Source Location Anonymity for Sensor Networks

Ali Abbasi
ECE Department, University of Tehran
a.abbasi@ece.ut.ac.ir

Ahmad Khonsari
ECE Department, University of Tehran
School of Computer Science, IPM, Iran
ak@ipm.ir

Mohammad S. Talebi
School of Computer Science, IPM, Iran
mstalebi@ipm.ir

***Abstract*—Motivated by applications like sensor, peer to peer networks there has been growing interest in monitoring large scale distributed systems. In these applications, *source location anonymity* is an attractive and critical security property. Most of prior works assumed a weak adversary model where the adversary sees only local network traffic, but here we consider source anonymity against a global eavesdropper. Attaining location unobservability under global attacker is very difficult and expensive to achieve, because sensor networks are very limited in resources. In this work we propose a distributed algorithm to mix real event traffic with carefully chosen dummy traffic to hide the real event traffic pattern. We assume that we have fixed amount of resources to send dummy traffic and we try to share it among sensors so as to maximize the *degree of anonymity* of the system. Through simulation, we illustrate that the proposed technique is efficient in protecting location information from the eavesdropper.**

***Index Terms*—security and privacy, source anonymity, sensor networks, optimization**

I. Introduction

Wireless sensor networks are envisioned to consist of thousands of devices, each capable of some limited computation, communication, and sensing abilities, operating in an unattended mode. These networks are an interesting architecture for inexpensive and scalable instrumentation of our physical environment, allowing continuous and real-time monitoring of large geographical areas[7].

Prototypes of sensor networks have been deployed for applications like habitat monitoring [9] and military surveillance and target tracking[2]. In these applications, when an event occurs in the field that sensor network covers, e.g. detecting an animal, sensors will report their observations on the event back to the base station. However, sensor networks are opposed with many security threats such as node compromise, routing disruption and false data injection.

Among all these threats, privacy is of special interest to us since it cannot be fully addressed by traditional security mechanisms, such as encryption and authentication. This is because the communication patterns of sensors can, by themselves, expose a great deal of contextual information. For example, delivering sensor data to the base station may disclose the locations of some critical events in the field, revealing valuable information [10].

Preserving source anonymity is a challenging problem because only inexpensive security mechanisms can be affordable. Sensors usually use low cost radio devices that employ standardized wireless communication technologies which permit an attacker to eavesdrop communications between sensors. An Attacker can monitor the communications of the network either by scattering his sensors or by using a powerful site whose coverage is not less than the network area.

Source location anonymity has attracted attention in recent years [8], [15] and [17]. But the proposals either have considered the problem under weak adversary model or concentrated on the behavior of a single node. For instance, in phantom routing [8], an attacker has coverage limited to one sensor, but if the attacker becomes more powerful, e.g. has a hearing range more than three times that of the sensors, the capture likelihood is as high as 97% [15], [17].

In this paper we aim at addressing source anonymity under global eavesdropper who have access to the whole network traffic. If the network traffic only contains real events, attaining source anonymity is very hard if not impossible. The idea to hide real event messages is to introduce carefully chosen dummy traffic so as to conceal real traffic fluctuations over the whole network. [15] proposes a probabilistic schema to include dummy traffic in which the total traffic generated by a node follows a probabilistic distribution all the time. However, it is a reasonable assumption that nodes with higher average traffic include more portion of real traffic which give the attacker the chance to find event sources by checking high traffic cells of network.

Our main idea is to maximize the uniformity of the traffic over the whole network as much as possible. We assume there is a fixed amount of resource in terms of communication capacity and we wish to assign it to sensors in a way to minimize the probability that an eavesdropper could detect a real event location.

The rest of the paper is organized as follows: Section II reviews related works. Section III defines the system model and problem definition. Section IV provides the optimal solution to the dummy traffic allocation problem. Section V presents a distributed resource allocation algorithm based on the problem's optimal solution. Section VI validates experimental evaluation of the proposed algorithm. Finally, Section VII concludes the study.

II. Related Works

Source location anonymity has attracted a great deal of interests in recent years [15], [12], [10]. Previous works have considered different adversary capabilities and attack models.

[12] and [8] proposed a random walk based phantom routing protocol to defend against an adversary who attempts to perform hop-by-hop traceback to the source location. Both of these two works consider local attacker model and under global attacker model, their performance will reduce considerably. Source location anonymity under global attacker was explored in [15], [17], [10]. In [17], to provide source unobservability, schemes like *ConstRate* or *ProbRate* are used by the sensors. However, to reduce overall network communication, proxies are introduced which proactively drop dummy messages on their way to the base station. [10], proposes two techniques that prevent the leakage of location information: periodic collection and source simulation. Periodic collection is a *ConstRate* scheme, while source simulation creates multiple candidate traces in the network to hide the traffic generated by real objects.

The basic idea of [15] is that every node in the network sends out dummy messages with intervals following a certain kind of probabilistic distribution. When a node detects a real event, it transmits the real event messages with intervals following the same distribution. In our work, we have made the same assumption about the operation of each individual sender, but we wish to maximize the degree of anonymity of the whole network as a collection of event reporters. Towards this, we used the measure proposed in [6], [14] as *the degree of anonymity* of the system. They used the notion of entropy from information theory as a measure for anonymity. It takes into account the probability distribution of the users performing certain actions, where the probabilities are assigned by an attacker after observing the system.

III. System Model and Problem Formulation

A. Network Model

We assume that there is a base station being responsible for monitoring the network. Sensor network is divided into cells where each pair of nodes in neighboring cells can communicate directly with each other. A cell is the minimum unit for detecting events. This defines an undirected graph. Each sensor node knows its location via its GPS or an specific localization method. Upon detection of an event, each sensor informs the base station about the event by sending a message compromising of control data such as event type, cell ID, etc.

B. Attack Model

Here, we consider anonymity against global eavesdroppers. By a global eavesdropper, we mean an attacker who has access to the whole traffic of the network. Moreover, the attacker is assumed to be external and passive [6]. An external attacker can only monitor communication channels between nodes, whereas an internal one may control nodes in the network. It is easy for the external global adversary to trace back to the source of the message if the encrypted message remains the same while being forwarded. Otherwise, he may perform more advanced attacks including rate monitoring and time correlation.

C. Privacy Evaluation Model

Measuring the anonymity of a system quantitatively is an issue of significant concern. The authors of [6] and [14] propose an information-theoretic measure to assess the anonymity degree of a system. The degree of anonymity takes into account the probability associated with each sender. In information theory, *information entropy* or simply *entropy* is a measure of the uncertainty associated with a random variable. Given a random variable X, the entropy of X is denoted by $H(X)$ and is defined as [5]:

$$H(X) = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{1}$$

Typically, the adversary would like to determine the probabilities of real events. Different cells may appear as having higher or lower probability p_i of observing real events, depending on the information obtained by the adversary. According to the maximum entropy principle, the maximum entropy occurs when the probabilities pursue a uniform distribution. Therefore, we deduce that the maximum effective anonymity for a set of N cells is achieved when all cells receive equal probability (i.e., $p_i = \frac{1}{N}$). In this case, all cells are indistinguishable towards the adversary with respect to observing real events. It is straightforward to show that the entropy in this case will be $\log_2 N$. For the sake of convenience, we denote the maximum entropy by H_M:

$$H_M = \log_2 N \tag{2}$$

We assume that the adversary has no information on the system a priori, hence, before an attack, the attacker assumes that the probability of all cells are uniformly distributed, for which the entropy is H_M. After an attack, the adversary approximate the probability of cells as a vector $\mathbf{p} = (p_1, \ldots p_N)$. Therefore, the amount of information he would gain is the difference between the entropy before and after the attack, i.e. $H_M - H(X)$. Based on this argument, the degree of anonymity is defined as

$$d = 1 - \frac{H_M - H(X)}{H_M} = \frac{H(X)}{H_M} \tag{3}$$

Now we are in a position to formulate the underlying problem. We assume that each sensor i has an average rate λ_i, and therefore its contribution to the total traffic of the network is $\frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i}$. Regarding this setting, it is rational for the adversary to assign $p_i = \frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i}$ as the probability of occurrence of an event in cell i. As mentioned earlier, to achieve anonymity, we mix real event traffic with dummy traffic to conceal patterns of real event traffic. If x_i denotes the allocated resource for sending dummy messages from sensor i, we get

$$p_i = \frac{\lambda_i + x_i}{\sum_{i=1}^{N} (\lambda_i + x_i)} \tag{4}$$

Considering the sensor network as a whole, we argue that the aggregate resources used for dummy traffic must be

constrained as

$$\sum_{i=1}^{N} x_i \leq C \tag{5}$$

where C is the total rate of the dummy traffics, i.e. the maximum affordable cost for sending dummy traffic. Based on this discussion, our objective is to choose dummy traffic rates so as to maximize the degree of anonymity subject to total dummy traffic constraint, so we come up to the problem

$$\min_{\mathbf{x}} \sum_{i=1}^{N} (\lambda_i + x_i) \log_2 \frac{\lambda_i + x_i}{\lambda + C} \tag{6}$$

subject to

$$\sum_{i=1}^{N} x_i \leq C \tag{7}$$

where $\lambda = \sum_{i=1}^{N} \lambda_i$ and $\mathbf{x} = (x_1, \ldots, x_N)$ is the vector representation of dummy traffics. We postpone solving problem (6) to the next section.

IV. Optimal Solution

Regarding problem (6), it is straightforward to show that its objective function is convex. Moreover, the constraint (7) is linear and therefore this optimization problem is a convex minimization problem and consequently admits a unique minimizer denoted by $\mathbf{x}^*$.

Although the objective of (6) is separable among nodes, its constraint (7) is coupled across the network. Such a constrained optimization problem can be efficiently solved using Interior Point Method, which necessitates the coordination among possibly all nodes of the networks, which is undesirable or infeasible. However, in the context of wireless ad-hoc and sensor networks, we are interested in distributive algorithms to solve (6).

Towards this end, we aim at solving problem (6) through its dual. In the sequel, we proceed to derive the dual problem of (6) and then present a distributively iterative algorithm as the solution to the dual problem.

A. Deriving Dual Problem

We start by writing the Lagrangian of problem (6), as follows

$$L(\mathbf{x}, \mu) = \sum_{i=1}^{N} (\lambda_i + x_i) \log_2 \frac{\lambda_i + x_i}{\lambda + C} + \mu \left(\sum_{i=1}^{N} x_i - C \right) \tag{8}$$

where $\mu > 0$ is the Lagrange multiplier associated with constraint (7). Using Karush-Kuhn-Tucker (KKT) conditions for convex optimization, to characterize the optimal values $\mathbf{x}^*$, we should find the stationary points of the Lagrangian and satisfy complementary slackness conditions. The complementary slackness conditions for optimal primal variable $\mathbf{x}^*$ and dual variable μ^*, are

$$\mu^* \geq 0 \tag{9}$$

$$\sum_{i=1}^{N} x_i^* \leq C \tag{10}$$

$$\mu^* \left(\sum_{i=1}^{N} x_i^* - C \right) = 0 \tag{11}$$

In order to find the stationary point of the Lagrangian, we solve

$$\nabla L(\mathbf{x}^*, \mu^*) = \mathbf{0} \tag{12}$$

where $\mathbf{0}$ is a vector with all zero. For the ith element of (12), we get

$$\frac{\partial L}{\partial x_i} = \log_2 \frac{\lambda_i + x_i}{\lambda + C} + \mu + \log_2 e \tag{13}$$

As a result, we come up to the following equation to obtain $\mathbf{x}^*$ as

$$x_i^* = (\lambda + C) 2^{-(\log_2 e + \mu)} - \lambda_i \tag{14}$$

B. Deriving Dual Problem

In order to solve problem (6) through its dual, we need to obtain the Lagrange dual function, or simply dual function. The Lagrange dual function $D(\mu)$ is defined as the minimum of the Lagrangian $L(\mathbf{x}, \mu)$ over the primal variable $\mathbf{x}$, for a given μ. Thus, $D(\mu)$ can be expressed as

$$D(\mu) = \min_{\mathbf{x}} L(\mathbf{x}, \mu) \tag{15}$$

Based on the results of the KKT condition mentioned above, minimization in (15) is already solved with $\mathbf{x}^*$ given by (14), which results in

$$D(\mu) = L(\mathbf{x}^*, \mu) \tag{16}$$

The dual problem is formulated as

$$\max_{\mu \geq 0} D(\mu) \equiv \max_{\mu \geq 0} L(\mathbf{x}^*, \mu) \tag{17}$$

Dual problem defined above can be solved using iterative methods. In order to obtain a distributed algorithm, we solve the dual problem (17) using Gradient Projection Method. We postpone solving (17) to the next subsection.

C. Solving The Dual

In this subsection, we solve the dual problem using Gradient Projection Method [3]. To solve the dual problem, Gradient Projection Method adjusts μ in the direction to the Gradient of dual function, i.e. $\nabla D(\mu)$. Precisely speaking, in the kth iteration step, $\mu^{(k)}$ is updated as follows

$$\mu^{(k+1)} = \left[\mu^{(k)} + \gamma \frac{dD(\mu^{(k)})}{d\mu} \right]^+ \tag{18}$$

where $[z]^+ = \max\{z, 0\}$ and γ is a sufficiently small constant step size. Using the Danskin's Theorem [3], the derivative of $D(\mu)$ is given by

$$\frac{dD(\mu)}{d\mu} = \sum_{i=1}^{N} x_i - C \tag{19}$$

Substituting (19) in (18), yields

$$\mu^{(k+1)} = \left[\mu^{(k)} + \gamma\left(\sum_{i=1}^{N} x_i^{(k)} - C\right)\right]^+ \tag{20}$$

where $x_i^{(k)}$ is the solution to (14) for a given $\mu^{(k)}$. In this equation, γ is chosen sufficiently small so as to guarantee the convergence.

V. Distributed Resource Allocation Algorithm

In this section, we propose a distributed algorithm based on the iterative solutions obtained in Section IV. Considering (20) and (14), it is clear that the iterative solution to the problem (6) can be regarded as a distributed algorithm.

In particular, each sensor i in the iteration step k benefits from all other nodes' current value x_i's, thanks to Flooding-like algorithms, to update the current dual variable $\mu^{(k)}$. Since all other nodes have access to such information too, they will obtain the same value for $\mu^{(k+1)}$ and therefore, we don't introduce additional notation to distinguish between the realized update process. Upon updating $\mu^{(k)}$, each sensor i calculates its dummy traffic $x_i^{(k)}$, accordingly. The above rule will proceed until reaching some predefined notions of convergence. This algorithm is listed as Algorithm 1.

A. Convergence Analysis

In this subsection, we investigate the convergence behavior of the proposed algorithm. As step size has an important role ons the convergence behavior of the update equation (20), we mainly focus on the effect of step size. The following theorem determines the necessary condition on the step size, under which Algorithm 1 converges to the neighborhood of the optimum of problem (17) and thereby that of problem (6).

Theorem 1: The iterative traffic assignment algorithm proposed by (14) and (20) converges to the neighborhood of the optimal point of problem (6) provided that

$$0 < \gamma < \frac{2}{\ln 2(C+\lambda)} \tag{21}$$

Proof: See Appendix I for proof. ■

VI. Experimental Evaluation

We have conducted simulation experiments to evaluate the performance of our proposed algorithm. We verify that the algorithms, locally executed on each node, may indeed achieve the desired global optimal resource allocation.

In our simulation scenario, we consider a sensor network consisting of 100 sensor nodes which randomly scattered over the unit square $[0,1] \times [0,1]$ area.

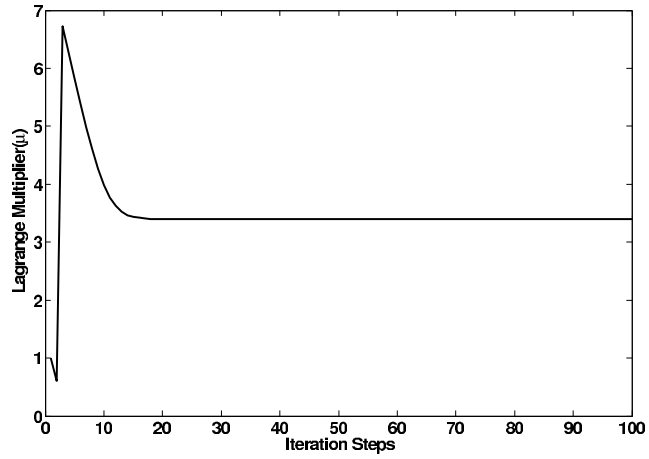

Fig. 1. Evolution of Lagrange Multiplier μ Using Algorithm 1

Algorithm 1. Distributed Dummy Traffic Allocation

Initialization

Initialize C $\forall i = 1..n$.

Main Loop

Do until $\max_i |x_i^{(k+1)} - x_i^{(k)}| < \epsilon$

1. At each sensor node, update the dual variable as following:

$$\mu^{(k+1)} = \left[\mu^{(k)} + \gamma\left(\sum_{i=1}^{N} x_i^{(k)} - C\right)\right]^+$$

2. Update x_i according to the following equation:
$x_i^{(k+1)} = (\lambda + C)2^{-(\log_2 e + \mu^{(k)})} - \lambda_i$

Algorithm 1. Distributed Dummy Traffic Allocation

We assume that each node i incessantly observes its cell. We assume that event occurrence in each cell obeys a Poisson distribution with the parameter λ_i. Step size is chosen to be $\gamma = 0.02$.

The most significant issues of interest is the evolutions of dual variables. Evolution of Lagrange multiplier μ for Algorithm 1 is depicted in Fig. 1. It is apparent from this figure that by spending less than 20 iteration steps, convergence was achieved and thereafter, μ had intangible variations.

Degree of Anonymity achieved under different cost constraints is demonstrated in Fig 2. It illustrates that by increasing the total amount of dummy traffic shared among sensors, a higher level of anonymity is attained; However, after allocation of some amount of resource between nodes we approximately reach the maximum anonymity.

VII. Conclusion

There have been many studies exploring various applications of WSNs such as monitoring. In this paper, we addressed the problem of *Source Location Anonymity* we encounter in such monitoring applications. Most of existing approaches on location privacy in sensor networks had assumed that the attacker has only a local monitoring capability. Here we

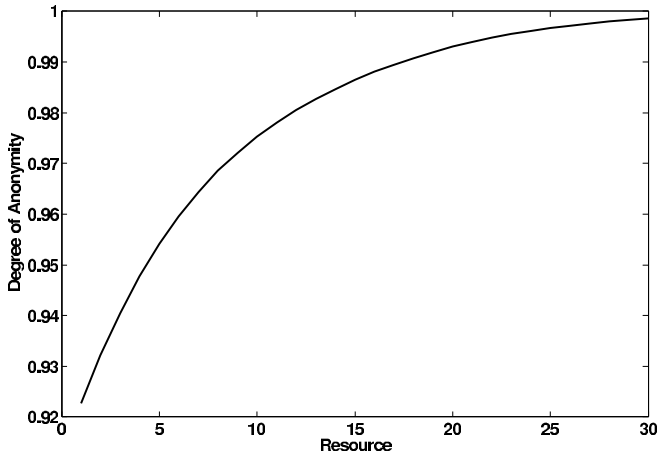

Fig. 2. Degree of Anonymity Under Different Allocated Resources

investigate the source anonymity problem under the global attacker model. To achieve source anonymity we propose a scheme to combine the real event traffic with carefully chosen dummy traffic. In the proposed method we have considered a resource-constrained sensor network, thus we formulate the problem toward attaining maximum degree of anonymity under a fixed affordable communication. We solved the this problem via its dual so as to achieve a distributed algorithm as the solution to the problem. The results extracted from the experimental evaluation demonstrated the achieved anonymity is much higher than the naive method of distributing resources among the sources and is more energy efficient than the constant rate scheme.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: A Survey," Computer Networks, vol. 38, no. 4, pp. 393-422, March 2002.

[2] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, et al. "A Line in The Sand: A Wireless Sensor Network for Target Detection, Classification, and Tracking," *Computer Networks*, pp. 605-634, 2004.

[3] Dimitri Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[4] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Gossip Algorithms: Design, Analysis, and Applications," In *INFOCOM*, pp. 1653-1664, 2005.

[5] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.

[6] C. Diaz, S. Seys, J. Claessens and B. Preneel, "Towards Measuring Anonymity," In *PET*, pp. 54-68, 2002.

[7] S. Gandhi, J. Hershberger and S. Suri, "Approximate Isocontours and Spatial Summaries for Sensor Networks," In *IPSN*, pp. 400-409, 2007.

[8] P. Kamat, Y. Zhang, W. Trappe and C. Ozturk, "Enhancing source location privacy in sensor network routing," In *ICDCS*, pp. 599-608, 2005.

[9] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, et al. "Wireless Sensor Networks for Habitat Monitoring," In *WSNA*, pp. 88-97, 2002.

[10] K. Mehta, D. Liu and M. Wright, "Location Privacy in Sensor Networks Against a Global Eavesdropper," In *ICNP*, pp. 314-323, 2007.

[11] Y. Ouyang, Z. Le, G. Chen, J. Ford and F. Makedon, "Entrapping Adversaries for Source Protection in Sensor Networks," In *WoWMoM*, pp. 23-34, 2006.

[12] C. Ozturk, Y. Zhang and W. Trappe, "Source-location Privacy in Energy Constrained Sensor Network Routing," In *SASN*, pp. 88-93, 2004.

[13] A. Pfitzmann and M. Hansen, "Anonymity, Unobservability, and Pseudonymity: A Consolidated Proposal for Terminology," Draft, July 2000.

[14] A. Serjantov and G. Danezis, "Towards an Information Theoretic Metric for Anonymity," In *PET*, pp. 259-263, 2002.

[15] M. Shao, Y. Yang, S. Zhu and G. Cao, "Towards Statistically Strong Source Anonymity for Sensor Networks," In *IEEE INFOCOM*, 2008.

[16] "The Free Haven Project, http://freehaven.net/anonbib/date.html, 2005.

[17] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar and G. Cao, "Towards Event Source Unobservability with Minimum Network Traffic in Sensor Networks," In *WiSec*, pp. 77-88, 2008.

APPENDIX I: PROOF OF THEOREM 1

According to the duality theory, whenever the strong duality is held, the duality gap is zero and the optima of the dual leads to that of primal. Hence it suffices to seek the condition on the step size under which (20) converges to a neighborhood of the dual-optimal point.

According to the Gradient Projection Method and Descent Lemma [3], to establish the convergence condition of the update equation, the step size must satisfy the following

$$\epsilon \leq \gamma \leq \frac{2-\epsilon}{M} \tag{22}$$

where M is the constant which establishes Lipschitz Continuity for the objective function. In order to satisfy Lipschitz Continuity for $D(\mu)$, it suffices to show that the second derivative of $D(\mu)$ is upper bounded. Considering (19), for the second derivative of $D(\mu)$, we have

$$\frac{d^2D}{d\mu^2} = \sum_{i=1}^{N} \frac{dx_i}{d\mu} \tag{23}$$

Using (14), for $\frac{dx_i}{d\mu}$, we have

$$\frac{dx_i}{d\mu} = \ln 2(\lambda + C)2^{-(\log_2 e+\mu)} \tag{24}$$

Therefore, for $\frac{d^2D}{d\mu^2}$, we get

$$\begin{aligned} \frac{d^2D}{d\mu^2} &= \sum_{i=1}^{N} \ln 2(\lambda + C)2^{-(\log_2 e+\mu)} \\ &= \sum_{i=1}^{N} \ln 2(x_i + \lambda_i) \\ &= (\ln 2)\left(\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} \lambda_i\right) \\ &\leq (\ln 2)(C + \lambda) \end{aligned} \tag{25}$$

Therefore, $\frac{d^2D}{d\mu^2}$ is upper bounded at least as follows

$$\frac{d^2D}{d\mu^2} \leq (\ln 2)(C + \lambda) \tag{26}$$

Therefore, for sufficiently small ϵ, we conclude

$$0 < \gamma < \frac{2}{\ln 2(C + \lambda)} \tag{27}$$

which completes the proof.