Towards a Statistical Context for Source Obscurity in Sensor Networks

[1]Shrikant, [2]Lilavati S.Samant

Department of Computer Engineering, Assistant Professor, SDIT, Mangalore
Department of Computer Engineering, Student (Mtech, CS&E),SDIT, Mangalore.
Email: lilavatisamant@gmail.com

Abstract—Source Obscurity in Sensor Network is the main aim of the paper. That is, unauthorized observers must be unable to detect the origin of events by judging the network traffic. This literature is provides a method for achieving source obscurity in wireless sensor network & converting real valued samples to binary codes.

Index Terms—Wireless Sensor Networks (WSN), source location, privacy, anonymity, nuisance parameters, coding theory

I. INTRODUCTION:

Wireless sensor networks have recently gained much attention in the sense that they can be readily deployed for many different types of missions. In particular, they are useful for the missions that are difficult for humans to carry out. For example, they are suitable for sensing dangerous natural phenomenon such as volcano eruption, biohazard monitoring, and forest fire detection. In addition to these hazardous applications, sensor networks can also be deployed for battle field surveillance, border monitoring, nuclear and chemical attack detection, intrusion detection, flood detection, weather forecasting, traffic surveillance and patient monitoring.

Sensor networks are deployed to sense, monitor, and report events of interest in a wide range of applications such as military, health care, and animal tracking. In many applications, such monitoring networks consist of energy constrained nodes that are expected to operate over an extended period of time, making energy efficient monitoring an important feature for unattended networks. In such scenarios, nodes are designed to transmit information only when a relevant event is detected (i.e., event-triggered transmission). Consequently, given the location of an event triggered node, the location of a real event reported by the node can be approximated within the node's sensing range. The locations of the combat vehicle at different time intervals can be revealed to an adversary observing nodes transmissions. There are three parameters that can be associated with an event detected and reported by a sensor node: the description of the event, the time of the event, and the location of the event. When sensor networks are deployed in untrustworthy environments, protecting the privacy of the three parameters that can be attributed to an event-triggered transmission becomes an important security feature in the design of wireless sensor networks

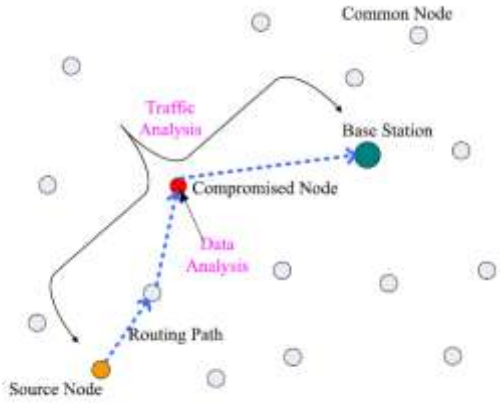


Fig. 2. Two scenarios for privacy attack in a WSN.

While transmitting the security of a message can be achieved via encryption primitives, hiding the timing and spatial information of reported events cannot be achieved via cryptographic means. Encrypting a message before transmission, for instance, can hide the context of the message from unauthorized observers, but the mere existence of the cipher text is indicative of information transmission. The source anonymity problem in wireless sensor networks is the problem of studying techniques that provide time and location privacy for events reported by sensor nodes. (Time and location privacy will be used interchangeably with source anonymity throughout the paper.) In the existing literature, the source anonymity problem has been addressed under two different types of adversaries, namely, local and global adversaries. A local adversary is defined to be an adversary having limited mobility

and partial view of the network traffic. Routing-based techniques have been shown to be effective in hiding the locations of reported events against local adversaries. A global adversary is defined to be an adversary with ability to monitor the traffic of the entire network. Against global adversaries, routing-based techniques are known to be ineffective in concealing location information in event-triggered transmission. This is due to the fact that, since a global adversary has full spatial view of the network, it can immediately detect the origin and time of the event-triggered transmission. The first step towards achieving source anonymity for sensor networks in the presence of global adversaries is to refrain from event-triggered transmissions. To do that, nodes are required to transmit fake messages even if there is no detection of events of interest (real events will be used to denote events of interest for the rest of the paper). When a real event occurs, its report can be embedded within the transmissions of fake messages. Thus, given an individual transmission, an observer cannot determine whether it is fake or real, assuming messages are encrypted. In the above approach, there is an implicit assumption of the use of a probabilistic distribution to schedule the transmission of fake messages.

However, the arrival distribution of real events is, in general, time-variant and unknown a priori. If nodes report real events as soon as they are detected(independently of the distribution of fake transmissions), given the knowledge of the fake transmission distribution statistical analysis can be used to identify outliers (real transmissions) with a probability higher than 1=2, as illustrated in Figure 1 (b)

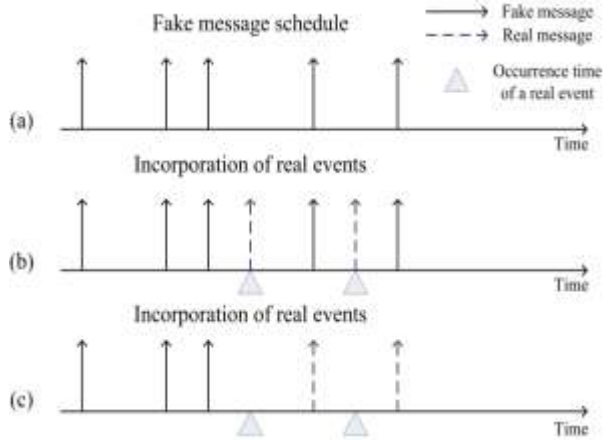


Fig.1. Ref [1] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran , 2010, (Fast Abstract). Different approaches for embedding the report of real events within a series of fake transmissions; (a)shows the pre-specified distribution of fake transmissions,(b) illustrates how real events are transmitted as soon as they are detected, (c) illustrates how nodes report real events instead of the next scheduled fake message.

In other words, transmitting real events as soon as they are detected does not provide source anonymity against statistical adversaries analyzing a series of fake and real transmissions.

One way to mitigate the above statistical analysis is illustrated in Figure 1(c). As opposed to transmitting real events as they occur, they can be transmitted instead of the next scheduled fake one. For example, consider programming sensor nodes to deterministically transmit a fake message every minute. If a real event occurs within a minute from the last transmission, its report must be delayed until exactly one minute has elapsed. This approach, however, introduces additional delay before a real event is reported. When real events have time-sensitive information, such delays might be unacceptable.

Reducing the delay of transmitting real events by adopting a more frequent scheduling algorithm is impractical for most sensor network applications since sensor nodes are battery powered and, in many applications, unchargeable. Therefore, a frequent transmission scheduling will drastically reduce the desired lifetime of the sensor network.

The Statistical Source Anonymity (SSA) problem in sensor networks is the study of techniques that prevent global adversaries from exposing source location by performing statistical analysis on nodes transmissions. Practical SSA solutions need to be designed to achieve their objective under two main constraints: minimizing delay and maximizing the lifetime of sensors' batteries.

II. MODEL ASSUMPTIONS

In this section, we describe the network and adversarial assumption that will be used in this paper.

2.1 Network Model

Communication is assumed to take place in a network of energy constrained sensor nodes. Nodes are deployed to sense events of interest and report them with minimum delay. Consequently, given the location of a certain node, the location of the reported event of interest can be approximated within the node's communication range at the time of transmission. When a node senses an event, it places information about the event in a message and broadcast an encrypted version of the message. To obscure the report of an event of interest, nodes are assumed to broadcast fake messages, even if no event of interest has been detected. Nodes are also assumed to be equipped with a semantically secure encryption algorithm, so that adversaries are unable to distinguish between the reports of events of interest and the fake transmissions by means of cryptographic tests.1 Furthermore, the network is assumed to be deployed in an unreachable environment and, therefore, the. conservation of nodes' energy is a design requirement

2.2 Adversarial Model

The adversarial model used in this paper is external, passive, and global. An external adversary is an

adversary who does not control any of the nodes in the network. As opposed to active adversaries injecting their own traffic or jamming the network, a passive adversary is only capable of observing the network traffic. A global adversary is an adversary who can monitor the traffic of the entire network and can determine the node responsible for the initial transmission reporting an event of interest. The adversary is assumed to know the locations of all nodes in the networks. The adversary is also assumed to know the distribution of fake message transmissions. Furthermore, the adversary is assumed capable of observing nodes transmissions over extended periods of times and performing sophisticated statistical analysis to compare the observed transmission with the known distribution of fake messages. The adversary, however, is not assumed able to break the security of the encryption algorithm and distinguish the report of event of interests via cryptographic tests.

III. PROPOSED FRAMEWORKS FOR STATISTICAL SOURCE OBSCURITY

In this section, source anonymity model for wireless sensor networks is being introduced. Intuitively, anonymity should be measured by the amount of information about the occurrence time and location of reported events an adversary can extract by monitoring the sensor network. The challenge, however, is to come up with an appropriate model that captures all possible sources of information leakage and a proper way of quantifying anonymity in different systems.

3.1 Interval Indistinguishability

Currently, statistical anonymity in sensor networks is modelled by the adversary's ability to distinguish between real and fake transmissions by means of statistical analysis. That is, given a series of transmissions of a certain node, the adversary must be unable to distinguish, with significant confidence, which transmission carries real information and which transmission is fake, regardless of the number of transmissions the adversary may observe. Consider now an adversary observing a sensor network over multiple time intervals. Assume that, during a given time interval, the adversary is able to notice a change in the statistical behaviour of transmission times of a certain node in the network. This distinguishable change in the transmission behaviour of the node can be indicative of the existence of real activities detected and reported by that node during that interval, even if the adversary was unable to distinguish between individual transmissions.

Consequently, in many applications, modelling source

Anonymity in sensor networks by the adversary's ability to distinguish between individual transmissions is insufficient to guarantee location privacy. It must be the case that an adversary monitoring the network over multiple time intervals, in which some intervals contain real event transmissions and the others do not, is unable to determine, with significant confidence, which of the intervals contain the real traffic. Formally, the notion of interval indistinguishability can be defined as follows.

Definition 1 (Interval Indistinguishability)[1]: Let IF denotes a time interval without any real event transmission (called the "fake interval" for the rest of the paper), and IR denotes a time interval with real event transmissions (called the "real interval" for the rest of the paper). The two time intervals are said to be statistically indistinguishable if the distributions of inter-transmission times during these two intervals cannot be distinguished with significant confidence.

3.2 Interval versus Event Indistinguishability

This section illustrates the relation between the traditional anonymity notion (i.e., individual event indistinguishability) and the proposed anonymity notion (i.e., interval indistinguishability). First, observe that as the length of intervals decreases, interval indistinguishability approaches event indistinguishability. If each interval consists of a single transmission, interval indistinguishability is equivalent to event indistinguishability. However, in the more general scenario, in which intervals contain more than a single transmission, interval indistinguishability implies indistinguishability of individual transmissions. To see this, assume a system satisfying interval indistinguishability but does not satisfy individual event indistinguishability. Since real and fake transmissions are distinguishable, given a fake interval and a real interval, the real interval can be identified as the one with the real transmission;

a contradiction to the hypothesis that the system satisfies interval indistinguishability. That is, if intervals are indistinguishable, then individual events within them must also be indistinguishable. In fact, the notion of interval indistinguishability is strictly stronger than the traditional notion individual event indistinguishability. That is, while interval indistinguishability implies individual indistinguishability, the converse is not true in general.

3.3 Mapping Statistical Source Anonymity to Binary Hypothesis Testing

In binary hypothesis testing, given two hypothesis, H0 and H1, and a data sample that belongs to one of the two hypotheses (e.g., a bit transmitted through a noisy communication channel), the goal is to decide to which hypothesis the data sample belongs. In the statistical strong anonymity problem under interval indistinguishability, given an interval of intertransmission times, the goal is to decide whether the interval is fake or real (i.e., consists of fake transmissions only or contains real transmissions).

That is, given two hypotheses (a real interval and a fake interval) and an observed data (an interval of inter-transmission times of a sensor node), the goal of the adversary is to determine to which hypothesis the

observed data belongs (i.e., whether the observed interval contains real event transmissions).

Remark 1: Although giving the adversary two intervals might seem too strong of an assumption, it is actually a practical one. To see this, note that the adversary can always observe multiple time intervals, two for instance. Then, all that is needed is to analyze these two observed intervals. If they are distinguishable, then it is likely that one of them is a real interval and the other is fake. Moreover, an adversary can discover the distribution of fake intervals by monitoring a node in the absence of real events. Then, all that is needed is to observe different time intervals. The more distinguishable a time interval from the known fake interval, the more likely it is to contain real events.

An adversary with the knowledge that a node is transitioning from one In either case, source anonymity can be breached. Real interval and the other is fake. Moreover, an adversary can discover the distribution of fake intervals by monitoring a node in the absence of real events.

IV. STATISTICAL GOODNESS OF FIT TESTS AND THE SSA PROBLEM

In the literature, statistical source anonymity is shown to be achieved via the use of statistical goodness of fit tests [11]. In this section, we describe the current use of statistical goodness of fit tests in designing anonymous sensor networks.

3.4 SSA Solutions Based on Statistical Goodness of Fit Tests

The statistical goodness of fit of an observed data describes how well the data fits a given statistical model. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the statistical model in question. Such measures can be used, for example, to test for normality of residuals, to test whether two samples are drawn from identical distributions, or to test whether outcome frequencies follow a specified distribution. Examples of well studied goodness of fit tests include, but are not limited to, the Anderson-Darling (A-D) test [26], the Kolmogorov-Smirnov (K-S) test [27], the Jarque-Bera (J-B) test 28].

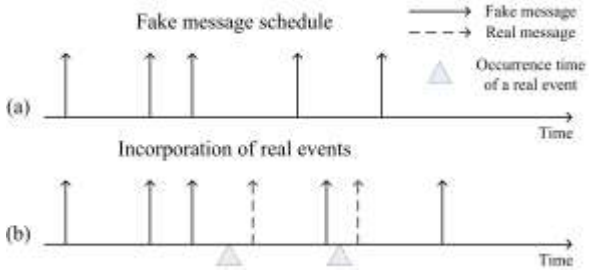
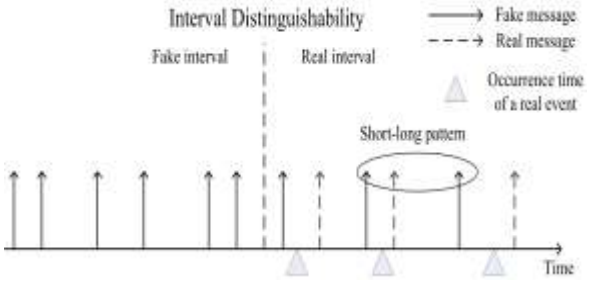


Fig. 3. Ref [1] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran , 2010, (Fast Abstract).

An illustration of solutions based on statistical Goodness of fit tests. Nodes transmit fake messages According to a pre-specified probabilistic distribution and maintain a sliding window of inter-transmission times. When a real event occurs, it is transmitted as soon as possible under the condition that the samples in the sliding window maintain the designed distribution. The transmission following the real transmission is delayed to maintain the mean of the distribution of inter-transmission times in the sliding window.



Fake events serve the same purpose they serve in algorithm AR, that is, they are used to hide the existence of real transmissions. Since there are no real events in fake intervals, however An illustration of interval distinguishability in the current state-of-the-art solutions based on statistical goodness of fit tests. Real events are transmitted sooner than what is determined by the probabilistic distribution, while the transmission following the real event is later than what is determined by the probabilistic distribution to fix the mean of the pre-defined distribution.

4.3 Correlation Analysis of SSA Solutions Based on Statistical Goodness of Fit Tests

The interpretation of the analysis is that each bit in a binary code representing a fake interval is independent of the all other bits, while bits in a binary code representing a real interval are correlated. More specifically, a binary code representing a real interval is likely to have more ‘01’

4.4 Nuisance Parameters

In statistical decision theory, the term “nuisance parameters” refers to information that is not needed for hypothesis testing and, further, can preclude a more accurate decision making. When performing hypothesis testing of data with nuisance parameters, it is desired (even necessary in some scenarios) to find an appropriate transformation of the data that removes or minimizes the effect of the nuisance information

before performing the hypothesis testing .

V. THE PROPOSED APPROACH

To improve anonymity, literature suggests introducing the same correlation of inter-transmission times during real intervals to inter-transmission times during fake intervals. That is, let the transmission procedure consists of two different algorithms:

AR and AF. In the presence of real events (i.e., in real intervals), algorithm AR is implemented. In the absence of real events (i.e., in fake intervals), algorithm AF is implemented. Algorithm AR is the same as the algorithm. In algorithm AF , the nodes generates two sets of events independently of each other: "dummy events", dummy events are generated to be handled as if they are real events. That is, dummy events are generated independently of fake messages and, upon their generation, their transmission times are determined according to the algorithm. The purpose of this procedure is to introduce the same correlation of real intervals into fake intervals. That is, not only the two sequences of intertransmission times will be statistically indistinguishable means of statistical goodness of fit tests [11] but also the binary codes representing fake and real intervals will have the same statistical behaviour**.**

The Anderson-Darlington test is used in both algorithms, AR and AF, to determine the transmission times of real events and dummy events, respectively.

5.1 Experimental Results and Anonymity Interpretations

This work is about developing A statistical framework using Java & Mysql. Since java is platform independent the same is used instead of traditional NS2 simulator. Using AD test in our algorithm AR and AF we have generated fake path by means of random values and real path will remain hidden. File will be encrypted and will be sent in various paths towards destination. Since message will be sent in various paths simultaneously adversaries will get confused. They will try to get the message and if they are on fake path they will fail in getting the message within a limited timeframe and message will reach destination successfully. Then adversary might try the other path thinking that it is real. Once message has been reached it will be useless decrypting it. Even if Adversaries are found to be on Real Path or if adversaries succeed in getting the real path they won't get the message since the code required to decrypt it, is with the recipient only and no one else. Message Packet Flow is encrypted in the binary format such that correlations analysis will fail to distinguish between real and fake path. Since the Message is sent at the same time via different paths events are indistinguishable resulting in anonymity of the node from where the message is come.

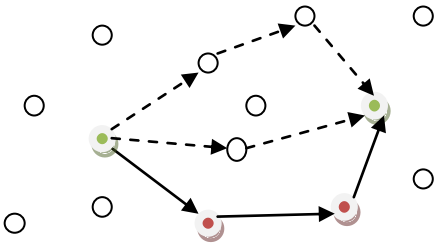


Fig4: Nodes in green color are Source and Destination respectively. All dotted lines are fake paths and the solid line indicates the real path from source to destination.

VI. PERFORMANCE

This statistical Framework will help to manually delay the message and to change the weights of the links connecting sensor nodes. It can be used from any location and on any platform. Source obscurity can be demonstrated successfully using this framework.

VII. CONCLUSION

The Source Obscurity can be achieved using the given Framework and Binary hypothesis concept is being implemented.

This Statistical Framework can be improved further for a moving target using more efficient Cryptographic techniques.

REFERENCES

[1] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "On Source Anonymity in Wireless Sensor Networks," in Proceedings of the 40th IEEE/IFIP International Conference on Dependable Systems and Networks–DSN'10. IEEE Computer Society, 2010, (Fast Abstract).

[2] "Statistical Framework for Source Anonymity in Sensor Networks," in Proceedings of the 53rd IEEE Global Communications Conference–GLOBECOM'10. IEEE Communications Society, 2010..

[3] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards Statistically Strong Source Anonymity for Sensor Networks," in Proceedings of the 27th Conference on Computer Communications–INFOCOM'08. IEEE Communications Society, 2008, pp. 466–474.

[4] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energyconstrained sensor network routing," in roceedings of the 2nd ACM workshop on Security of Ad hoc and Sensor Networks–SASN'04. ACM,2004, pp. 88–93.

[5] Y. Li and J. Ren, "Source-location privacy through dynamic routing in wireless sensor networks," in Proceedings of the 29th Conference on Computer Communications – INFOCOM'10. IEEE Communications Society, 2010, pp. 1–9.

[6] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks," in Proceedings of the 20th IEEE International Parallel & Distributed Processing Symposium– IPDPS'06. IEEE Computer Society, 2006, pp. 1–8.

[7] B. Hoh and M. Gruteser, "Protecting Location Privacy Through Path Confusion," in

Proceedings of the 1st IEEE/CreatNet International Conference on Security and Privacy for Emerging Areas in Communications Networks–SecureComm'05. IEEE Communications Society, 2005, pp.194–205.

[8] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," in Proceedings of the first ACM conference on Wireless network security–WiSec'08. ACM, 2008, pp. 77–88.

[9] N. Li, N. Zhang, S. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: A state-of-the-art survey," Elsevier Journal on Ad Hoc Networks, vol. 7, no. 8, pp. 1501–1514, 2009.

[10] Y. Li and J. Ren, "Preserving source-location privacy in wireless sensor networks," in Proceedings of the 6th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks–SECON'09. IEEE Communications Society, 2009, pp. 493–501.

[11] S. Goldwasser and S. Micali, "Probabilistic encryption," Journal of Computer and System Sciences, vol. 28, no. 2, pp. 270–299, 1984

[12] M. Stephens, "EDF statistics for goodness of fit and some comparisons," Journal of the American Statistical Association, vol. 69, no. 347, pp.730–737, 1974.