelation techniques or attacks, which take into account a traffic broadcasted by a base station. This might provide an adversary with new possibilities to reveal a base station location.

A proper IDS placement strategy has to be chosen in order to minimize energy consumption and maximize an IDS effectiveness. For example, as a criteria of effectiveness we may choose a number of IDS nodes, a volume of analyzed traffic and accuracy of detection. It is more "effective" to deploy IDSs in traffic concentration points. Hence, by performing the rate monitoring attack an adversary can find out a location of IDS nodes. In general, taking into account the placement strategy an adversary can significantly reduce a set of nodes, which may run an IDS and hence significantly increase a probability to locate these nodes. We are not aware of placement strategies, which take into account a location privacy issue. It means that an adversary might find a large subset of IDS nodes and compromise them in order to subvert detection results. Obviously, the traffic analysis countermeasures can be applied in this case. However, they can be

more resource consuming than a design, which counts with the location privacy issue from the beginning. Therefore, we think that a location privacy should be considered as one of parameters of IDS placement strategy effectiveness.

The problem with the placement strategy holds also for base stations. In most cases, when designing a placement strategy, efficiency is the only design consideration and its actual meaning is dependent on a particular application. Sometimes we need a low latency, whereas in other case we prefer a strong energy awareness. Anyway, the location privacy needs are neglected. Suppose the adversary knows the area of deployment and the placement strategy. In that case he/she can focus his/her attention only to the particular places, where the probability of the base station being placed is the highest.

6.3 A WSN Model

The strength of proposed decorrelation techniques has to depend on a variety of input parameters. As long as these parameter are not taken into account, the decorrelation technique either underestimates or overestimates some threats. The underestimating of threats may lead to an easy compromise of the whole network and their overestimating may lead to a wastage of energy due to excessive prevention, detection or reaction mechanisms, which in turn significantly decrease a total WSN lifetime. We think that used WSN models are too simplified and techniques proposed under these models can be employed only under very limited circumstances. Ideally, the required strength of decorrelation technique should change gradually according to the changes of input parameters. In this section we present some parameters, which we think have to be taken into account while designing traffic analysis countermeasures.

There is a need to consider a presence of several base stations since they can share responsibility and mitigate network traffic patterns. Authors propose sophisticated decorrelation mechanisms assuming a presence of only one base station. We believe, that in some cases it would be more profitable to use several base stations instead of the decorrelation technique at all. Both situations are extremes and will be realistic only under very limited circumstances. The more realistic case is when there are several base stations and a decorrelation technique is employed. In order to ensure a wider application range of decorrelation technique, its strength should change according to the number of available base stations. Intuitively, the higher number of base station is, the more mitigated traffic patterns are and hence the less complex decorrelation technique we need. There is a need to find a trade-off between a number of base-station and a complexity of decorrelation technique used.

Different WSN applications require different sending data rates. These data rates may be so small that it takes much more time to perform traffic analysis and locate a base station than a total lifetime of a WSN, which does not employ any decorrelation technique. So, the strength of decorrelation technique depends on the sending data rate parameter and should vary for different WSN applications with different sending rates.

A number of nodes and their density also have an impact on the required strength of decorrelation technique. We can imagine a WSN network, where a number of nodes is not high and the deployed area is relatively small for adversary to find out a location of base station(s). In that case there is no sense to employ any decorrelation technique at all. In the section 4 there are discussed decorrelation techniques, which are based on the multi-parent routing scheme. However, the density of the nodes might be so low that no multi-parent routes will exist. In that case the decorrelation techniques also will not be profitable at all. We conclude, that a number of nodes and their density should be also taken into account when designing traffic analysis countermeasures.

There is a need to deeply analyze relations between different parameters and to propose an adequate WSN model. The more precise model we make, the more effective solution we will be able to obtain. We are aware of the fact that very precise model might not lead to any solution at all as we will not be able to solve the problem under this model. However, we believe that currently available models are too simplified and are not adequate enough.

7 Summary

We examined identity issues in WSNs and showed the need to ensure a location privacy of important nodes, especially base stations. State-of-the-art traffic analysis techniques and countermeasures against them were briefly described. We tried to identify shortcomings and blanks of the presented countermeasures. We also outlined problems of hiding a location identity of the important nodes. It comes out, that this area is not adequately examined yet and new mechanisms have to be designed. We identified several open questions, which, we hope, will encourage interesting discussions. We think that a field of identity and location privacy in the WSNs has a great potential for the future research.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks 38(4), 393–422 (2002)
2. Deng, J., Han, R., Mishra, S.: Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks. In: DSN 2004: Proceedings of the 2004 International Conference on Dependable Systems and Networks, Washington, DC, USA, p. 637. IEEE Computer Society, Los Alamitos (2004)
3. Deng, J., Han, R., Mishra, S.: Countermeasures against traffic analysis attacks in wireless sensor networks. In: SECURECOMM 2005: Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks, Washington, DC, USA, pp. 113–126. IEEE Computer Society, Los Alamitos (2005)
4. Deng, J., Han, R., Mishra, S.: Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks. Pervasive and Mobile Computing 2(2), 159–186 (2006)
5. Hong, X., Wang, P., Kong, J., Zheng, Q., Liu, J.: Effective probabilistic approach protecting sensor traffic. In: Military Communications Conference, vol. 1 (2005)

6. Kong, J., Hong, X.: Anodr: anonymous on demand routing with untraceable routes for mobile ad-hoc networks. In: MobiHoc 2003: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, pp. 291–302. ACM, New York (2003)
7. Nezhad, A.A., Makrakis, D., Miri, A.: Destination Controlled Anonymous Routing in Resource Constrained Multihop Wireless Sensor Networks. In: Wireless Sensor and Actor Networks, December 2007. IFIP International Federation for Information Processing, vol. 248, pp. 83–94. Springer, Boston (2007)
8. Pfitzmann, A., Hansen, M.: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management - a consolidated proposal for terminology (February 2008)
9. Seys, S., Preneel, B.: Arm: Anonymous routing protocol for mobile ad hoc networks. In: AINA 2006: Proceedings of the 20th International Conference on Advanced Information Networking and Applications, Washington, DC, USA, vol. 2, pp. 133–137. IEEE Computer Society, Los Alamitos (2006)
10. Zhu, B., Wan, Z., Kankanhalli, M.S., Bao, F., Deng, R.H.: Anonymous secure routing in mobile ad-hoc networks. In: LCN 2004: Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, Washington, DC, USA, pp. 102–108. IEEE Computer Society, Los Alamitos (2004)