Future Generation Computer Systems 37 (2014) 108–116

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Secure and energy-efficient data aggregation with malicious aggregator identification in wireless sensor networks[☆]

Hongjuan Li[a], Keqiu Li[a,*], Wenyu Qu[b], Ivan Stojmenovic[c]

[a] *School of Computer Science and Technology, Dalian University of Technology, Dalian, 116024, China*

[b] *School of Information Science and Technology, Dalian Maritime University, Dalian, 116026, China*

[c] *SITE, University of Ottawa, Ontario, K1N 6N5, Canada*

HIGHLIGHTS

- We propose a secure and energy-efficient data aggregation scheme.
- Our scheme can detect malicious aggregators with a constant per node communication overhead.
- Theoretical analysis and extensive simulations indicate that our scheme outperforms the state-of-art secure aggregation scheme.
- We provide some extension directions to refine our scheme.

ARTICLE INFO

Article history:
Received 16 April 2012
Received in revised form
11 November 2013
Accepted 3 December 2013
Available online 24 December 2013

Keywords:
Wireless sensor networks
Data aggregation
Security
Malicious aggregator identification

ABSTRACT

Data aggregation in wireless sensor networks is employed to reduce the communication overhead and prolong the network lifetime. However, an adversary may compromise some sensor nodes, and use them to forge false values as the aggregation result. Previous secure data aggregation schemes have tackled this problem from different angles. The goal of those algorithms is to ensure that the Base Station (BS) does not accept any forged aggregation results. But none of them have tried to detect the nodes that inject into the network bogus aggregation results. Moreover, most of them usually have a communication overhead that is (at best) logarithmic per node. In this paper, we propose a secure and energy-efficient data aggregation scheme that can detect the malicious nodes with a constant per node communication overhead. In our solution, all aggregation results are signed with the private keys of the aggregators so that they cannot be altered by others. Nodes on each link additionally use their pairwise shared key for secure communications. Each node receives the aggregation results from its parent (sent by the parent of its parent) and its siblings (via its parent node), and verifies the aggregation result of the parent node. Theoretical analysis on energy consumption and communication overhead accords with our comparison based simulation study over random data aggregation trees.

1. Introduction

Wireless sensor networks (WSNs) are becoming increasingly popular to provide solutions to many security-critical applications such as wildfire tracking, military surveillance, and homeland security [1]. In sensor networks, thousands of sensor nodes collectively monitor an area. As all the sensor nodes in an area usually detect common phenomena, there is high redundancy in the raw data. To save energy and prolong network lifetime, an efficient way is to aggregate the raw data before they are transmitted to the base station as the sensor nodes are resource limited and energy constrained.

Data aggregation [2–6] is an essential paradigm to eliminate data redundancy and reduce energy consumption. During a typical data aggregation process, sensor nodes are organized into a hierarchical tree rooted at the base station. The non-leaf nodes act as aggregators, fusing data collected from their child nodes and forwarding the aggregated results towards the BS. However, data aggregation is challengeable in some applications due to the fact that the sensor nodes are vulnerable to physical tampering, which may lead to the failure of data aggregation. The sensor nodes are often deployed in hostile and unattended environments, and are not made tamper-proof due to cost considerations. So they might be captured by an adversary, which may arbitrarily tamper with the data to achieve its own purpose. Therefore, an important issue in applying data aggregation is to avoid such tampering so that the base station can get the correct data aggregation result.

[☆] This work is supported by the National Natural Science Foundation of China under Grant nos. 90718030, 90818002, 60973155, and 60973117, and New Century Excellent Talents in University (NCET) of the Ministry of Education of China.

[*] Corresponding author.
E-mail address: keqiu@dlut.edu.cn (K. Li).

0167-739X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.future.2013.12.021

To meet this challenge, some work has been done [7–12] in the area of secure data aggregation. For example, Chan et al. [7] put forward a secure hierarchical in-network aggregation scheme that provides favorable and impressive security properties. This scheme can verify whether or not tampering has occurred on the path between a leaf and the root [7]. Nevertheless, it cannot pinpoint the exact node where the tampering has happened in the case of tampering. To the best of our knowledge, none of the existing work is able to identify the nodes that tamper with the intermediate aggregation results.

To overcome this deficiency, we present a secure and energy-efficient data aggregation scheme termed MAI [13] to effectively locate the malicious aggregators in wireless sensor networks. To accomplish malicious aggregator identification, MAI performs aggregation recalculation. Since data aggregation is executed on the path from a leaf node to the base station, each node can verify its parent's aggregation by recalculating the aggregation result according to the results obtained from its siblings. If an inconsistency occurs, the parent node is flagged as a malicious node; otherwise, it is a normal one. Another characteristic of the scheme is that the aggregation and verification can be executed interactively. A parent node executes data aggregation only after the verification on its child nodes is completed. If any child node is identified to be malicious, the aggregation stops. This can avoid unnecessary wrong data transmissions and further reduce the energy consumption. Moreover, the verification procedure is a localized one, which results in a low communication overhead.

The rest of the paper is organized as follows: In Section 2, we overview some related work on secure data aggregation. Section 3 describes our system model and the attack model. In Section 4, we give a detailed description on the proposed MAI. Theoretical analysis and discussions are also presented in this section to further explain our scheme. Section 6 reports the simulation results. Finally, we summarize our work and conclude the paper in Section 7.

2. Related work

Data aggregation has the benefit to achieve bandwidth and energy efficiency. There has been extensive research [14–16] on data aggregation in various application scenarios. These aggregation schemes have been designed without security in mind. However, wireless sensor networks are likely to be deployed in hostile environments such as the battlefield, where an adversary may compromise nodes and manipulate the data.

Secure data aggregation [17,18] is a hot research problem in some applications. Basically, there are two types of aggregation models, i.e., the single-aggregator model and the multiple-aggregator model.

The authors in [8,9] investigated secure data aggregation for the single-aggregator model. The secure information aggregation (SIA) protocol presented by Przydatek et al. [8] was the first one to propose the aggregate–commit–prove framework. In this model, the BS is the only aggregator. Du et al. [9] proposed a scheme using multiple witness nodes as additional aggregators to verify the integrity of the aggregated result. As for the single-aggregator model, the corresponding schemes do not provide per-hop aggregation.

The multiple-aggregator model employs more than one aggregator. Hu and Evans [12] presented a secure aggregation protocol that is resilient to single aggregator compromise. However, this protocol cannot deal with the situation where there exist two consecutive colluding compromised aggregator nodes in the tree. Yang et al. [10] proposed SDAP, which utilizes a novel probabilistic grouping technique to dynamically subdivide an aggregation tree into subtrees of similar sizes, each of which reports its aggregation result. Suspicious groups participate in an attestation process to prove the correctness of its group aggregation. Due to the statistical nature, SDAP may not be able to detect the attacks that slightly change the intermediate aggregation results.

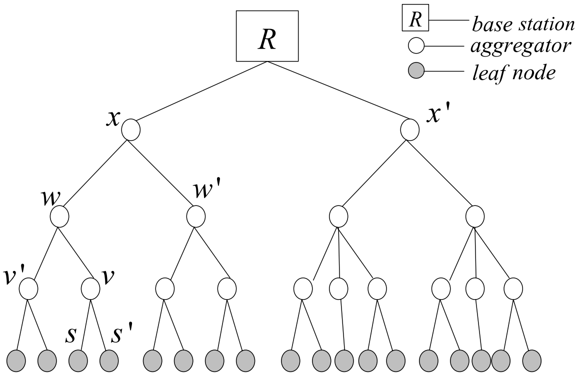

Fig. 1. An example aggregation tree.

In the privacy-preservation domain, Castelluccia et al. [19] proposed a new homomorphic encryption scheme in which the aggregation is carried out by aggregating the encrypted data at intermediate sensors without decrypting them, resulting in a higher level privacy. He et al. [20] proposed two privacy-preserving data aggregation schemes CPDA and SMART for additive aggregation functions.

3. Network model and attack model

In this section, we introduce the preliminary knowledge, including the network model and the attack model.

3.1. Network model

We model a wireless sensor network as a graph consisting of a set of n resource-limited sensor nodes $U = \{u_1, u_2, \dots, u_n\}$, each of which has an unique identifier ID_{u_i}. In addition, a resource-enhanced BS R is deployed to connect the sensor network to the outside infrastructure, e.g. the Internet. We assume that a topological tree rooted at R is constructed to perform the data aggregation. There are three types of nodes in the sensor network: leaf nodes, intermediate nodes, and the base station. The leaf nodes are collecting sensor readings. An intermediate node acts as an aggregator, aggregating the data transmitted from its child nodes and forwarding the aggregation result to its parent node. The base station is the node where the final result is aggregated. An example of such an aggregation tree is shown in Fig. 1. One method for constructing such an aggregation tree can be found in TAG [6].

Our scheme assumes that the network utilizes an identity-based public key crypto-system, which is also used in [21]. Each sensor node $u \in U$ is deployed with a private key, K_u^{-1}, and other nodes can calculate u's public key based on its ID, i.e., $K_u = f(\mathrm{ID}_u)$. Traditionally, it is assumed that public key systems exceed the memory and computational capacity of the sensor nodes. However, public key cryptography on new sensor hardware may not be as prohibitive as is traditionally assumed [21].

We further assume that the sensor nodes have the ability to perform symmetric-key encryption and decryption as well as to compute a collision-resistant cryptographic hash function. We also assume that there is a reliable transmission mechanism such that the packets transmitted in our scheme will not be lost.

3.2. Attack model

In this paper, we focus on defending against the attacks tampering with the intermediate aggregation results to make the