Ad Hoc Networks 16 (2014) 28–45

Contents lists available at ScienceDirect

Ad Hoc Networks

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

A practical approach for provenance transmission in wireless sensor networks

S.M. Iftekharul Alam [a,*], Sonia Fahmy [b]

[a] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
[b] Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 13 April 2013
Received in revised form 16 October 2013
Accepted 3 December 2013
Available online 12 December 2013

Keywords:
Provenance
Trust framework
Energy-efficiency
Sensor networks

ABSTRACT

Assessing the trustworthiness of sensor data and transmitters of this data is critical for quality assurance. Trust evaluation frameworks utilize *data provenance* along with the sensed data values to compute the trustworthiness of each data item. However, in a sizeable multi-hop sensor network, provenance information requires a large and variable number of bits in each packet, resulting in high energy dissipation due to the extended period of radio communication. In this paper, we design energy-efficient provenance encoding and construction schemes, which we refer to as Probabilistic Provenance Flow (PPF). Our work demonstrates the feasibility of adapting the Probabilistic Packet Marking (PPM) technique in IP traceback to wireless sensor networks. We design two bit-efficient provenance encoding schemes along with a complementary vanilla scheme. Depending on the network size and bit budget, we select the best method based on mathematical approximations and numerical analysis. We integrate PPF with provenance-based trust frameworks and investigate the trade-off between trustworthiness of data items and transmission overhead. We conduct TOSSIM simulations with realistic wireless links, and perform testbed experiments on 15–20 TelosB motes to demonstrate the effectiveness of PPF. Our results show that the encoding schemes of PPF have identical performance with a low bit budget (∼32-bit), requiring 33% fewer packets and 30% less energy than PPM variants to construct provenance. With a twofold increase in bit budget, PPF with the selected encoding scheme reduces energy consumption by 46–60%.

1. Introduction

New micro-sensors have enabled wireless sensor networks (WSNs) to gather real-time data from the physical world [1,2]. Planet-wide sensor networks [3,4], sensor networks for large-scale urban environments [5], and physical infrastructure systems [6] indicate potential deployments of multi-hop networks consisting of hundreds of sensor nodes. In such networks, data produced by the sensors are collected at the base station and made available to decision makers for further analysis. As the quality of decision making is critically dependent on the quality of transmitted information [5], trustworthiness of information and information-transmitting nodes is important [7]. In a multi-hop network, *provenance* includes knowledge of the originator and processing path of data since its generation. While a few provenance-based trust evaluation frameworks have been proposed [8,9], they do not consider *energy dissipation* due to provenance transmission.

Provenance of a data item can be represented by a tree that is embedded as meta-data with the item, and updated along the path used to forward the item to the base station [9]. In this case, every intermediate node carries provenance of length proportional to the hop count between that node and the originator of the data item.

[*] Corresponding author. Tel.: +1 765 337 6447.
E-mail addresses: alams@purdue.edu (S.M.Iftekharul Alam), fahmy@cs.purdue.edu (S. Fahmy)alams@purdue.edu (S.M.I. Alam), fahmy@cs.purdue.edu (S. Fahmy).

1570-8705/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.adhoc.2013.12.001

In a network with a large diameter (hop count), this increased meta-data length results in an extended period of radio communication and energy dissipation at every intermediate node. We consider a real deployment of a 46-hop network [10] in our simulations, and observe that aggregated energy dissipation of the network increases by 27% when a traditional trust framework is employed. Although large networks can be hierarchically organized [11], they still require a significant number of hops [12], with non-negligible energy usage for provenance transmission.

Provenance encoding and construction is similar in nature to the well-known *IP traceback* problem [13,14]. IP traceback aims to determine the forwarding paths of spoofed packets in the Internet. Among the many proposed solutions to this problem, Probabilistic Packet Marking (PPM) can most easily be adapted to WSNs [15]. We have shown that direct application of PPM to WSNs is infeasible since it embeds a single node identifier in each packet, and hence requires a large number of packets to construct the forwarding path [16]. Instead, we propose a new approach, Probabilistic Provenance Flow (PPF), where a connected subgraph of the forwarding path is probabilistically embedded into a packet. PPF includes three new bit-efficient provenance encoding schemes that quickly construct provenance of an arbitrarily large multi-hop network.

We integrate a simple but robust scheme into PPF to handle topological changes. Since encoded provenance is matched against previously constructed provenance graphs at the base station, we can reduce decoding errors to negligible levels and speed up convergence. We also integrate PPF with provenance-based trust frameworks and explore how trust scores evolve faster with data items having dissimilar provenance than with the items having shared provenance. This study exposes the trade-off between trustworthiness or provenance dissimilarity of data items and transmission overhead: making provenance more dissimilar increases transmission overhead. We investigate this trade-off and propose a solution to provide decision makers with a tunable parameter to control the extent of provenance dissimilarity and transmission overhead.

We perform extensive simulations using TOSSIM to demonstrate the performance of PPF in a highly dynamic and asymmetric network. We further evaluate PPF using a testbed consisting of 15–20 TelosB motes in different settings. Our simulation and testbed results show that PPF with the selected encoding scheme can consume up to 46–60% less energy and converge with 45% fewer packets than the traditional approach, which significantly increases the network life-time.

The remainder of this paper is organized as follows. We formulate the problem of energy-efficient provenance transmission in Section 2. Section 3 discusses related work. Section 4 explains three different encoding schemes for PPF. We discuss the corresponding approaches to decode and construct provenance in Section 5. Section 6 discusses integration of PPF with provenance-based trust frameworks. In Section 7, we examine the parameter selection for one of the encoding methods. We analyze the bit requirements for embedding provenance using all encoding schemes in Section 8. Sections 9 and 10 present simulation and testbed results, respectively. Finally, Section 11 concludes the paper.

2. Problem formulation

2.1. Network model

We consider a multi-hop WSN where changes in topology due to failure or mobility can occur, but are infrequent. We make the following assumptions regarding the network and traffic:

(1) A Base Station (BS) acts as a central command authority and the root of a routing tree. It has no resource constraints and cannot be compromised by an attacker.
(2) Sensor nodes monitor their surroundings and periodically report to the base station or their designated cluster head (if any).
(3) Multiple sensors are used to monitor an event. Within a particular time window, independent observations obtained at cluster heads (if any) or the base station from different sensors are concerned with the same event.
(4) A provenance-based trust management method such as [8,9] is used in the application layer to evaluate and manage trust in an adaptive manner. More details can be found in [17].

2.2. Problem statement

We consider a network of N nodes, where the maximum length (depth) of any forwarding path (tree) is L. Assume that the maximum number of bits that can be used to embed provenance information in a single packet is B. Based on this bit budget, there is an integer $m, 1 < m \leqslant L$ such that at most m consecutive node identities (that is, $m - 1$ consecutive edges) can be embedded into a single packet. We must perform the following operations:

(1) **Provenance embedding**: In a forwarding tree $G = (V, E)$ rooted at the base station, each node $n_i \in V$ makes an independent decision whether to embed its identity into the packet, starting a connected sub-graph, with probability p_i. We need to design a provenance embedding method to carry a partial path $P =< n_{i_1}, n_{i_2}, \ldots n_{i_m} >$ into a single packet where $n_{i_j} \in V, 1 \leqslant j \leqslant m$ and $(n_{i_k}, n_{i_{k+1}}) \in E, 1 \leqslant k \leqslant m - 1$. This problem is a simple extension of the edge sampling approach in IP traceback [13].
(2) **Provenance construction**: At the base station, we must construct the entire provenance tree $G = (V, E)$ by exploiting partial path information collected from a number of received packets, with an upper bound on the number of packets required to construct the provenance.
(3) **Provenance evolution**: After topological changes, e.g., due to failures or mobility, we must bound the time that it takes to reflect the changes in the constructed provenance.