Computer Networks 146 (2018) 125–137

Contents lists available at ScienceDirect

Computer Networks

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

Towards optimal source location privacy-aware TDMA schedules in wireless sensor networks

Jack Kirton*, Matthew Bradbury, Arshad Jhumka

Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom

ARTICLE INFO

Article history:
Received 27 March 2018
Revised 23 July 2018
Accepted 10 September 2018
Available online 15 September 2018

Keywords:
Genetic algorithm
Wireless sensor networks
TDMA
Data aggregation schedule
Source location privacy

ABSTRACT

Source Location Privacy (SLP) is becoming important for wireless sensor networks where the source of messages is kept hidden from an attacker. In this paper, we conjecture that similar traffic perturbation to altering the routing protocol can be achieved at the link layer through assignment of time slots to nodes. This paper presents a multi-objective optimisation problem where SLP, schedule latency and final attacker distance are criteria. We employ genetic algorithms to generate Pareto-optimal schedules using two fitness criteria, examining the Pareto efficiency of selecting either and confirming the efficiency by performing simulations which show a near optimal capture ratio.

1. Introduction

Wireless sensor networks (WSNs) have enabled novel classes of applications such as monitoring and tracking. Asset monitoring is a task performed by some WSNs, where some node(s) detects the presence of an asset and transmits data about the asset back through the network to a base station node known as the sink. Even though sensitive data may be encrypted, the process of sending the message back to the sink, called *convergecast*, implicitly reveals the location of such an asset, as a potential attacker can trace back over the network traffic to the source of traffic, to ultimately capture the asset. *Source location privacy* (SLP) is the property of keeping the source's location within the network secret so as to prevent the capture of the asset.

The idea motivating this area of research was originally developed in [1,2] as the panda hunter game. The panda hunter game is based upon the premise of using WSNs to monitor the population of endangered animals (in this case pandas) over a large area of their natural habitat, such as in [3,4]. One of the typical problems facing endangered animals are poachers. While the data being transmitted through the network is encrypted (providing *content*-based privacy), the *context* in which the data is broadcast must also be protected. Context can be defined as a collection of attributes derived from the environmental and temporal situation in which the data was broadcast. This means that typical content-based privacy solutions (such as encryption) are not sufficient to solve context-based privacy issues. Hence, there is a need for context-specific privacy solutions.

The majority of existing work on SLP is focused on providing a solution at the *routing* layer of the network stack. In these works, the primary objective is to perturb the original (convergecast) routing protocol such that the resulting protocol (i) still routes data to the sink and (ii) the attacker cannot capture the source while backtracking on the network traffic. Based on the notion that traffic can be engineered to provide for SLP, we focus on achieving a similar objective while focusing at the link layer. Using the link layer as opposed to the routing layer has the advantage of typically sending less control messages [5] and as such reducing network energy consumption. Specifically, in WSNs, TDMA-based MAC protocols are often used in cases where timeliness is required. A TDMA-based MAC protocol works by splitting the timeline into slots and then allocating slots to nodes in such a way that message transmissions do not result in collision. Thus, it becomes possible to impose a given traffic pattern on the network based on slot allocation, providing a basis for traffic engineering at the link layer level. Specifically, each valid slot assignment for the network will provide a different pattern of traffic during operation. The principle then is that the slot assignment can be performed in such a way as to provide SLP within a class of convergecast protocol known as *data aggregation scheduling* (DAS). DAS works by constructing an aggregation tree rooted at the sink and the slot allocated to a parent is strictly greater than the slot of any of its children. Thus, a parent node will propagate aggregated information after collecting messages from all of its children.

* Corresponding author.
E-mail addresses: J.D.Kirton@warwick.ac.uk (J. Kirton), M.Bradbury@warwick.ac.uk (M. Bradbury), H.A.Jhumka@warwick.ac.uk (A. Jhumka).

https://doi.org/10.1016/j.comnet.2018.09.010
1389-1286/

Developing an optimal and valid TDMA schedule is known to be NP-complete [6]. However, we propose to add SLP as another optimization criterion in the design of optimal TDMA schedules. Our aim is thus to generate SLP-aware slot assignments utilising an evolutionary method in order to produce schedules that are valid for the network and also provide SLP.

We thus make the following contributions:

1. We map the SLP problem onto a GA problem.
2. We present suitable *crossover*, *mutation* and *selection* operators to expedite the generation of optimised schedules.
3. We use the notion of Pareto optimality to compare the various generated schedules and we use two different *fitness functions* for analysis.
4. We perform simulations in both ideal and realistic environments, showing metrics about the generated solutions such as near optimal capture ratio and high packet delivery ratio.
5. We examine those solutions that lie on the Pareto frontier and determine the optimal solution of those generated for a specific network configuration.

The remainder of the paper is organised into eight sections. Section 2 explores other works performed in this area. Section 3 provides models used during this work. Section 4 outlines the specification for DAS. The genetic algorithm and operators are detailed in Section 5. Section 6 defines Pareto efficiency and explains its utility. Section 7 explains the experiments that shall be run and the results are analysed in Section 8. Finally, Section 9 summarises our contributions.

2. Related work

2.1. Source location privacy

Phantom routing was first introduced in [1,2], alongside the panda hunter game. Phantom routing is a two stage protocol, firstly transmitting the message from the source along a directed random walk to a phantom node and secondly using the routing protocol to continue the transmission to the sink. Two variants were proposed for altering the routing protocol in stage two; PRS [2] used flooding and PSRS [1] used single-path routing. There has been much work on providing SLP since [7].

Since the introduction of the phantom routing concept, more work has been created to improve the first stage of the protocol (i.e. the directed random walk). Two solutions that attempt to prevent the random walk from turning back on itself are GROW [8], which stores previously visited nodes in a bloom filter, and [9], which uses location angles. While some phantom routing work has focused on improving the selection of nodes that take part in the random walk [10], others have used delay to prevent the attacker making positive moves towards a source [11].

Several issues with Phantom Routing have been investigated. One such issue is the performance degradation associated with the use of multiple sources [12]. Another issue is that traffic-analysis of phantom routes that collide with the network boundary can allow prediction of the source's location [13].

An alternative technique utilises fake sources to misdirect the attacker. A fake source sends encrypted and padded messages that appear identical to those produced by a real source. The idea is to attract the attacker to one of the fake sources rather than the real one. A fake source technique was proposed in [1] but was deemed to have a poor performance and said that it was not worth investigating this class of solutions further. This was contested by Jhumka et al. [14], which implemented an alternative technique that provides high levels of SLP. This class of solution was further expanded in [15,16] to make it applicable to a wider variety of areas and to determine parameters online. A drawback to utilising fake sources is that they often demand considerably higher energy usage than routing-based techniques.

Solutions exist that combine the use of both phantom routing and fake sources. Long et al. [17] generates a routing tree where leaf nodes would be fake sources that send messages up the tree. A further example is fog routing [18] where the network is split into *fogs*, which creates routing loops that trap the attacker indefinitely. Within the fogs, fake sources are also used. PEM [19] generates fake sources that perform a walk about during execution.

The solutions presented thus far focused on the local (distributed) eavesdropper [20] where, at any point in time, the attacker only gathers information about its current neighbourhood and then moves to gain further information. A global attacker is a stronger attacker with a view of the entire sensor network. The attacker could either operate their own sensor network deployed over the target network [21] or use long range radios to eavesdrop on all traffic. Two solutions that provide perfect privacy against global attackers are Periodic Collection [22], where every node broadcasts periodically, and ProbRate/FitProbRate [23], where nodes broadcast periodically but the rate at which they do so is based on an exponential distribution.

Cross-layer techniques are those that employ more than one layer of the network stack to provide SLP. Typically, only the routing layer is used, as is the case for phantom routing, where a cross-layer solution would additionally employ another layer, typically the MAC/link layer. Cross-layer techniques are far less common than the others. In [24] beacon frames at the MAC layer are modified to carry data to aid in moving messages away from the source, in a similar fashion to phantom routing. After the message has been propagated far enough, a standard routing method is used to deliver the message to the sink. A second method was proposed whereby the message would first be routed to a *pivot* node on the first round of beaconing, sending the message further away on the second round before finally flooding to the sink. Another technique [5] employs SLP for TDMA DAS networks, where slot allocations are altered at the MAC protocol level in order to attract the attacker along a diversionary route.

Our solution can be considered a hybrid between the techniques that provide privacy against a global attacker and local attacker. The solution will have all nodes periodically broadcasting, but the pattern of broadcasts is done in such a way that a route is created for a local attacker to follow.

2.2. Genetic algorithms

Genetic algorithms have been used for a wide variety of purposes in the sensor networks field. They have been used to find optimal parameters for routing protocols, such as LEACH [25], where it was used to determine weightings for combining multiple heuristics to determine which node became the next cluster head. The goal of this was to increase network lifetime by balancing energy loss between nodes more effectively. In [26], they went so far as to produce a new routing protocol created by a GA that is comparable to LEACH in order to use the least energy possible for those networks that harvest energy from the environment rather than batteries.

Genetic algorithms have also been used in the deployment of WSNs. The maximum coverage sensor deployment problem (MCSDP) is the problem of finding the minimum number of nodes required to cover a certain area [27]. Additionally, further work has been performed such that the deployment is augmented to find a solution that maximises the network lifetime by reducing energy requirements [28].

The allocation of TDMA time slots using genetic algorithm-related methods has previously been investigated [6,29,30] and finding an optimal TDMA schedule has been shown to be NP-