The Journal of Systems and Software 85 (2012) 152–166

Contents lists available at ScienceDirect

The Journal of Systems and Software

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

DyDAP: A dynamic data aggregation scheme for privacy aware wireless sensor networks

Sabrina Sicari[a], Luigi Alfredo Grieco[b], Gennaro Boggia[b], Alberto Coen-Porisini[a,*]

[a] Dipartimento di Informatica e Comunicazione, Università degli Studi dell'Insubria, via Mazzini, 5, 21100 Varese, Italy
[b] Politecnico di Bari, via Orabona, 4, 70125 Bari, Italy

ARTICLE INFO

Article history:
Received 20 September 2010
Received in revised form 27 July 2011
Accepted 27 July 2011
Available online 3 August 2011

Keywords:
Wireless sensor networks
Privacy
Anonymity
End-to-end data aggregation
Congestion control

ABSTRACT

End-to-end data aggregation, without degrading sensing accuracy, is a very relevant issue in wireless sensor networks (WSN) that can prevent network congestion to occur. Moreover, privacy management requires that anonymity and data integrity are preserved in such networks. Unfortunately, no integrated solutions have been proposed so far, able to tackle both issues in a unified and general environment. To bridge this gap, in this paper we present an approach for dynamic secure end-to-end data aggregation with privacy function, named DyDAP. It has been designed starting from a UML model that encompasses the most important building blocks of a privacy-aware WSN, including aggregation policies. Furthermore, it introduces an original aggregation algorithm that, using a discrete-time control loop, is able to dynamically handle in-network data fusion to reduce the communication load. The performance of the proposed scheme has been verified using computer simulations, showing that DyDAP avoids network congestion and therefore improves WSN estimation accuracy while, at the same time, guaranteeing anonymity and data integrity.

1. Introduction

Wireless sensor networks (WSN) technologies support data collection and distributed data processing by means of very small sensing devices (Akyildiz et al., 2007) with limited computation and energy capabilities. WSN are used in many contexts such as telemedicine, surveillance systems, assistance to disabled and elderly people, environmental monitoring, localization of services and users, industrial process control, and systems supporting traffic monitoring/control in urban/suburban areas, military and/or anti-terrorism operations.

Data transmission is one of the most power demanding tasks in WSN and therefore messages forwarding policies and processing techniques should be carefully designed to extend network lifetime. For example, a very common technique consists in keeping WSN nodes silent as long as no relevant information is detected in the monitored area (Dardari et al., 2007). Then, when some information is detected, wireless interfaces are turned on and sensors transmit the sensed data. However, the main drawback of such an approach is that it can cause network congestion (Akan and Akyildiz, 2005) and as a consequence some information may get lost.

To avoid network congestion one solution consists in using proper in-network algorithms (e.g., see Bagaa et al., 2007; Mahimkar and Rappaport, 2004; Castelluccia et al., 2005; Hu and Evans, 2003) that can reduce significantly the number of bytes exchanged across the WSN by aggregating data (Younis et al., 2006; Fasolo et al., 2007; Mastrocristino et al., 2010). In fact, in many situations, what is needed are aggregated measures, such as the average temperature of a region and the average humidity.

Another important issue in WSN is represented by privacy that may be violated by tampering of sensors and/or traffic due to the nature of the wireless channel and its deployment in uncontrolled environments. Thus, privacy aware mechanisms are crucial for several WSN applications such as localization and telemedicine. Among the different aspects characterizing privacy, anonymity is an important requirement for a privacy aware system that aims at protecting the identity of the individuals whose data are handled by the system. Moreover, it may be necessary to take into account privacy also in some application contexts in which data referring to individuals are not directly handled by WSN. For example, in home networks, sensor nodes may collect a large amount of data that may reveal habits of individuals, violating in this way their privacy.

However, the low power resources and the limited computational and storage capabilities of sensor nodes impose severe constraints on how privacy requirements can be satisfied.

* Corresponding author.
E-mail addresses: sabrina.sicari@uninsubria.it (S. Sicari), a.grieco@poliba.it (L.A. Grieco), g.boggia@poliba.it (G. Boggia), alberto.coenporisini@uninsubria.it (A. Coen-Porisini).

0164-1212/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2011.07.043

Considering together data aggregation and privacy issues is a very challenging problem, falling in the main research area of secure WSN (Grieco et al., 2009). Many solutions addressing at the same time aggregation and security aspects such as confidentiality, integrity, authentication, and availability can be found in literature (for an exhaustive and very comprehensive view of this topic see Ozdemir and Xiao, 2009). However, to the best of our knowledge, no solution is able to encompass privacy and end-to-end secure data aggregation.

This paper presents an approach that couples a privacy management policy with an original aggregation algorithm able to deal with end-to-end encrypted data. The approach, named DyDAP (Dynamic Data Aggregation with Privacy functions), is based on the privacy model proposed in previous works (Coen-Porisini et al., 2007, 2010a,b). Such a model, defined in UML (OMG, 2007a,b), represents a general schema that can be easily adopted in different contexts. DyDAP adapts such a schema to the context of WSN by introducing concepts, such as nodes, data, actions that are needed to define a privacy policy along with the existing relationships among them. Moreover, DyDAP integrates an original aggregation algorithm designed by exploiting linear discrete time control theory (Mastrocristino et al., 2010; Astrom and Wittenmark, 1995). Using DyDAP, each node periodically evaluates the amount of data to aggregate in order to control the level of its transmission queue. The aggregation process, which merges spatial correlated data working on encrypted information, involves only linear operations and allows the sink node to estimate the confidence level of the aggregated data.

The main goals fulfilled by our approach are: (i) anonymity management; (ii) data integrity check; (iii) data aggregation to reduce the network load; (iv) end-to-end secure data aggregation. The effectiveness of the approach has been studied using computer simulations. Notice that simulations are used to evaluate aspects such as congestion, data loss, number of transmitted messages, accuracy, while they are not useful to show that security/privacy goals are satisfied. In fact for these latter aspects we rely on the fact that the techniques used (such as encryption) guarantee "by construction" that such goals are met. In conclusion we show that DyDAP avoids network congestion and improves WSN estimation accuracy, while, at the same time, guaranteeing anonymity management and data integrity.

The rest of the paper is organized as follows. Section 2 introduces the foundations for modeling privacy in the context of WSN. Section 3 describes the reference scenario and the foundations on which DyDAP is based. Section 4 presents DyDAP in detail, while Section 5 reports the simulation results that show the effectiveness of DyDAP. Section 6 presents the most relevant related works. Finally, Section 7 draws some conclusions and provides hints for future works.

2. Modeling privacy policies for wireless sensor networks

A privacy policy defines the way in which data referring to individuals can be collected, processed, and diffused according to the rights that individuals are entitled to.

The rest of the paper adopts the terminology introduced by the EU directive (Directive, 1995). Notice that the terminology is as much technology-independent as possible and therefore, beside the original definition, we provide the needed refinements in order to support the definition of privacy mechanisms in WSN communications:

- *Personal data* means any information related to an identified or identifiable natural person (referred to as *data subject* or *subject*). In the context of WSN, they represent the data sensed by the nodes of the network; in other words, nodes play the role of *subjects* since they receive information from the environment in which they are located.
- *Processing of personal data* (*processing*) means any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction. In the context of WSN, activities such as sensing data, receiving/transmitting messages, aggregating data, encrypting/decrypting data and verifying data integrity can be considered as *processing*.
- *Controller* means the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data. In a WSN, nodes play the role of controllers of the network, since they verify the processing actions that involve sensed data.
- *Processor* means a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller. In a WSN, the role of processor is played by the nodes of the network.
- Data *subject consent* (*consent*) means any freely given specific and informed indication of his/her wishes by which the data subject signifies his/her agreement to personal data relating to him/her being processed.

As a distinctive feature of a privacy policy, the processor is required to state for what *purpose* data are processed. A purpose can be defined either as a high-level activity (e.g., "monitoring", "tracking") or as a set of actions (e.g., "determine the average temperature", "evaluate the humidity").

In addition processing actions may be executed under specific *obligations*. An obligation is a set of actions that the processor and/or the controller guarantees to carry out at the end of the processing activities. For example, a node that measures the temperature of the ground may be required to send an alert message whenever the temperature is less than a given threshold.

In a privacy aware system, subjects have to grant their consent before any processing can occur on their data. In the context of WSN we assume that the consent is implicitly given by the nodes of the network. This means that every node accepts that its data may undergo different processing activities each of which consisting in a set of processing actions (and obligations) that may occur on the other nodes of the network. Notice that the above assumption requires that the system modeler adopts adequate mechanisms to assure that nodes can trust each other.

2.1. The UML model

Starting from the general definitions, a conceptual model of privacy has been provided, using UML, in previous works (Coen-Porisini et al., 2007, 2010a,b). In the following we provide a short overview of such a conceptual model along with the refinements needed to take into account the specificities of WSN. The structural aspects are defined using UML classes and their relationships such as associations, dependencies, and generalizations. Fig. 1 depicts a class diagram that provides a high level view of the basic structural elements of the model, while Fig. 2 shows a refined class diagram in which all the needed concepts are introduced.

A *WSN_Privacy Policy* consists of three types of classes: *Node*, *Data*, and *Action*. Thus, an instance of class *WSN_PrivacyPolicy* is characterized by specific instances of *Node*, *Data*, and *Action*, and by the relationships among such entities. In what follows we focus on such basic classes.