Towards Fake Sources for Source Location Privacy in Wireless Sensor Networks with Multiple Sources

Joanna F. Laikin, Matthew Bradbury, Chen Gu and Matthew Leeke
Department of Computer Science, University of Warwick,
Coventry, United Kingdom, CV4 7AL
E-mail: {j.f.laikin, m.bradbury, c.gu.1, m.leeke}@warwick.ac.uk

***Abstract*—Wireless sensor networks (WSNs) are regularly used in asset monitoring applications, where the location of an asset or assets must be kept private. Providing location privacy for such an asset is tantamount to protecting the location of a source node from an attacker who is attempting to locate it. Although no solution exists to provide source location privacy over an extended period, it has been shown that attackers can be sufficiently inhibited by prominent approaches that use either a phantom node, via which protocol messages are routed, or nodes assigned to be fake sources, each of which then broadcast fake messages. However, the applicability of fake source approaches to networks where location privacy must be maintained for multiple sources has yet to be considered. This paper addresses this issue by analysing a representative fake source algorithm in the context of multiple sources, presenting simulation results that demonstrate the shortcomings of the approach and identifying the underlying limitations to pave the way for the development of algorithms capable of accounting for multiple sources.**

***Keywords*-Context Privacy; Fake Source; Location; Multiple Sources; Wireless Sensor Networks**

I. Introduction

Wireless sensor networks (WSNs) are frequently cited as an enabling technology for the Internet of Things (IoT), as well as being fundamental to solving problems in application domains such as animal conservation, healthcare and military intelligence [1], [2], [3]. The need to monitor assets, such as endangered animals or military personnel, is common across such application domains. As monitored assets are invariably valuable, consideration must be given to the way in which sensed information is communicated during monitoring. In many situations it is desirable for the location of the asset being monitored to be kept private.

The source location privacy (SLP) problem focuses on ensuring that the location of a source node or asset can only be observed or inferred by those intended to observe or decipher it [4]. WSNs operate in a broadcast medium, which means attackers can intercept messages and use the knowledge gained to locate assets. Three broad categories of techniques exist for solving the SLP problem in the context of a distributed local eavesdropper: (i) fake source techniques that allocate nodes to act like sources but send fake messages that are indistinguishable from real messages [5], (ii) routing-based techniques [4], and (iii) hybrid techniques that combine both routing and fake sources [6], [7]. In this paper we focus on the fake source approach, though much of the analysis presented is similarly relevant to routing-based and hybrid approaches.

Much research in the application of fake sources has focused on the development of algorithms that preserve SLP whilst minimising network energy consumption [7], [8]. Work has also focused on examining heuristics for the selection of fake sources, a problem that is known to be NP-complete, in order to develop adaptive algorithms that are suitable for practical deployment [5], [9]. Despite this body of work, little has been done to assess the applicability of the fake source approach in networks with multiple sources, a characteristic of an increasing number of WSN application scenarios and a requirement in the context of the IoT.

The need to preserve the location privacy of multiple assets is a practical issue that must be overcome before the use of WSNs can become widespread in monitoring applications, not least because single asset scenarios can rarely justify the considerable expense of deploying a large-scale WSN. The work presented in this paper moves to address this challenge, since it provides the results and analysis that demonstrate the immediate applicability of current fake source algorithms to inform the adaptation or development of algorithms to account for multiple information sources.

A. Contributions

In this paper we contribute to the understanding of existing fake source algorithms for source location privacy in networks containing multiple information sources. In doing this we:

- Provide privacy and energy focused simulation results for fake source algorithms operating in networks with multiple information sources.
- Demonstrate that, in the worst case, communicating with no context-privacy preserving algorithm can yield better privacy than a fake source algorithm in networks with multiple sources.
- Identify shortcomings in the considerations of fake source algorithms that lead to the location privacy of any single source being compromised.

The remainder of this paper is as follows. In Section II we provide a survey of related research. The network and attacker models are detailed in Section III. Section IV provides details of the simulation experiments and the fake source algorithm to be analysed. The results generated by these experiments are analysed and discussed in Section V, before Section VI concludes the paper with a summary of outcomes.

II. Related Work

Seminal research in SLP first defined the problem of location privacy for WSNs [4], [10]. The authors of [4] developed widely adopted formalisations of the problem and explored a set of privacy-preserving algorithms. They proposed the *fake source* approach to solving the SLP problem, whilst also acknowledging its poor performance in terms of maintaining location privacy and high energy consumption requirements. The problem subsequently attracted much research interest, often motivated by the needs of specific problem domains [11], [12], [13]. It has since been shown that, for some categories of attacker model, fake sources can provide high levels of SLP whilst realising a trade-off between security and network energy consumption [7], [8].

Alternative approaches to the SLP problem have been proposed, including prominent techniques that build on work in phantom routing [4]. The majority of this work has focused on altering the nature of the random walks used in routing [14], [15], [16], with some research seeking, as with fake sources, to balance privacy and energy consumption [17]. However, as it has been shown that many phantom routing-based techniques are vulnerable to correlation-based source identification, routing traceback, and reducing source space, the use of fake sources remains an active area of research.

Despite much interest in the fake source approach, many practical issues must be addressed before the technique can be applied to a wider spectrum of monitoring problems. Work in the past decade has addressed the issue of having multiple sink nodes, often under strong network model assumptions and with a view to minimising energy consumption [18], [19]. However, little research has considered the implications of using state-of-the-art fake source approaches in networks with multiple sources, the issue that is the focus of this paper.

III. Models

A. Network Model

A wireless sensor node has a unique identifier and a limited set of computational capabilities. It is equipped with a radio transmitter for communication. A WSN is a set of wireless sensor nodes with communication links between pairs of nodes. We assume that all nodes in the network have the same communication range. The nodes in direct communication range with a node n are known as the neighbours of n.

There exists a distinguished node in the network, known as the *sink*, which is responsible for collecting data and which acts as a link between the WSN and the external world. Other nodes sense data and route it attached to messages along a computed route to the sink. We assume the network is event-triggered, i.e., when a node senses an object, it starts sending messages periodically to the sink. We assume the message to be encrypted and that the source nodes include their ID in messages. Using the ID the sink can infer an asset's location. We do not assume that WSN nodes have access to GPS.

B. Attacker Model

It was proposed in [20] that the strength of an attacker for WSNs can be factored along two dimensions, namely *presence* and *actions*. Presence captures the network coverage of the attacker, while actions capture the attacks the attacker can launch. We assume a *mobile distributed eavesdropper* attacker based on the patient adversary, introduced in [4]. Such an attacker is reactive in nature and initially starts at the sink. In general, the attacker need not start at the sink, this simply guarantees they will receive a message. When the attacker is located at a node n and receives a message from a neighbour node m, the attacker will move to m, that 1-hop neighbour of n from which he received the message, if that message had not been received before. We assume the attacker has the same communication range as the network nodes. To detect if a message has been received before, we assume that an attacker has access to the message type, sequence number and source ID. When an attacker receives a message, they can either move one step in an inferred direction. Repeating this action for a number of times may enable the attacker to capture ab asset based on a traceback of the traffic flow to the asset.

Once the source has been found, the attacker will no longer move. We assume that the attacker has the capability to detect the direction of message arrival and and a large amount of memory to keep track of information such as messages that have been heard. This is commensurate with the attacker models used in [5], [21], [22]. Our aim in providing SLP is to make using the context of messages too expensive for attackers, such that they instead choose to perform alternate attacks such as a brute force search of the network.

The adopted model may be considered relatively weak compared to the more powerful attacker models proposed in [20]. This weaker model has been chosen because some of the stronger models include behaviours that are unlikely for the attacker to perform as he will not be able to extract further information from the network. The authors of [20] developed a total order on the set of actions an attacker can perform on a WSN. The order is as follows:

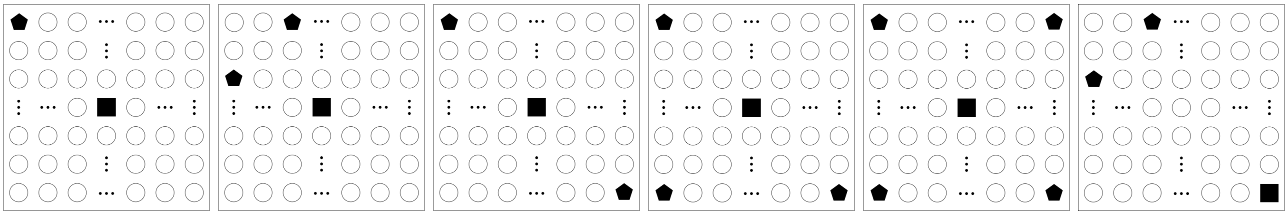
eavesdrop → *crash* → *disturbing* → *limited passive* →
passive → *reprogramming*

Aside from eavesdropping, an attacker could attempt to disrupt the functioning of the network, e.g., a DoS attack, but this would limit the amount of useful information the attacker could gather. Alternatively, the attacker could try to modify a node's software in order to learn more information about the network, but this is a time-consuming process, that would lead to the safety period being exceeded, i.e., the asset will have moved location. In both cases, the attacker will either not learn anything useful or the action will take too long.

IV. Experimental Setup

A. Network Simulation and Configurations

The TOSSIM (v2.1.2) simulation environment was used in all experiments [23]. TOSSIM is a discrete event simulator



(a) (b) (c) (d) (e) (f)

Fig. 1: Network configurations for multiple source simulation experiments.

TABLE I: Link Layer Model parameters.

Name	Value
PATH_LOSS_EXPONENT	4.7
SHADOWING_STANDARD_DEVIATION	3.2
D0	1.0
PL_D0	55.4
NOISE_FLOOR	-105
S	[(0.9, -0.7); (-0.7, 1.2)]
WHITE_GAUSSIAN_NOISE	4

capable of accurately modelling sensor nodes and the modes of communications between them. The fake source protocol implemented was the adaptive algorithm proposed in [22], since this provided best-in-class privacy whilst obviating the need for an extensive exploration of parameters.

A square grid network layout of size $n \times n$ was used, where $n \in \{11, 15, 21\}$, i.e., networks with 121, 225, and 441 sensor nodes. These network sizes where chosen to maintain comparability with previous work in source location privacy [22]. The node neighbourhoods were generated using Link Layer Model with the parameters shown in Table I, which gives a small chance of asymmetric links occurring. The network configurations are shown in Figure 1, where nodes, sources and the sink are represented by circles, pentagon and squares respectively. These network configurations are commensurate with those in [5] and remained consistent across network sizes.

B. Safety Period and Broadcast Rates

The overall objective of any WSN-based SLP solution is to ensure that an asset is *never* captured through the WSN. However, two issues arise: (i) if the asset is not mobile, the attacker can take as long as it requires to perform an exhaustive search of the network, and (ii) if the asset is mobile, performing an exhaustive search of the network is unsuitable as the attacker may focus on a location only to find the asset has moved. Knowing this, the SLP problem can only be considered when it is time-bounded.

This notion of time-boundedness has been termed *safety period*. There are two competing definitions of safety period. The first, used primarily by routing-based techniques, e.g., [4], is where the safety period is defined as the time required to capture the asset. The aim of these techniques is to maximise the safety period, i.e., the higher the time to capture, the higher the level of SLP provided [4]. On the other hand, the second notion of safety period is used where it is desirable to bound the amount of time over which SLP is considered. Specifically, SLP is then said to be provided if an attacker fails to capture a source within the safety period. The safety period is set such that it has a value greater than the time to capture.

Allowing the safety period to be greater than the capture time means that the level of SLP obtained is bounded below. Specifically, denoting safety period by τ_s and capture time by τ_c, if an asset is not captured at τ_s, then it means that the asset is not captured at τ_c meaning that the SLP with τ_s is lower bounded. The safety period intuitively captures the maximum time an asset will be at a given location before its next movement. Often, this can be obtained from previous data gathering to know more about such mobile assets. For a given network size and source rate, using flooding as a base routing protocol, we calculate the average time it takes the attacker to detect the real source, i.e., capture the asset. We term this the capture time. The safety period is twice the capture time.

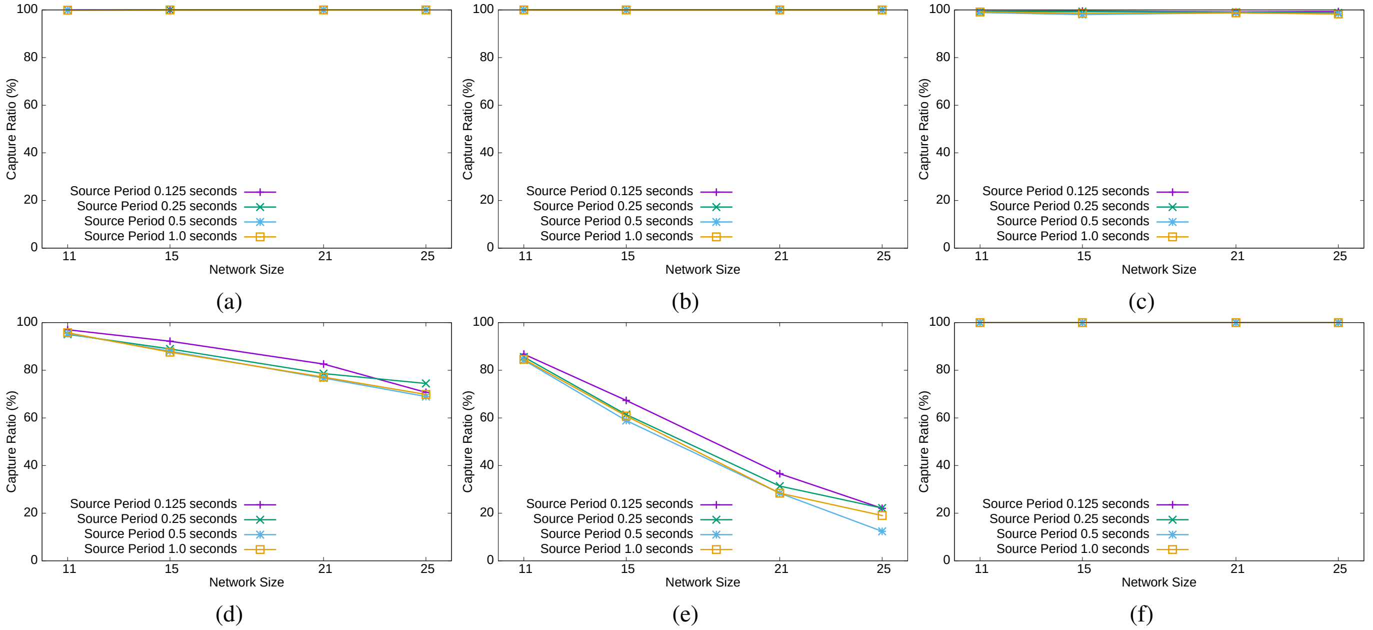
The rate at which messages were generated from each source was varied, with results for 1, 2, 3 and 4 messages per second shown in Section V as source periods of 1, 0.5, 0.25 and 0.125 seconds respectively. Results for other messages rates was gathered, but only these two are shown for brevity. A total of 500 repeats were performed for each fake source experiment. Nodes were located 4.5 meters apart. The node separation distance was determined experimentally, based on observing the pattern of transmissions in the simulator. This separation distance ensures that messages (i) pass through multiple nodes from source to sink, (ii) can move only one hop at a time, and (iii) will usually only be passed to horizontally or vertically adjacent nodes.

V. Results

A. Protectionless Privacy

To evaluate the privacy afforded by a source location privacy algorithm, the inherent privacy provided by a network must be understood. To do this we perform simulations for each network configuration with no fake source protocol.

Figure 2 shows capture ratio against networks size for protectionless network configurations with varying message broadcast rates. As expected, Figure 2(a) shows that, in the absence of any privacy preserving algorithm, the capture ratio for a single source network is consistently 100% across all network sizes. Similarly, Figures 2(b), 2(c) and 2(f) show that the capture ratio is either 100% or close to it for all networks with two sources, implying that the addition of a



(a) (b) (c)

(d) (e) (f)

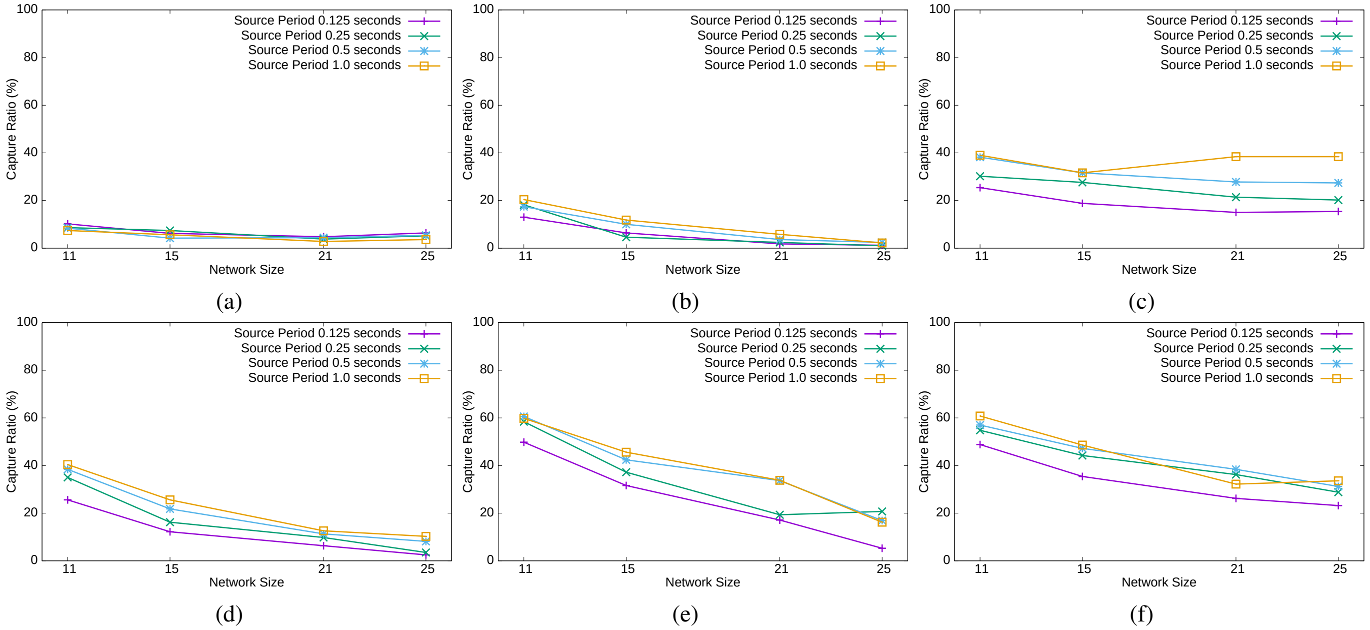
Fig. 2: Capture ratio plotted against network size for protectionless network configurations with varying message broadcast rates. Figures 2(a)-(f) correspond to the network configurations shown in Figures 1(a)-(f) respectively.

further source has done little to hinder the attacker in their pursuit of a source capture. The variation in network traffic caused by multiple sources can produce a push-pull effect on that attacker, similar to that observed in [22], that results in less informed decision making. This effect can be seen in Figure 2(c), where the attacker occasionally fails to make a capture in a two source network, leading to a capture ratio of less than 100%. Such high capture ratios can not be seen in Figures 2(d) and 2(e), which implies that having more than two sources can make it more challenging for an attacker to capture a source node. It also serves to validate the scale of the simulations, demonstrating that sources can not be trivially located in multiple source networks of these sizes whilst informing our view of the lower bound on practical network size. Indeed, despite the multiple source configurations in Figures 2(d) and 2(e) demonstrating inherent privacy, the capture ratios remain high and form an appropriate baseline.

B. Privacy Preservation

Capture ratio is a metic used to indicate the level of SLP afforded. Capture ratio is calculated as the ratio of the number of runs in which a source is captured within the safety period to the total number of runs. A capture ratio of 0% is indicative of the highest possible levels of privacy, whilst a capture ratio of 100% implies the that no privacy is being afforded. Research presented in [22] demonstrated that the algorithm analysed achieves capture ratios of 0-7% in single source networks.

Figure 3 shows the capture ratio against networks size with varying broadcast rates for each network configuration. The capture ratios shown in Figure 3(a) are commensurate with those achieved in [5], providing confidence in the efficacy of approach and the simulation results presented for multiple source networks. It is informative to contrast Figures 3(b), 3(c) and 3(f), which are configurations of two sources that achieve differing capture ratios. In 3(f) the capture ratio remains below other two source configurations. Here the relative position of the sources unifies them in their goal of dissuading an attacker. The sources are in close enough proximity, given a particular network scale, to allow the fake algorithm to be constructive in fake source selection. In contrast, the configuration in Figure 3(c) creates a situation where the creation of fake sources is destructive in realising the privacy preservation of each source. This situation is exasperated in Figure 3(f), which suffers from the sink being at a network extremity. The problem of providing SLP in this situation has been acknowledged [5]. This effect is only amplified where sources in close proximity increase the information on the network for an attacker to use in making inferences, as is the case in Figure 3(f). The capture ratios for configurations with more than two sources, shown in Figures 3(d) and 3(e), indicate that the fake source algorithm generally improved privacy. Moreover, the best case privacy afforded in these cases is not dissimilar to that of a single source network, at around 4% and 6% for the largest three and four source networks respectively. The performance depicted in Figure 3(d) is encouraging in this regard, though it is crucial to note that the average case performance for configurations with multiple sources, whilst better in general, can deteriorate due to a privacy preserving algorithm. For example, in the case of sources with source period of 0.25-1.0 in Figure 3(e), the afforded privacy is commensurate or worse than the same configurations for the protectionless cases in Figure 2(e). This worst case suggests that fake source algorithms founded on shared assumptions must be reconsidered to account multiple sources.



(a) (b) (c)

(d) (e) (f)

Fig. 3: Capture ratio plotted against network size for each network configuration with varying message broadcast rates. Figures 3(a)-(f) correspond to the network configurations shown in Figures 1(a)-(f) respectively.

C. Energy Efficiency

Broadcasting is typically the most energy consuming task that a sensor node can perform. This makes the number of messages broadcast is a reasonable indicator of network energy consumption when considering algorithms independently of deployment scenario, sensor technology or broadcast protocol.

Figure 4 shows the mean number of messages broadcast per node per second against networks size with varying broadcast rates for each network configuration. In an event-triggered network it is intuitive that increasing the source period results in a commensurate increase in the mean number of messages broadcast regardless of configuration. It is less apparent that having multiple sources in a network should reduce the mean number of messages broadcast, as seen in Figure 4. This reduction is, in part, associated with the reduced number of messages broadcast in the restricted regions reserved around each source node. The number of the these regions increases linearly with the number of sources, though regions may overlap to produce areas of constructive or destructive interference. Reducing the number of messages broadcast as more source are added is a desirable characteristic, though this efficiency must be balanced against privacy. Indeed, as these results indicate, the development of an energy efficient, fake source algorithm for multiple sources remains an open problem.

D. Discussion

The results presented motivate the development of fake source algorithms that are applicable in the context of networks with multiple sources, not least because the worst-case privacy afforded by current generation algorithms can deteriorate to levels below that of a protectionless network. In the design of these algorithms, the results imply that it is not only the number of sources that impact afforded privacy, the relative locations of the sources are significant in determining the likelihood of their privacy being preserved. Knowledge of relative locations would, even for adaptive fake source algorithms, provide opportunities to obviate the destructive interference caused by overlapping broadcast restrictions, remedy network extremity source location and utilise the energy efficient characteristics of current algorithms. This positional knowledge should be foremost in the design of fake source algorithms.

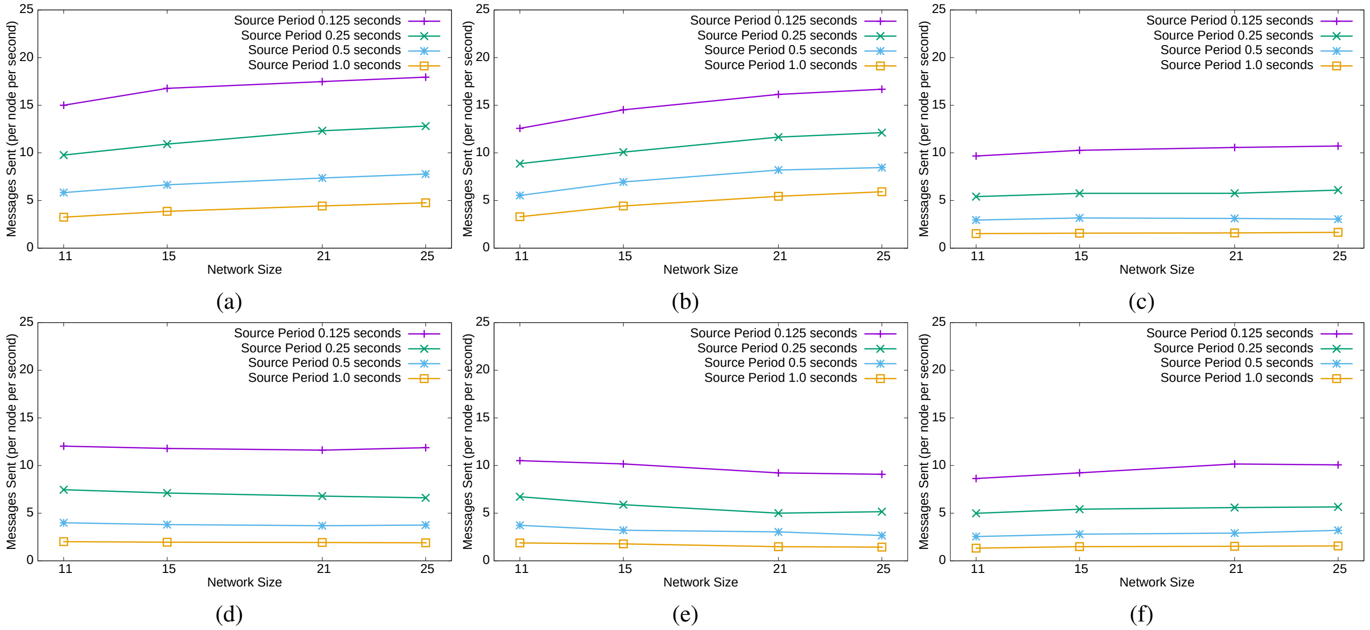
It should be noted that the results presented are based on the consideration of a limited set of network configurations, chosen for consistency with existing literature or to expose an underlying performance characteristic. Similarly, the algorithm was selected for its alignment with the body of existing fake source algorithms and its known level of privacy provision.

VI. Conclusion

Privacy preserving algorithms for WSNs with multiple sources are becoming fundamental to problems in a wide range of application domains. This paper has presented results for a representative fake source algorithm operating in networks with multiple information sources. These results demonstrated that, in the worst case, using protectionless communication can provide better privacy than a fake source algorithms in networks with information multiple sources, as well as being used to identify shortcomings in the considerations of current fake source algorithms, particularly with regard to their accounting for the relative locations of sources.

References

[1] A.-M. Badescu and L. Cotofana, “A wireless sensor network to monitor and protect tigers in the wild,” *Ecological Indicators*, vol. 57, pp. 447–451, 2015.



(a) (b) (c)

(d) (e) (f)

Fig. 4: Mean messages broadcast per node per second plotted against network size for each network configuration with varying message broadcast rates. Figures 4(a)-(f) correspond to the network configurations shown in Figures 1(a)-(f) respectively.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless sensor networks for habitat monitoring,” in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*. New York, NY, USA: ACM, 2002, pp. 88–97.

[3] A. Milenković, C. Otto, and E. Jovanov, “Wireless sensor networks for personal health monitoring: Issues and an implementation,” *Computer Communications*, vol. 29, no. 13-14, pp. 2521–2533, Aug. 2006.

[4] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, “Enhancing source-location privacy in sensor network routing,” in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*. IEEE, Jun. 2005, pp. 599–608.

[5] A. Jhumka, M. Bradbury, and M. Leeke, “Fake source-based source location privacy in wireless sensor networks,” *Concurrency and Computation: Practice and Experience*, vol. 27, no. 12, pp. 2999–3020, 2015.

[6] H. Chen and W. Lou, “On protecting end-to-end location privacy against local eavesdropper in wireless sensor networks,” *Pervasive and Mobile Computing*, vol. 16:A, pp. 36–50, January 2015.

[7] J. Long, M. Dong, K. Ota, and A. Liu, “Achieving source location privacy and network lifetime maximization through tree-based diversionary routing in wireless sensor networks,” *IEEE Access*, vol. 2, no. 1, pp. 633–651, June 2014.

[8] A. Jhumka, M. Leeke, and S. Shrestha, “On the use of fake sources for source location privacy: Trade-offs between energy and privacy,” *The Computer Journal*, vol. 54, no. 6, pp. 860–874, 2011.

[9] A. Jhumka, M. Bradbury, and M. Leeke, “Towards understanding source location privacy in wireless sensor networks through fake sources,” in *Proceedings of the 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, Jun. 2012, pp. 760–768.

[10] C. Ozturk, Y. Zhang, and W. Trappe, “Source-location privacy in energy-constrained sensor network routing,” in *Proceedings of the 2nd ACM Workshop on Security of Ad-hoc and Sensor Networks*. New York, NY, USA: ACM, October 2004, pp. 88–93.

[11] M. Conti, J. Willemsen, and B. Crispo, “Providing source location privacy in wireless sensor networks: A survey,” *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1238–1280, January 2013.

[12] R. Rios, J. Lopez, and J. Cuellar, *Foundations of Security Analysis and Design VII: FOSAD 2012/2013 Tutorial Lectures*. Springer International Publishing, September 2014, ch. Location Privacy in WSNs: Solutions, Challenges, and Future Trends, pp. 244–282.

[13] N. Li, N. Zhang, S. K. Das, and B. Thuraisingham, “Privacy preservation in wireless sensor networks: A state-of-the-art survey,” *Ad Hoc Networks*, vol. 7, no. 8, pp. 1501–1514, 2009.

[14] P. Spachos, D. Toumpakaris, and D. Hatzinakos, “Angle-based dynamic routing scheme for source location privacy in wireless sensor networks,” in *Proceedings of the 79th IEEE Vehicular Technology Conference*, May 2014, pp. 1–5.

[15] P. Kumar, J. Singh, P. Vishnoi, and M. Singh, “Source location privacy using multiple-phantom nodes in WSN,” in *Proceedings of the IEEE Region 10 TENCON Conference*. IEEE, Nov. 2015, pp. 1–6.

[16] L. Zhang, “A self-adjusting directed random walk approach for enhancing source-location privacy in sensor network routing,” in *Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing*. New York, NY, USA: ACM, 2006, pp. 33–38.

[17] Y. Xi, L. Schwiebert, and W. Shi, “Preserving source location privacy in monitoring-based wireless sensor networks,” in *Proceedings of the 20th International Parallel and Distributed Processing Symposium*. IEEE, Apr. 2006, pp. 355–363.

[18] M. Soyturk, “A novel stateless energy-efficient routing algorithm for large-scale wireless sensor networks with multiple sinks,” in *Proceedings of the 2006 IEEE Annual Wireless and Microwave Technologuy Conference*. IEEE, December 2006, pp. 1–5.

[19] E. Lee, S. Park, F. Yu, and S.-H. Kim, “Back to results communication model and protocol based on multiple static sinks for supporting mobile users in wireless sensor networks,” *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1652–1660, August 2010.

[20] Z. Benenson, P. M. Cholewinski, and F. C. Freiling, *Wireless Sensors Networks Security*. IOS Press, April 2008, ch. Vulnerabilities and Attacks in Wireless Sensor Networks, pp. 22–43.

[21] A. Thomason, M. Leeke, M. Bradbury, and A. Jhumka, “Evaluating the impact of broadcast rates and collisions on fake source protocols for source location privacy,” in *Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, Jul. 2013, pp. 667–674.

[22] M. Bradbury, M. Leeke, and A. Jhumka, “A dynamic fake source algorithm for source location privacy in wireless sensor networks,” in *Proceedings of the 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, Aug. 2015, pp. 531–538.

[23] P. Levis, N. Lee, M. Welsh, and D. Culler, “Tossim: accurate and scalable simulation of entire tinyos applications,” in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*. New York, NY, USA: ACM, November 2003, pp. 126–137.