Ad Hoc Networks 7 (2009) 1501–1514

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Privacy preservation in wireless sensor networks: A state-of-the-art survey

Na Li [a,*], Nan Zhang [b], Sajal K. Das [a], Bhavani Thuraisingham [c]

[a] Department of Computer Science and Engineering, The University of Texas at Arlington, Box 19015, 416 Yates St., Room 300, Nedderman Hall, Arlington, TX 76019-0015, United States
[b] Department of Computer Science, The George Washington University, 801 22nd Street NW, Suite 704, Washington DC 20052, United States
[c] Department of Computer Science, Erik Jonsson School of Engineering & Computer Science, The University of Texas at Dallas, 800 W. Campbell Road, MS EC31, Richardson, TX 75080, United States

ARTICLE INFO

Available online 22 April 2009

Keywords:
Wireless sensor network
Privacy

ABSTRACT

Much of the existing work on wireless sensor networks (WSNs) has focused on addressing the power and computational resource constraints of WSNs by the design of specific routing, MAC, and cross-layer protocols. Recently, there have been heightened privacy concerns over the data collected by and transmitted through WSNs. The wireless transmission required by a WSN, and the self-organizing nature of its architecture, makes privacy protection for WSNs an especially challenging problem. This paper provides a state-of-the-art survey of privacy-preserving techniques for WSNs. In particular, we review two main categories of privacy-preserving techniques for protecting two types of private information, data-oriented and context-oriented privacy, respectively. We also discuss a number of important open challenges for future research. Our hope is that this paper sheds some light on a fruitful direction of future research for privacy preservation in WSNs.

1. Introduction

In recent years, wireless sensor networks (WSNs) have drawn considerable attention from the research community on issues ranging from theoretical research to practical applications. Special characteristics of WSNs, such as resource constraints on energy and computational power, have been well defined and widely studied [3]. What has received less attention, however, is the critical *privacy concern* on information being collected, transmitted, and analyzed in a WSN. Such private information of concern may include *payload data* collected by sensors and transmitted through the network to a centralized data processing server. For example, a patient's blood pressure, sugar level and other vital signs are usually of critical privacy concern when monitored by a medical WSN which transmits the data to a remote hospital or doctor's office. Privacy concerns may also arise beyond data content and may focus on *context information* such as the location of a sensor initiating data communication. Note that an alert communication originating from a patient's heart monitor in the medical WSN is enough for an adversary to infer that the patient suffers from heart problem. Effective countermeasure against the disclosure of both *data* and *context-oriented* private information is an indispensable prerequisite for the broad application of WSNs to real-world applications.

Privacy protection has been extensively studied in various fields related to WSN such as wired and wireless networking, databases and data mining. Nonetheless, the following inherent features of WSNs introduce unique challenges for privacy preservation in WSNs, and prevent the existing techniques from being directly transplanted:

- *Uncontrollable environment*: Sensors may have to be deployed to an environment uncontrollable by the defender, such as a battlefield, enabling an adversary to launch physical attacks to capture sensor nodes or deploy counterfeit ones. As a result, an adversary may retrieve private

* Corresponding author. Tel.: +1 8172727409.
E-mail addresses: na.li@mavs.uta.edu (N. Li), nzhang10@gwu.edu (N. Zhang), das@exchange.uta.edu (S.K. Das), bhavani.thuraisingham@utdallas.edu (B. Thuraisingham).

1570-8705/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2009.04.009

keys used for secure communication and decrypt any communication eavesdropped by the adversary.

- *Sensor-node resource constraints*: A battery-powered sensor node generally has severe constraints on its ability to store, process, and transmit the sensed data. As a result, the computational complexity and resource consumption of public-key ciphers is usually considered unsuitable for WSNs. This introduces additional challenges for privacy preservation.
- *Topological constraints*: The limited communication range of sensor nodes in a WSN requires multiple hops in order to transmit data from the source to the base station. Such a multi-hop scheme demands different nodes to take diverse traffic loads. In particular, a node closer to the base station (i.e., data collecting and processing server) has to relay data from nodes further away from base station in addition to transmitting its own generated data, leading to higher transmission rate. Such an unbalanced network traffic pattern brings significant challenges to the protection of context-oriented privacy information. Particularly, if an adversary holds the ability of global traffic analysis, observing the traffic patterns of different nodes over the whole network, it can easily identify the sink and compromise context privacy, or even manipulate the sink node to impede the proper functioning of the WSN.

The unique challenges for privacy preservation in WSNs call for the development of effective privacy-preserving techniques. In this paper, we provide a state-of-the-art survey of existing privacy-preserving techniques in WSNs. We review two main categories of privacy-preserving techniques for protecting two types of private information, data-oriented and context-oriented privacy, respectively. In the category of data privacy, we mainly discuss how to enable the aggregation of sensed data without violating the privacy of the data being collected and guarantee the privacy of data query initiated by users of the network. For context-based privacy, we analyze the protection of location privacy and temporal private information. In addition, we build a table to compare different techniques in terms of their effectiveness in practical applications. Last but not the least, we discuss some interesting and challenging open issues on this topic, which are expected to shed light on a fruitful direction of future research on privacy preservation in WSNs.

The rest of the paper is organized as follows. We review privacy-preserving techniques in related fields in Section 2. In Section 3, we introduce our taxonomy of privacy-preserving techniques in WSNs. Sections 4 and 5 address techniques for data and context-oriented privacy protection, respectively. In Section 6, we evaluate and compare the performance of different privacy-preserving techniques. Section 7 outlines the open challenges for future research, followed by final remarks in Section 8.

2. Related work

Research on issues related to WSNs requires multi-disciplinary studies spanning networking, databases, distributed computing, etc. To properly understand the challenges of privacy preservation in WSNs and the techniques necessary to address such challenges, it is important to first examine the privacy issues and privacy-preserving techniques in such related fields as databases, data mining and wireless networks, which we briefly review as follows.

In the field of database and data mining, privacy concerns may arise from three types of systems [36]: The first is an information sharing system which involves two or more mutually untrusted parties. The objective is to guarantee that no private information beyond the minimum necessary is disclosed during information sharing. Cryptographic secure multi-party computation techniques are usually used for this type of systems [1,11,35,38]. The second is a data collection system where one centralized data collector/analyzer collects and mines data from multiple distributed data providers. Random perturbation techniques [2,15,33,34] are usually applied to protecting privacy in these systems. The third type is a data publishing system, the objective of which is to publish data to support data analytical application without compromising the anonymity of individual data owners. k-anonymity [27] and l-diversity based algorithms [19,20] are proposed for privacy protection in these systems.

Privacy issues have also been extensively studied in the domain of generic networking. Location privacy is of particular concern with the pervasive development of advanced wireless device, like PDA, and with the advent of location-based service (LBS). In an LBS system, a user holding a wireless device queries the LBS server to obtain the nearest restaurant or hospital to the user. Nonetheless, the user would not willingly disclose his/her real location. To address such location privacy concerns, ANONYMIZER, a trusted-third-party based framework, was proposed [13]. With this framework, a user first sends his/her location to the centralized anonymizer which then queries the LBS server with not the user's real location but a cloaking region which covers not only the user but also a number of other users. This technique prevents the LBS server from distinguishing one user from many others. However, it is unlikely to be practical for two reasons: First, it is not reasonable to assume the existence of a trustable third party. Second, even if such a trusted third party exists, it creates a single point-of-failure for the system because if the third party is compromised, the privacy preservation over the whole system will completely collapse. To remove the requirement of a trusted third party, a private-information-retrieval based technique was proposed [12]. Nonetheless, this technique also suffers from significant computation and communication overhead. Besides the direct disclosure of user's location from query payload, traffic flow information may also breach location privacy. In particular, the server may pinpoint the location of a user based on the user's IP address. In order to provide traffic flow confidentiality, Tor [10] is designed, as the second generation onion routing [22], and becomes a popular anonymous communication network which consists of thousands of Tor routers to relay user traffic to or from LBS server.

3. Taxonomy of privacy-preserving techniques for WSNs

In this section, we introduce our taxonomy of privacy-preserving techniques for WSNs. Fig. 1 depicts the complete taxonomy which includes the problems defined (as elipses) as well as the techniques proposed (as rectangles) in the literature. One can see from the figure that there are two main types of privacy concerns, data-oriented and context-oriented concerns. Data-oriented concerns focus on the privacy of data collected from, or query posted to, a WSN. On the other hand, context-oriented concerns concentrate on contextual information, such as the location and timing of traffic flows in a WSN. Data- and context-oriented privacy concerns may be violated by data- and traffic-analysis attacks, respectively. Fig. 2 depicts a simple illustration of the two types of attacks. One can see from Fig. 2 that, in the case of data analysis attack, a malicious node of the WSN abuses its ability of decrypting data to compromise the payload being transmitted. In traffic-analysis attacks, a third-party adversary does not have the ability to decrypt data payloads. Instead, it eavesdrops the wirelessly transmitted data and tracks the traffic flow information hop-by-hop. In the following, we provide a brief overview of the problem definitions and main challenges for these two categories, respectively.

3.1. Overview of data-oriented privacy protection

Data-oriented privacy protection focuses on protecting the privacy of data content. Here "data" refer to not only *sensed data* collected within a WSN but also *queries* posed to a WSN by users. In the above-mentioned medical WSN example, private information may include temperature and blood pressure collected from the WSN, or queries on these vital signs posed to the WSN.

There are two types of adversaries which may compromise data-oriented privacy. One is an *external* adversary which eavesdrops the data communication between sensor nodes in a WSN. This type of adversaries can be effectively defended against using the traditional techniques of cryptographic encryption and authentication.

The second type is an *internal* adversary which is also a participating node of the WSN, but has been captured and manipulated by malicious entities to compromise private information. Since a participating node is allowed to decrypt data legally, the traditional encryption and authentication techniques may no longer be effective. Thus, the main challenge for protecting data-oriented privacy is to prevent an internal adversary from compromising the private information, while maintaining the normal operation of the WSN.

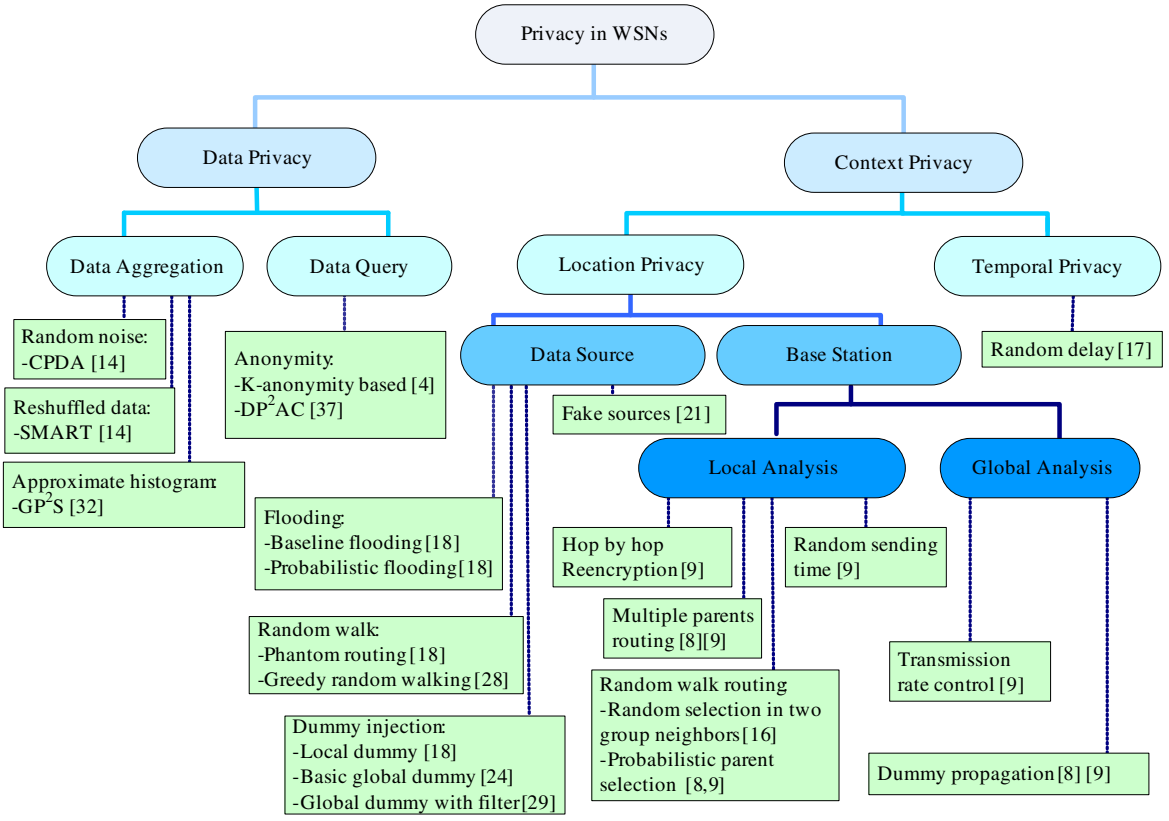

Fig. 1. Taxonomy of privacy-preserving techniques for WSNs.

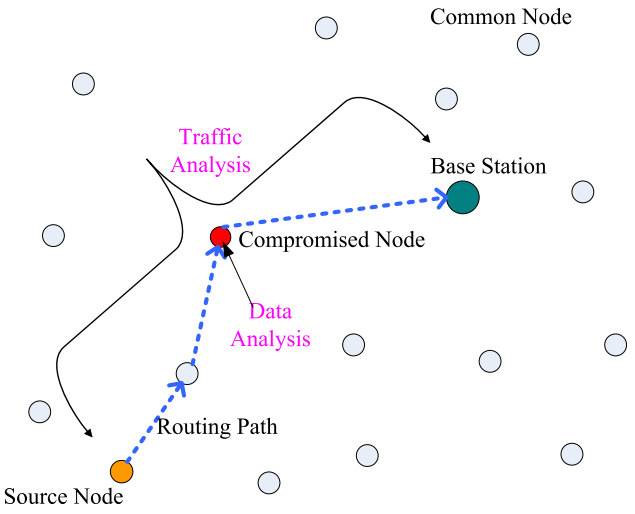

Fig. 2. Two scenarios for privacy attack in a WSN.

3.2. Overview of context-oriented privacy protection

Context-oriented privacy protection focuses on protecting contextual information, such as the *location* [7–9,16,18,21,24,28,29] and *timing* [17] information of traffic transmitted in a WSN. Location privacy concerns may arise for such special sensor nodes as the data source [18,21,24,28,29] and the base station [7–9,16]. As we mentioned in Section 1, an adversary with knowledge of the location of the data source or base station location may be able to infer the content of the data being transmitted or destroy the sensor network.

Timing privacy, on the other hand, concerns the time when sensitive data is created at data source, collected by a sensor node and transmitted to the base station [17]. This type of privacy is also of primary importance, especially in the mobile target tracking application of WSNs, because an adversary with knowledge of such timing information may be able to pinpoint the nature and location of the tracked target without learning the data being transmitted in the WSN [17]. Furthermore, the adversary may be able to predict the moving path of the mobile target in the future, violating the privacy of the target.

Similar to data-oriented privacy, context-oriented privacy may also be threatened by both external and internal adversaries. Nonetheless, existing research has mostly focused on defending against external adversaries, because such adversaries may be able to compromise context privacy easily by monitoring wireless communication. Within the category of external adversaries, one can further classify adversaries into two categories, *local attackers* and *global attackers*, based on the strength of attacks an adversary is capable of launching. Local attackers can only monitor a local area within the coverage area of a WSN, and therefore have to analyze traffic *hop-by-hop* to compromise traffic context information. On the other hand, a global attacker has the capability (e.g., a high-gain antenna) of monitoring the global traffic in a WSN. One can see that a global attacker is much stronger than a local one.

We would like to remark that the classification of context-oriented privacy into the two categories of location and timing privacy only reflects the current state-of-the-art, and should not be treated as a comprehensive classification. Context-oriented privacy also extends to other concerns, such as the hiding of time frequency of communication between sensor nodes in a WSN, because such frequency may expose information about the traffic flow in a WSN. In the above-mentioned medical WSN, knowing that a patient's blood pressure sensor is having frequent communication with its neighboring nodes is enough for an adversary to infer that the patient may be suffering from high or low arterial pressure.

4. Data-oriented privacy protection techniques

Recall from Section 3 that two types of adversaries threatening data-oriented privacy are external and internal adversaries. To defend against external adversaries, cryptographic techniques such as encryption and authentication can be used effectively. In particular, a multi-level privacy protection scheme was proposed in [25] to assign different encryption keys to sensed data which require different levels of privacy protection, in order to properly defend against external adversaries.

Compared with external adversaries, the internal adversaries can be more powerful because they consist of nodes/base stations captured and controlled by malicious entities and therefore may have knowledge of encryption keys used in a WSN. As mentioned in Section 3, there are two types of data privacy we would like to protect against internal adversaries: the privacy of data being collected and the privacy of queries being posed to a WSN. For the first class privacy, a straight forward approach to defend against internal adversaries is to apply *end-to-end* encryption between the data source and the base station. With this approach, no intermediate node, including the internal adversaries, can compromise the privacy of data being

transmitted without knowing the key shared by only the two end nodes. Nonetheless, this approach also impedes the normal operation of a WSN. In particular, it renders data aggregation (i.e., node-by-node selection of transmitted data to reduce traffic volume) infeasible in intermediate nodes, due to their incapability of decrypting the transmitted data. This leads to a significant amount of additional traffic being transmitted, jeopardizing the energy consumption of a power-constrained WSN. Thus, a main challenge for privacy-preserving data aggregation is to defend against internal adversaries under hop-by-hop (i.e., link) encryption, where nodes that are directly communicating with each other share a private key for encryption and decryption.

For protecting the privacy of queries posed to a WSN, the base station must collect not only the data useful for answering a query, but also other data available in a WSN, in order to prevent an adversary from inferring the private query based on the data being accessed. Thus, a main challenge for protecting query privacy is to minimize the amount of dummy tuple while maintaining the privacy of queries being issued [4]. In the following, we will present the existing techniques for protecting the privacy of data aggregated and queries posed, respectively.

4.1. Privacy protection during data aggregation

Data aggregation is designed to substantially reduce the volume of traffic being transmitted in a WSN by fusing or compressing data in the intermediate sensor nodes (called *aggregators*). It is an important technique for preserving resources (e.g., energy consumption) in a WSN. Interestingly, it is also a common and effective method to *preserve* private data against an external adversary, because the process compresses large inputs to small outputs at the intermediate sensor nodes. However, as mentioned above, the usage of data aggregation also leads to vulnerability against internal adversaries. In particular, if an aggregator is compromised, it may perform either passive or active attacks:

- For passive attacks, the malicious aggregator properly follows the protocols defined by a WSN, with the only exception that it may record all intermediate computation and communication. Since an aggregator is supposed to perform the aggregation operations (e.g., max/min), it has the capability of decrypting the transmitted data and compromising the content privacy.
- Furthermore, an aggregator may launch active attacks to *inject* bogus data or *tamper with* raw data, leading to the invalidity of collected data at base station. Unlike the passive attacks which aim to compromise the confidentiality of private data, active attacks aim to destroy the integrity of collected data. Existing techniques defend against such active attacks by utilizing trust-based mechanisms to confirm the aggregated result [6,30,31].

In the following, we review the existing techniques for privacy-preserving data aggregation [14,32]. A common idea of existing techniques is to perturb the data being transmitted in a WSN. Two privacy-preserving techniques were proposed in [14]: (i) cluster-based private data aggregation (CPDA), which adds random seeds into the original data, and (ii) slice-mixed aggregation (SMART), which chops one data item into pieces and then rebuilds data package after exchanging those pieces randomly. Both approaches rely on the cooperation of neighbors to hide individual raw data, and they concentrate on the SUM aggregation operation. On the other hand, in [32], a collusion-resilient privacy-preserving data-aggregation technique is proposed, generalizing the transmitted data to support multiple aggregation operations, such as MIN/MAX, Median, etc. In the following, we review these three techniques in detail.

4.1.1. Cluster-based privacy data aggregation (CPDA)

The basic idea of CPDA is to introduce noise to the raw data sensed from a WSN, such that although an aggregator can obtain accurate aggregated information but not individual data points. This is similar to the data perturbation approach extensively used in privacy-preserving data mining. However, unlike in privacy-preserving data mining where noises are independently generated (at random) and therefore leads to imprecise aggregated results, the noises in CPDA are carefully designed to leverage the cooperation between different sensor nodes, such that the precise aggregated values can be obtained by the aggregator.

In particular, CPDA classifies sensor nodes into two categories: cluster heads and cluster members. There is a one-to-many mapping between the cluster heads and cluster members. The cluster heads are responsible for directly aggregating data from cluster members, with the communication secured by a different shared key between any pair of communicating nodes. Consider a cluster with one head and n members. Let the cluster head be denoted as s_0, and let all other sensors be denoted as $s_1,\ldots, s_n$, respectively. Let P_i be the private data values belonging to a sensor s_i. Each node is pre-assigned a non-zero number, A_i, which is known to all other members in the same cluster. Furthermore, each sensor s_i generates n private random numbers, $R_1^i,\ldots,R_n^i$ and calculates the following numbers:

$$V_j^i = P_i + R_1^i A_j + R_2^i (A_j)^2 + \ldots + R_n^i (A_j)^n \tag{1}$$

where $0 \leqslant j \leqslant n$. Then s_i will send V_j^i to s_j, which is encrypted by the shared key between $s_i$and s_j. After exchanging data, every sensor s_j calculates a value F_j as follows

$$F_j = \sum_{i=0}^{n} V_j^i = P + R_1 A_j + R_2 (A_j)^2 + \ldots + R_n (A_j)^n \tag{2}$$

where $P = \sum_{i=0}^{n} P_i$ and $R_k = \sum_{i=0}^{n} R_k^i$ for $1 \leqslant k \leqslant n$. Then s_j broadcasts F_j to the cluster head s_0 which calculates the sum P according to $U = G^{-1}F$ where G^{-1} is the inverse of the matrix

$$G = \begin{pmatrix} 1 & A_0 & \ldots & (A_0)^n \\ 1 & A_1 & \ldots & (A_1)^n \\ \ldots & \ldots & \ldots & \ldots \\ 1 & A_n & \ldots & (A_n)^n \end{pmatrix}. \tag{3}$$

Here $F = (F_0, F_1, F_2, \ldots F_n)^T$ and $U = (P, R_1, R_2, \ldots R_n)^T$ where $X^T$is the transpose of a matrix X. So P is the first

element of U. Note that the matrix G is of full rank because A_i are distinct numbers. One can see that the cluster heads now learn the sum of private data from their cluster members. Nonetheless, the cluster heads cannot compromise the privacy of their respective members separately.

The random data mentioned in CPDA could be regarded as noise which assists to hide individual raw data items from being known by the cluster head. Due to the cooperation of nodes in the cluster, the negative effect of noise is eliminated, and the sum calculated is exactly the same as its original value.

4.1.2. Slice-mixed aggregation (SMART)

SMART is another solution to protect individual data in the SUM aggregation. The main idea is to slice original data into pieces and recombine them randomly. There are three steps in this scheme. In the first step (Slicing), each sensor randomly selects J neighbor nodes within h hops to form a set S. Then it slices its data into J pieces, keeps one of those pieces for itself and then sends the other $(J-1)$ encrypted pieces to $(J-1)$ sensors randomly selected from the set S. In the second step (Mixing), after a sensor receives pieces of data from some other sensors, it decrypts the data using the key shared with the data sender. Each sensor waits for a while to make sure that all of the round aggregation data have already been sliced and received separately. In the third step (Aggregation), the intermediate sensor aggregates all pieces of data and transmits it towards the base station. Figs. 3–5 give a clear illustration of the basis idea of SMART. In this example, the five sensors are denoted as s_1 to s_5, respectively. Let d_{ii} be the piece of data kept by s_i, where let d_{ij} mean the piece of data transmitted from s_i to s_j and r_i means the data aggregated by s_i.

4.1.3. Generic privacy-preservation solutions for approximate aggregation (GP^2S)

The basic idea of GP^2S is to *generalize* the values of data transmitted in a WSN, such that although individual data content cannot be decrypted, the aggregator can still obtain an accurate estimate of the histogram of data distribution, and thereby approximate the aggregates. In particular, before transmission, each sensor node first uses an integer range to replace the raw data. Then, with certain granularity, the aggregator plots the histogram for data collected, and then estimates aggregates such as MIX/MAX and Median according to the algorithms proposed in [32].

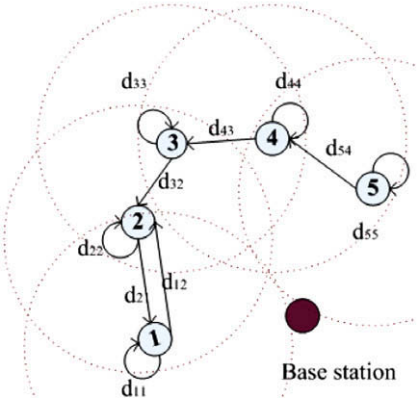

Fig. 3. Slicing $(J=2, h=1)$.

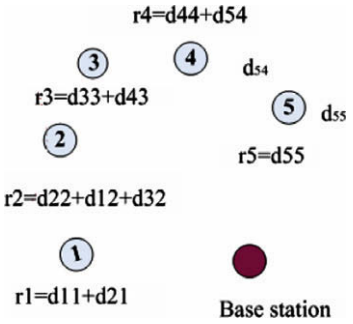

Fig. 4. Mixing: $(J=2, h=1)$.

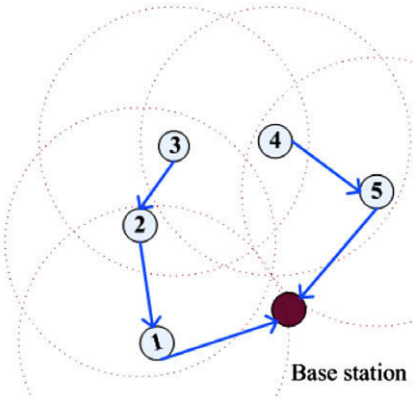

Fig. 5. Aggregating.

Similar to the generalization mechanism of GP^2S, a bucketing scheme was proposed in [26] to first bucketize the mixture of original data (into a few bins) before storing them at storage nodes compromisable by an adversary. Before transmitting the data to the storage nodes, the data source first encrypts the data using the key shared with the base station, and then attaches to the encrypted data a tag indicating the bucket which the data falls into. When the base station needs to process a range query over the data stored in the storage nodes, instead of storage nodes decrypting every data point and returning them, the base station first obtains an approximate result based on the tags corresponding to the query range. Then, through decrypting at the base station, real data set in need could be derived accurately.

4.2. Private date query

Orthogonal to the protection of private data being collected in a WSN, the query issued to a WSN (to retrieve the collected data) is often also of critical privacy concerns. In the previous example of medical WSN, if an adversary learns that queries have been frequently issued to the part

of WSN that covers a patient's house, the adversary can certainly infer the health of the patient is getting more attention probably due to his/her health problems. Private data query represents significant challenges to the design of a resource-constrained WSN. From the perspective of energy preservation, query processing should be restricted to as small a targeted range as possible. Nonetheless, narrowing down the range also increases the chance for the query to be inferred by adversaries.

To address this challenge, a target-region transformation technique was proposed in [4] to fuzzy the target region of the query according to pre-defined transformation functions. The main idea of the transformation function is to map one region into m regions, such that the target region cannot be distinguished from the other uninteresting regions. Multiple transformation functions such as uniform, randomized and hybrid were introduced in [4].

In the Union Transform (UT), the interesting region of each query is transformed to the set of all regions that appear in a query sequence. For instance, let (Q_1, Q_2, Q_3, Q_4) be a query sequence with target regions $(0,0)$, $(2,3)$, $(4,2)$ and $(1,1)$, respectively. No matter which of those four queries is carried on, the four target regions are queried simultaneously. Fig. 6 illustrates the mapping between query sequence and query regions under the UT scheme.

In the Randomized Transform (RT), each query is mapped into a randomized set of regions involving the target region. Considering the same example as above, Fig. 7 shows the association between queries and their query regions under this scheme. Regions with the same specific lines represent the query regions of the corresponding query. Here each query maps four regions involving only one real target. Finally, the hybrid Transform (HT) is a combination of UT and RT.

In comparison with the data query problem in the domain of databases, these schemes are similar to the k-anonymity algorithms (e.g., [27]) which hide the target's real identity using other $(k-1)$ similar objects so that it is impossible for the adversary to distinguish the target from the $(k-1)$ other objects. But the cost of such an anonymity-based technique is high in a WSN, because query dissemination and data collection in the uninteresting

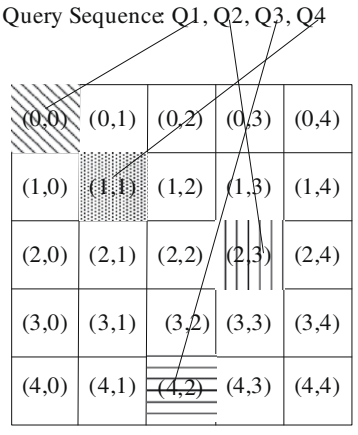

Fig. 6. Each query maps with all target.

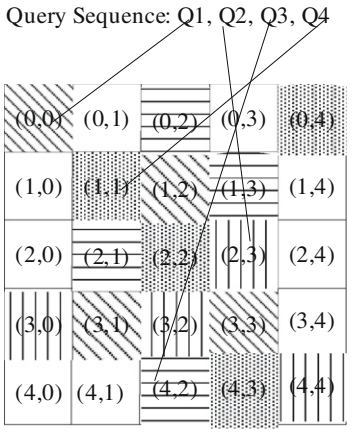

Fig. 7. Each query randomly maps m regions in the sequence ($m = 4$).

regions consume a large amount of energy. The larger the region saturated with query, the more effectively does safeguarding privacy work, but at the cost of the more energy. Consequently, a proper tradeoff between energy consumption and privacy protection is important.

Another query protection technique was proposed in [37]. Instead of enlarging the query regions as proposed in [4], this technique intends to disconnect the mapping between a user's identity and the query issued by the user. With this technique, a user is able to access data in a WSN after purchasing a certain amount of *tokens* from the WSN owner with blind signatures [5]. Such a token not only controls access to the sensed data, but also hides the user's identity and thereby guarantees query privacy.

5. Context-oriented privacy protection techniques

In this section, we review the existing techniques for context-oriented privacy protection that address the protection of private information related to the characteristics of traffic being saturated in a WSN. In particular, the existing techniques in the literature have focused on protecting two types of contextual information: location of special sensor nodes such as the data source and the base station, and timing of the generation of sensitive data.

5.1. Location privacy

A major challenge for context-oriented privacy protection is that an adversary may be able to compromise private information even without the ability of decrypting the transmitted data. In particular, since hop-by-hop transmission is required to address the limited transmission range of sensor nodes, an adversary may derive the locations of base station and data source by observing and analyzing the traffic patterns between different hops (to track down the base station and/or the data source). To address this challenge, the objective of context-oriented privacy protection is to hide the real traffic pattern. Various traffic-pattern-obfuscation techniques have been proposed in the literature.

In the following, we review the existing techniques for protecting the locations of data source and base station. A common idea of such techniques is to introduce random factors to packet routing, in order to increase the uncertainty of traffic patterns observed by the adversary and to counteract the adversarial traffic analysis attacks. Note that, after all, it is the routing path that exposes both types of private information. Thus, protecting such private information should start from hiding routing information against traffic analysis. With this common idea, the cost associated with privacy protection is the delay of delivering data packets, the (possible) reduction of successful delivery probability, or even a huge amount of energy consumption.

5.1.1. Location privacy of data source

Before reviewing the existing techniques for protecting the location privacy of the data source, let us first briefly discuss the "Panda Hunter Game", a classic formed problem, based on which (data-source) location privacy in WSN has been extensively studied in the literature [18]. In the Panda Hunter Game, a large number of *panda-detecting sensors* are deployed in a panda habitat. After sensors detect a panda, they will generate event messages and transmit them toward the base station. Meantime, a *panda-hunter* also attempts to identify the location of the data source to find the panda.

In this game, the objective of the defender is two-fold: On one hand, the defender needs to properly obtain the panda's moving information in order to enable biological research. On the other hand, it also intends to hide the location information from being known by the panda hunters which have the ability to eavesdrop the wireless communication between different sensor nodes, but do not have the key to decrypt the payload. The objective of the panda hunter is to compromise the location of the data source (and thereby the panda) by analyzing the traffic flow in the WSN.

Since a local adversary may only be able to monitor the traffic within a small local area, generally, there are two approaches for adversary to start the attack, arbitrarily choosing a place to stay in the network to monitor traffic or staying around base station with the prior knowledge about the location of the base station. In the following, we discuss four existing techniques, *flooding* [18], *Random walk* [18,28], *dummy injection* [18,24,29], and *fake data sources* [21], against the disclosure of the location of data source in WSNs.

5.1.1.1. Baseline and probabilistic flooding mechanisms. The basic idea of baseline flooding is for each sensor to broadcast the data it receives from one neighbor to all of its other neighbors. The premise of this approach is that all sensors participate in the data transmission so that it is unlikely for an attacker to track a path of transmission back to the data source [18]. However, the effectiveness of baseline flooding on privacy protection critically depends on the number of nodes on the transmission path between the data source and the base station. If the path is too short, after an adversary detects the arrival of the first packet at the base station, the adversary can infer that the routing path of this packet is the shortest path between the data source and the base station. Then, at a later time, the adversary can track back from the last forwarding sensor along the routing path to the data source. Furthermore, the flooding consumes significant amount of energy in the whole network and hence the lifetime of the WSN may be substantially reduced.

To address the side effect of baseline flooding, probabilistic flooding is proposed in [18], in which not all sensors are involved in forwarding data. Instead, each node forwards/broadcasts a packet it receives with a pre-determined probability. One can see that this scheme may not only significantly save energy but also effectively limit the adversary's ability to deterministically track back to the data source. Nonetheless, the reception of data by the base station is not guaranteed due to the randomness involved in this approach.

5.1.1.2. Random walk mechanisms. As one of random walk approaches, *Phantom Routing* is proposed in [18]. Figs. 8 and 9 illustrate the basic idea of Phantom Routing. Data

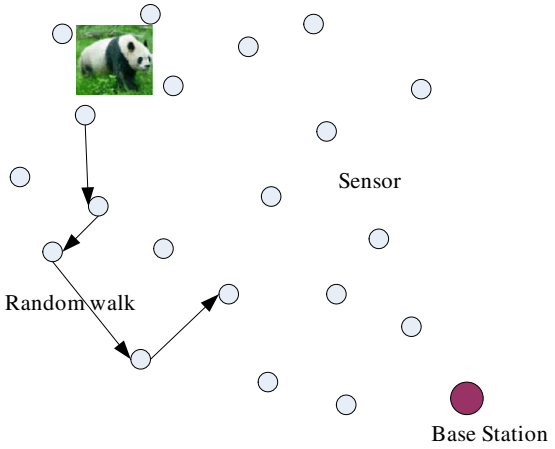

Fig. 8. Random walk, $h = 4$.

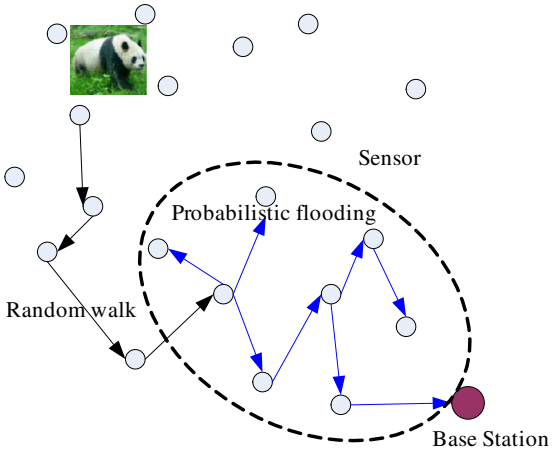

Fig. 9. Probabilistic flooding.

first performs a few steps of random walk from data source, and then, by employing probabilistic flooding scheme, it is transmitted towards base station. The premise of this approach is that even if an adversary is able to track back along the routing path, it would only be able to figure out the terminal node of the random walk instead of the original data source.

Unfortunately, as indicated in [28], the pure random walk approach is not statistically secure for protecting the location of the data source. In particular, it can be shown that a pure random walk tends to stay around the real source [23]. To improve Phantom flooding, a two-way *greedy random walk* (*GROW*) scheme was proposed in [28]. Figs. 10 and 11 give a vivid illustration of running GROW. In this scheme, a random path with a given number of hops is initiated from base station first. Sensors located on the path will serve as the *receptors*. And then, each packet from a source is then randomly forwarded until it reaches one receptor. At that point, the packet is forwarded to the base station through the path pre-established by base station.

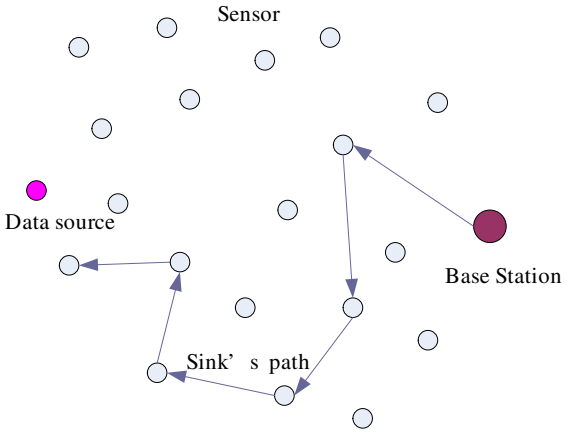

Fig. 10. Sink builds the receptor path.

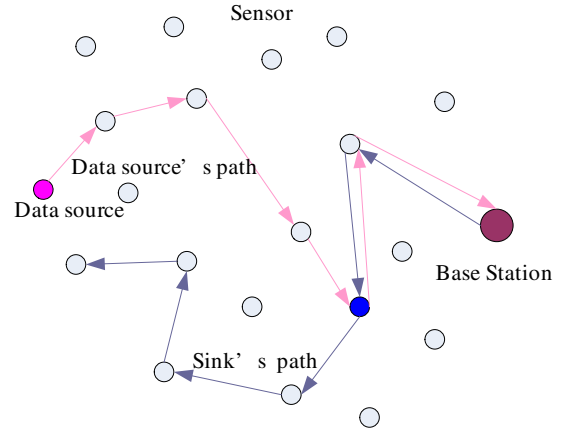

Fig. 11. Data randomly walk, intersect with a receptor and follow receptor path.

5.1.1.3. Dummy data mechanism. To further protect the location of the data source, fake data packets can be introduced to perturb the traffic patterns observed by the adversary. In particular, a simple scheme called *Short-lived Fake Source Routing* was proposed in [18] for each sensor to send out a fake packet with a pre-determined probability. Upon receiving a fake packet, a sensor node just discards it. Although this approach perturbs the local traffic pattern observed by an adversary, it also has limitations on privacy protection. Specifically, to maintain the energy-efficiency of the WSN, the length of each path along which fake data is forwarded is only one hop, therefore, an adversary is able to quickly identify fake paths and eliminate them from consideration.

Note that this technique is ineffective against global adversaries that can monitor the transmission rate of each sensor node and thereby identify those that are only sending out dummy data. To address this problem, one possible approach is to globally inject dummy data as well as keeping the transmission of real data the same as that of dummy data. However, this approach may introduce significant delay to the data transmission process. To cope with this concern, various techniques have been proposed in the literature. For example, a special distribution of data transmission interval was suggested in [24], such that as long as every sensor node follows such a pre-determined distribution, the delay of real data can be minimized without allowing an adversary to identify the real traffic. Two other schemes, Proxy-based Filtering Scheme (PFS) and Tree-based Filtering Scheme (TFS), were proposed in [29] to filter partial dummy data without threatening the source privacy. The whole network is divided into cells. The proxies are responsible for relaying real data from cells around them and filtering dummy data. After filtering all dummy packages from cells and buffering real data, the proxy will send data package, including buffered real data and new generated dummy data, at the same rate of transmission. Thanks to the filtering procedure, a large number of dummy data are removed so that energy consumption on overhead communication is greatly reduced.

5.1.1.4. Fake data sources mechanism. The basic idea of fake data source is to choose one or more sensor node to *simulate* the behavior of a real data source in order to confuse the adversaries [21]. The more the fake sources, the better can the identity of real source be protected. Such a technique will also incur more power consumption in the WSN. Furthermore, a significant challenge for the design of this technique is how to simulate the behavior of data sources without being detected. This is an open problem.

5.1.2. Location privacy of base station

Besides protecting the data-source location, another important challenge for context-oriented privacy is the proper hiding of the base station's location. In a WSN, a base station is not only in charge of collecting and analyzing data, but also used as the gateway connecting the WSN with outside wireless or wired network. Consequently, destroying or isolating the base station may lead to the malfunction of the entire network. In the following, we

review the existing privacy-preserving techniques [7–9,16] for the location of the base station. In particular, we discuss techniques for defense against local and global adversaries.

5.1.2.1. Defense against local adversaries. The main challenge for defending the location of base station against local adversaries is two-fold: First, the location information of the base station is usually included in the data payload being transmitted. Thus, one must provide the payload confidentiality to hide the base station's location. Second, a local adversary may be able to infer the parent–child relationship (i.e., which node within two communicating ones is nearer to the base station) from the time interval between receiving and sending data at each sensor. Such information can enable an adversary to efficiently trace down to the base station along the data transmission route. For the first challenge, traditional security techniques could work effectively. To handle the second challenge, the following four techniques have been proposed in the literature [8,9,16].

- *Changing data appearance by re-encryption:* Similar to anonymous routing systems (e.g., [10]) in traditional wired networks, a data re-encryption scheme was proposed in [9] such that a packet is re-encrypted hop-by-hop when transmitted through the routing path. This scheme eliminates the disclosure of base station location through changing the appearance of data.
- *Routing with multiple parents:* In [8,9], a multiple-parent scheme was introduced to balance the traffic load between parents and children, such that an adversary cannot easily identify which one is nearer to the base station. Also, this scheme allows each sensor to randomly select one of its multiple parents to transmit data towards the base station. This also makes it more difficult for the adversary to identify the location of the base station by tracing the data transmission.
- *Routing with random walk:* In [16], a simple random walk divides the neighbors of a sensor into two lists – closer and further lists – according to the hop count from the base station. When sensor forwards data, it randomly selects a next hop neighbor from one of those two lists. In the random walk schemes [8,9], each node selects its parent as next hop with probability P_r or randomly forwards data with probability $(1 - P_r)$, which drastically reduces the probability of successful analysis by the adversary. In this sense, the scheme in [16] could be regarded as a special case of the strategy in [8,9] for $P_r = 1/2$. However, the cost of random walk is the delay of delivering the data package to the base station as well as the extra energy consumption, since it is possible to select a node as next hop which is further away from sink. Or even, there is no guarantee to the arrival of data at the base station.
- *Decorrelating parent–child relationship by randomly selecting sending time* [9]: An adversary can figure out the relation of parent–child through the short time interval between sending data at a sensor and receiving data at its neighbor node. In order to de-correlate this specific relation, the period of time T is divided into m slots when there are one parent and $(m - 1)$ children for a sensor. The sensor assigns time slots to its children and each sensor will transmit its data at a random time within its slot.

The above privacy-preserving techniques aim to prevent an adversary from identifying the base station. There has also been work on how to improve the robustness of a WSN after the base station has been detected. In particular, a multiple-base station scheme was proposed in [7]. Even when one of the base stations is destroyed, the others can continue to collect data and maintain the normal functions of the WSN. This scheme apparently enhances the robustness of the system, but offers no additional protection for the base stations. Thus, it only delays, without eliminating, the threat from a local adversary on identifying the location of the base stations.

5.1.2.2. Defense against global adversaries. The above-mentioned techniques are ineffective against a global adversary which is capable of monitoring all traffic transmitted in the WSN. In particular, since an adversary is able to compute the transmission rate of each sensor node, it can easily identify the base station according to the specific topology of WSNs as discussed in Section 1. To defend against global adversaries, the injection of dummy traffic and/or dummy sensor nodes is required. We review the existing techniques in the following:

- *Hiding traffic pattern by controlling transmission rate.* The asymmetric data traffic flow in a WSN facilitates adversary to find the base station because a sensor close to the base station needs to not only send its own data but also relay data from sensors further away from the base station, and therefore features a high transmission rate. A privacy-preserving technique was proposed in [9] to maintain the same transmission rate among all sensors by controlling delay of real data.
- *Propagating dummy data.* In [8,9], under the assumption that an adversary cannot distinguish real data from fake data, a fake-packet injection scheme was proposed to prevent an adversary from identifying the real data transmission pattern. In particular, if a sensor detects that its neighbor is transmitting a real packet to the base station, the sensor generates a fake packet with probability P_c, and forwards it to one of its neighbors. In [8,9], the authors introduced two fractal propagation methods. One scheme is to control the probability P_c according to the data forwarding rate. The higher the rate is, the lower is P_c. The other scheme aims to simulate a high transmission-rate area to mislead the adversary to believe it as the base station. The side effect of fractal propagation is additional energy consumption, due to the fake data being transmitted over the network. Consequently, one has to make a proper tradeoff between energy consumption and privacy protection.

5.2. Temporal privacy problem

Let us now address the second type of context-based privacy, namely temporal privacy. Consider again the Panda Tracking application. When sensors detect the target,

they generate event messages and forward them to the base station. If an adversary can identify a specific time when an event message is generated, the adversary could use the information to track the target, or even predict the target's next move. Here the attacker's confidence depends on the assumption that the delay time (t) of data passing through each intermediate sensor along the routing path is the same. When the adversary eavesdrops the message near the base station, it can first obtain the arrival time z and then deduct from it the multiplication of average delay t and the hop count h, according to the equation $x = z - h^*t$. The main idea of the counteraction in [17] is to locally buffer the data for a random period of time at the intermediate sensors located along the routing path, such that an adversary cannot accurately estimate the original generation time of the message.

Apparently, the random delay leads to increasing costs of buffer space at the intermediate nodes. How to make a tradeoff between the protection of timing privacy and the efficiency of buffer space is of primary concern. For this purpose, a rate-controlled adaptive delaying scheme is proposed in [17] to adjust the delay distribution as a function of the incoming traffic rate and the available buffer space.

6. Comparisons

In the previous sections, we followed a depth-first approach to review the existing techniques for each category of problems respectively. In this section, we horizontally compare all privacy-preserving techniques that have been reviewed in this paper. Table 1 depicts a comprehensive comparison of the performances of privacy-preserving techniques in WSNs. We evaluate their performances in terms of four metrics: privacy, accuracy, delay time, and power consumption. *Privacy* refers to the degree of privacy protection provided by the reviewed techniques. The *accuracy* measure covers two perspectives: (i) the accuracy of the data obtained by the base station; and (ii) the availability of the (intended) data to the base station (i.e., whether the data can be delivered to the base station). The *delay time* includes both the computation and communication time of data transmission at the intermediate sensors. Finally, the *power consumption* measure focuses on the additional messages required for transmission (i.e., additional energy consumed) in the WSN.

One can see from Table 1 that the benefit of privacy protection usually comes at the cost of other metrics. For example, all data aggregation techniques shown in Table 1 (i.e., CPDA, SMART and GP^2S) can provide perfect protection for the privacy of data collected from individual sensors. Nonetheless, all of these techniques also have deficiencies on other measures. For CPDA and SMART, the cost is mainly on the power consumption due to the exchange of sliced data. On the other hand, for GP^2S, the cost comes from the accuracy of data being aggregated, as only approximate values of aggregate functions, e.g., SUM, MIN and MEDIAN, can be obtained. One can see that a main challenge for the future design of privacy-preserving data aggregation techniques is how to make a proper tradeoff between three metrics: privacy, power consumption, and accuracy.

Another observation from Table 1 concentrates on the context-oriented privacy-preserving techniques. In particular, with the three techniques for location privacy protection (i.e., *random walk routing*, *flooding* and *dummy injecting*), the main cost of privacy protection is on power consumption. These techniques have to transmit a large amount of additional traffic due to the flooding of real data or the injection of dummy data. Such overhead leads to significant power consumption. Thus, a main challenge for the future design of context-oriented privacy-preserving techniques is to minimize the communication overhead which causes the costly power consumption.

7. Open problems

While some work has been proposed in privacy protection in WSNs, there are still many open research issues in need of future research. Here we list some important ones.

- In the scenario of Panda Hunting, the existing work [18] only addresses systems with a single mobile target (panda). An interesting open problem would be to tackle the cases with multiple mobile targets.
- A significant challenge is to effectively protect the location of a mobile base station. Admittedly, by intuition, the mobility of base station is supposed to provide some protection to its location privacy against external attack, however, it has to update its location to all or part of nodes in the network so that they could forward data towards it, which implicitly creates more chance to pin point it through internal attack. Accordingly, how to protect the location privacy of a mobile base station as well as guaranteeing the system functionality is challenge of primary importance.
- In [9], re-encryption technique is proposed to change the appearance of data in order to eliminate the possibility of figuring out base station by tracking one data package in a dense network. But in a sparse WSN, it is still highly likely for adversary to infer the moving of one data package without the disturbing of dense data traffic, in spite of disguising the data package by re-encrypting it. Therefore, how to make re-encryption technique work effectively in a sparse network is still challenging.
- Considering the promising applications of WSNs to our real physical world, new types of contextual privacy information may come out, which are possibly combined with such potential research fields as social network. We are looking forward to the coming of them and their counteractions which gradually enhance the privacy preservation in WSNs.

8. Final remarks

In this paper, we have presented a state-of-the-art survey on privacy-preserving techniques in WSNs. We discussed the existing privacy-preserving techniques in two categories, which address data-oriented and

Table 1
An overview of solutions.

	Privacy	Accuracy	Delay time	Power consumption
CPDA	Depend on the security of shared key	Get the exact sum if no data is lost	Cummunicate data assisting with calculation	Exchange a large amount of communication data for calculating the aggregation result
SMART	100% protection for privacy	Get the exact sum if no data is lost	Slice and recombine data	Transmit and receive slices of data
GP^2S	100% protection for privacy	Approximately plot data histogram	No extra delay time except aggregating	No extra energy consumption on overhead and extra communication
K-Anonymity based	Depend on the parameter k	Get query results from one interesting cell and (k-l) other uninteresting cells	Query and collect data in uninteresting cells	Query and collect data in uninteresting cells
DP^2AC	Depend on the λ-bit random integer generated at base station	Get accurate query results with verified access right	Verify the token based on publish-key mechanism, also token-reuse detection	Consume energy on distributed token-reuse-detcction
Flooding to protect data source	For baseline flooding, easily figure out the shortest path between sink and data source For probabilistic flooding, depend on the preset pribabilily	For baseline flooding, guarantee data arrival For probabilistic flooding, not guarantee data arrival	For probabilistic flooding, not guatantee the data transmission along the shortest path	Flood data over the whole network
Random walk to protect data source	Distract attention to the real data source	For Phantom, 100% data arrival For GROW, data arrival depends on intersection of two random walk	For Phantom, depend on the hops of the walk For GROW, depend on the randomness of walk	Consume energy on random walking
Dummy injection to protect data source	Disturb the traffic pattern of the whole network	No influence on data arrival and accuracy	Delay real data to keep its transmission rate follow the same distribution as that of fake data	Inject fake data
Fake data sou ice	Distract attention from the real source	No influence on data arrival and accuracy	No extra delay	Generate fake data from fake sources
Hop by hop re-encryption	Re-encrypt data link to link	No influence on data arrival and accuracy	Spend time on encrypting and decrypting data at each intermediate node	No extra energy consumption on overhead and extra communication
Multiple-parent routing	Route data from multiple routing paths	No influence on data arrival and accuracy	No extra delay	No extra energy consumption on overhead and extra communication
Random walk routing to protect base station	Depend on the probability of choosing the parent as next hop	Not guarantee data arrival Depend on the probabilistic selection	Depend on the selection of next hop	Random walking
Random sending time	Make parent–child relationship ambiguous	No influence on data arrival and accuracy	Randomly select transmission time in each slot	Less energy consumption on synchronization for slot control
Transmission rate control	Hide global traffic pattern by making transmission rate the same	No influence on data arrival and accuracy	Buffer data to keep transmission rate the same ove the whole network	No extra energy consumption on overhead and extra communication
Dummy propagation to protect base station	Injecting fake data	No influence on data arrival and accuracy	No extra delay	Inject fake data
Random delay for temporal privacy	Destroy the deduction of approximate generation time of data	No influence on data arrival and accuracy	Randomly buffer data at intermediate sensors	No extra energy consumption on overhead and extra communication

context-oriented privacy concerns. The data-oriented techniques address the protection of private data transmitted in the WSN and that of sensitive queries executed. Context-oriented techniques, on the other hand, protect the location of the data source and the base station as well as the timing of the generation and transmission of sensitive data. Also, we attempted to compare the existing techniques in terms of such metrics as privacy, accuracy, delay and power consumption. Through comprehensive analysis of privacy issues in WSNs, ranging from problem definitions to the existing techniques, we depicted a complete picture of the state-of-the-art on privacy preservation in WSNs. Furthermore, based on the existing work, we listed a number of open issues which may intrigue the interest of researchers for future work. It is our hope that this paper sheds lights on a fruitful direction of future research on privacy preservation for WSNs.

Acknowledgement

The authors gratefully acknowledge the insightful comments of the anonymous reviewers which helped improve

the quality of the paper significantly. This work is partially supported by the AFOSR Grant A9550-08-1-0260, NSF Grants IIS-0326505, CNS-0721951, CNS-0852673, CCF-0852674, and IIS-0845644. The work of S.K. Das is also supported by (while serving at) the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] R. Agrawal, A. Evfimievski, R. Srikant, Information sharing across private databases, in: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, 2003, pp. 86–97.
[2] R. Agrawal, R. Srikan, Privacy-preserving data mining, in: Proceedings of the 2000 ACM SIGMOD on Management of Data, Dallas, TX USA, May 15–18, 2000, pp. 439–450.
[3] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a Survey, Computer Networks 38 (4) (2002) 393–422.
[4] B. Carbunar, Y. Yu, L. Shi, M. Pearce, V. Vasudevan, Query privacy in wireless sensor networks, in: 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2007 (SECON'07), 18–21 June 2007, pp. 203–212.
[5] D. Chaum, Blind signatures for untraceable payments, in: Advances in Cryptology – Crypto'82, Springer-Verlag (1983), 1982, pp. 199–203.
[6] E.D. Cristofaro, J. Bohli, D.Westho, FAIR: fuzzy-based aggregation providing in-network Resilience for real-time wireless sensor networks, to appear, in: Proceedings of the Second ACM Conference Wireless Network Security (Wisec), 2009.
[7] J. Deng, R. Han, S. Mishra, Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks, in: IEEE International Conference on Dependable Systems and Networks (DSN 2004), Florence, Italy, 28 June–1 July 2004, pp. 637-646.
[8] J. Deng, R. Han, S. Mishra, Countermeasures against traffic analysis attacks in wireless sensor networks, in: First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SecureComm) 2005, September 2005, pp. 113–126.
[9] J. Deng, R. Han, S. Mishra, Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks, Pervasive and Mobile Computing Elsevier 2 (2) (2006) 159–186.
[10] R. Dingledine, N. Mathewson, P. Syverson, Tor: the second-generation Onion router, in: Proceedings of the 13th USENIX Security Symposium, 2004.
[11] M.J. Freedman, K. Nissim, B. Pinkas, Efficient private matching and set intersection, in: Advances in Cryptography, Proceedings of Eurocrypt 2004, 2004, pp. 1–19.
[12] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, Tan, Kian-Lee, Private queries in location based services: anonymizers are not necessary, in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, Vancouver, Canada, 2008, pp. 121–132.
[13] M. Gruteser, D. Grunwald, Anonymous usage of location-based services through spacial and temporal cloaking, in: Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys), 2003, pp. 31–42.
[14] W.B. He, X. Liu, H. Nguyen, K. Nahrstedt, T. Abdelzaher, PDA: privacy-preserving data aggregation in wireless sensor networks, in: Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May 2007, pp. 2045–2053.
[15] Z. Huang, W. Du, B. Chen, Deriving private information from randomized data, in: Proceedings of 2005 ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2005), 2005, pp. 37–48.
[16] Y. Jian, S.G. Chen, Z. Zhang, L. Zhang, Protecting receiver-location privacy in wireless sensor networks, in: Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May 2007, pp. 1955–1963.
[17] P. Kamat, W.Y. Xu, W. Trappe, Y.Y. Zhang, Temporal privacy in wireless sensor networks, in: Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS 2007), June 2007, pp. 23–23.
[18] P. Kamat, Y.Y. Zhang, W. Trappe, C. Ozturk, Enhancing source-location privacy in sensor network routing, in: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS 2005), June 2005, pp. 599–608.
[19] K. LeFevre, D.J. DeWitt, R. Ramakrishnan, Workload-aware anonymization, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 277–286.
[20] A. Machanavajjhala, D. Kifer, J. Gehrke, M. Venkitasubramaniam, l-Diversity: privacy beyond k-anonymity, in: ACM Transactions on Knowledge Discovery from Data, 1 (2007).
[21] K. Mehta, D.G. Liu, M. Wright, Location privacy in sensor networks against a global eavesdropper, in: Proceedings of the IEEE International Conference on Network Protocols (ICNP 2007), October 2007, pp. 314–323.
[22] M.G. Reed, P.F. Syverson, D.M. Goldschlag, Anonymous connections and onion routing, IEEE Journal on Selected Areas in Communications, 16 (4) (1998) 482–494.
[23] S.M. Ross, Stochastic Processes, second ed., John Wiley & sons, Inc., 1996.
[24] M. Shao, Y. Yang, S. Zhu, G. Cao, Towards statistically strong source anonymity for sensor networks, in: Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May 2007, pp. 1298–1306.
[25] M. Shao, S. Zhu, W. Zhang, G. Cao, pDCS: security and privacy support for data-centric sensor networks, in: Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May 2007, pp. 1298–1306.
[26] B. Sheng, Qun Li, Verifiable privacy-preserving range query in two-tiered sensor networks in: Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM 2008), 13–18 August 2008, pp. 46–50.
[27] L. Sweeney, K-anonymity: a model for protecting privacy, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 2 (2) (2002) 557–570.
[28] Y. Xi, L. Schwiebert, W.S. Shi, Preserving source location privacy in monitoring-based wireless sensor networks, in: Proceedings of the 20th International Parallel and Distributed Processing Symposium (IPDPS 2006), April 2006.
[29] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, G. Cao, Towards event source unobservability with minimum network traffic in sensor networks, in: Proceedings of the first ACM Conference on Wireless Network Security (WiSec), 2008, pp.77–88.
[30] Y. Yang, X. Wang, S. Zhu, G. Cao, SDAP: a secure hop-by-hop data aggregation protocol for sensor networks, in: ACM Transactions on Information and System Security (TISSEC 2008), 11 (4) (2008).
[31] W. Zhang, Y. Liu, S.K. Das, P. De, Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach, Elsevier Pervasive and Mobile Computing 4 (5) (2008) 658–680.
[32] W.S. Zhang, C. Wang, T.M. Feng, GP^2S: generic privacy-preservation solutions for approximate aggregation of sensor data, concise contribution, in: Proceedings of the Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom), Hong Kong, P.R.C., March 17–21, 2008, pp.179–184.
[33] N. Zhang, S. Wang, W. Zhao, A new scheme on privacy preserving association rule mining, in: Proceedings of the 8th European conference on Principles and practice of knowledge discovery in databases, 2004, pp. 484–495.
[34] N. Zhang, S. Wang, W. Zhao, A new scheme on privacy-preserving data classification, in: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2005, pp. 374–383.
[35] N. Zhang, W. Zhao, Distributed Privacy Preserving Information Sharing, in: Proceedings of the 31st International Conference on Very Large Data Bases, 2005, pp. 889–900.
[36] N. Zhang, W. Zhao, Privacy-Preserving Data Mining Systems, IEEE Computer 40 (2007) 52–58.
[37] R. Zhang, Y. Zhang, K. Ren, DP^2AC: distributed privacy-preserving access control in sensor networks, to appear in: Proceedings of the 28th IEEE International Conference on Computer Communications (INFOCOM 2009), pp.1298–1306.
[38] N. Zhang, W. Zhao, Privacy protection against malicious adversaries in distributed information sharing systems, IEEE Transactions on Knowledge and Data Engineering, 20 (8) (2008).

Na Li received her B.S. degree in Computer Science & Technology from Nankai University, Tianjin, China, in 2005. From 2005–2007, She was a research assistant in Computer Network Information Center, Chinese Academy of Sciences, Beijing, China. Now she is a PhD student at Computer Science and Engineering Department in the University of Texas at Arlington. Her interests involve routing and scheduling techniques, QoS and managing strategies, security and privacy in Wireless Sensor Network and Ad Hoc Network.

Nan Zhang is an Assistant Professor of Computer Science at the George Washington University. He received the B.S. degree from Peking University in 2001 and the Ph.D. degree from Texas A&M University in 2006, both in computer science. His current research interests include security and privacy issues in databases, data mining, and computer networks, in particular privacy and anonymity in data collection, publishing, and sharing, privacy-preserving data mining, and wireless network security and privacy.

Sajal K. Das is a University Distinguished Scholar Professor of Computer Science and Engineering and the Founding Director of the Center for Research in Wireless Mobility and Networking (CReWMaN) at the University of Texas at Arlington (UTA). He is currently a Program Director in the Computer and Network Systems division at the US National Science Foundation. He is also E.T.S. Walton Professor of Science Foundation of Ireland; a Visiting Professor at the Indian Institute of Technology (IIT) at Kanpur and IIT Guwahati; an Honorary Professor of Fudan University in Shanghai and International Advisory Professor of Beijing Jiaotong University, China; and a Visiting Scientist at the Institute of Infocomm Research (I2R), Singapore. His current research interests include wireless sensor networks, mobile and pervasive computing, design and modeling of smart environments, pervasive security, smart health care, resource and mobility management in wireless networks, mobile grid computing, biological networking, applied graph theory and game theory. He has published over 400 papers and over 35 invited book chapters in these areas. He holds five US patents in wireless networks and mobile Internet, and coauthored the books"Smart Environments: Technology, Protocols, and Applications" (Wiley, 2005) and"Mobile Agents in Distributed Computing and Networking (Wiley, 2008). Dr. Das is a recipient of several Best Paper Awards in such conferences as EWSN'08, IEEE PerCom'06, and ACM MobiCom'99. He received 2009 IEEE Technical Achievement Award for pioneering contributions to wireless mobile and sensor networks. He is also a recipient of the IEEE Engineer of the Year Award (2007), UTA Academy of Distinguished Scholars Award (2006), University Award for Distinguished Record of Research (2005), College of Engineering Research Excellence Award (2003), and Outstanding Faculty Research Award in Computer Science (2001 and 2003). He is frequently invited as keynote speaker at international conferences and symposia. Dr. Das serves as the Founding Editor-in-Chief of Pervasive and Mobile Computing (PMC) journal, and Associate Editor of IEEE Transactions on Mobile Computing, ACM/Springer Wireless Networks, IEEE Transactions on Parallel and Distributed Systems, and Journal of Peer-to-Peer Networking. He is the founder of IEEE WoWMoM and co-founder of IEEE PerCom conference. He has served as General or Technical Program Chair as well as TPC member of numerous IEEE and ACM conferences.

Bhavani Thuraisingham joined The University of Texas at Dallas (UTD) in October 2004 as a Professor of Computer Science and Director of the Cyber Security Research Center in the Erik Jonsson School of Engineering and Computer Science. She is an elected Fellow of three professional organizations: the IEEE (Institute for Electrical and Electronics Engineers), the AAAS (American Association for the Advancement of Science) and the BCS (British Computer Society) for her work in data security. She received the IEEE Computer Society's prestigious 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management." She was quoted by Silicon India Magazine as one of the top seven technology innovators of South Asian Origin in the USA in 2002. Dr. Thuraisingham has published over 80 journal papers and is the author of 8 books. Her research focussed on applying information management technologies for data security and national security. She was educated in the United Kingdom both at the University of Bristol and the University of Wales.