Statistical Framework for Source Anonymity in Sensor Networks

Basel Alomair*, Andrew Clark*, Jorge Cuellar†, and Radha Poovendran*
*Network Security Lab (NSL),University of Washington, Seattle, Washington
†Siemens Corporate Technology, München, Germany
Email: {alomair,awclark,rp3}@uw.edu, jorge.cuellar@siemens.com

***Abstract*—In this work, we investigate the security of anonymous wireless sensor networks. To lay down the foundations of a formal framework, we develop a new model for analyzing and evaluating anonymity in sensor networks. The novelty of the proposed model is twofold: first, it introduces the notion of "interval indistinguishability" that is stronger than existing notions; second, it provides a quantitative measure to evaluate anonymity in sensor networks. The significance of the proposed model is that it captures a source of information leakage that cannot be captured using existing models. By analyzing current anonymous designs under the proposed model, we expose the source of information leakage that is undetectable by existing models and quantify the anonymity of current designs. Finally, we show how the proposed model can lead to a general and intuitive direction for improving the anonymity of current designs.**

I. Introduction and Related Work

In sensor networks, nodes are deployed to capture and report relevant events. A topic that has been drawing increasing research attention in wireless sensor networks is source location privacy [1]–[7]. (Source anonymity and source location privacy will be used synonymously for the rest of the paper.) Given the adversary's knowledge of the locations of sensor nodes in a network, determining the individual nodes reporting the occurrence of real events can translate to the exposure of the location of the real events themselves. Applications in which hiding the occurrence of real events can be critical include, but are not limited to, the deployment of sensor nodes in battlefields and the classic Panda-Hunter Game [1], [2], [5], [6].

In such applications, where source location privacy is of critical importance, special attention must be paid to the design of the node transmission algorithm so that monitoring sensor nodes does not reveal critical source information. One of the major challenges for the source anonymity problem is that it cannot be solved using traditional cryptographic primitives. Encrypting nodes' transmissions, for instance, can hide the contents of plaintext messages, but the mere existence of ciphertexts is indicative of information transmission.

In the presence of a global adversary, who is able to monitor the traffic of the entire network, routing-based solutions has been shown to leak private source information [4]. An intuitive approach to report a real event without revealing, to a global adversary, its location information is to program nodes to transmit *fake messages* even if there are no real events to be reported [4]. When real events occur, they can be embedded within the transmissions of fake messages. This intuitive approach, however, does not completely solve the location privacy problem. When fake transmissions are scheduled according to certain probabilistic distributions, statistical analysis can be used to distinguish between real and fake transmissions if real events are transmitted as they arrive. This intuitive approach is illustrated in Figure 1.

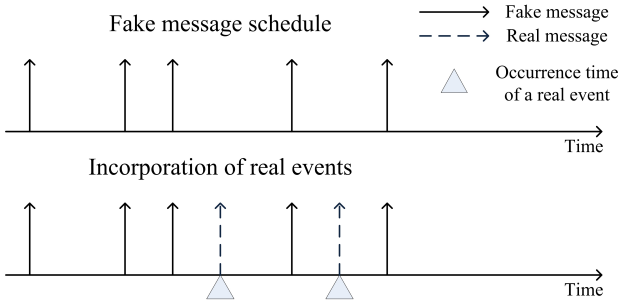

Fig. 1. An illustration of the intuitive approach. The node is programmed to transmit fake messages so that real events are hidden within the fake transmissions.

By realizing the problem with the intuitive approach, the solution becomes trivial. As opposed to transmitting real events as they occur, they can be transmitted instead of the next scheduled fake one. For example, sensor nodes can be programmed to transmit an encrypted message every minute. If there is no event to report, the node transmits a fake message. If a real event occurs within a minute from the last transmission, it must be delayed until exactly one minute after the last transmission has elapsed. This algorithm, trivially, provides source anonymity since an adversary monitoring a node will observe one transmission every minute and, assuming the semantic security of the underlying encryption, the adversary has no means of distinguishing between fake and real events. Figure 2 depicts an example of this trivial solution.

The trivial solution, however, has a major drawback: reporting real events must be delayed until the next scheduled transmission. (In the above example, the average latency of transmitting real events will be half a minute.) When real events have time-sensitive information, this latency might be unacceptable. Reducing the latency of transmitting real events by adopting a more frequent scheduling algorithm is impractical for most sensor network applications. This is mainly because sensor nodes are battery powered and, in many applications, are unchargeable. Consequently, a more

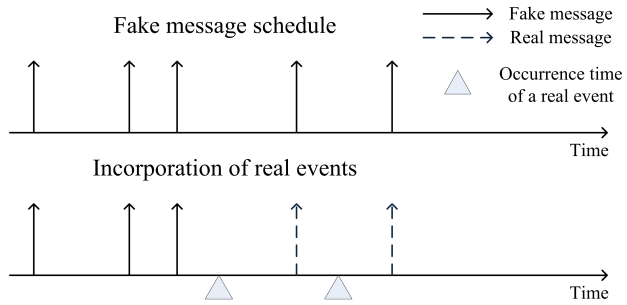

Fig. 2. An illustration of the trivial solution. Real event must be delayed until the next scheduled fake transmission.

frequent scheduling algorithm can exhaust nodes' batteries rather quickly, rendering sensor nodes useless.

Furthermore, a transmission scheduling based on any pre-specified probabilistic distribution, not necessarily deterministic as in the above example, will suffer the same problem discussed above: slower rates lead to longer latencies and faster rates lead to shorter battery lives. Consequently, practical solutions are designed to achieve the objective of source anonymity under two main constraints: minimizing latency and maximizing the lifetime of sensors' batteries. To make things even more complex, the arrival rate and distribution of real events can be time varying and unknown in advance. Clearly, in the trivial solution, no pre-specified probabilistic distribution for fake transmissions can satisfy both constraints for arbitrary time-variant distribution of real event arrivals.

The current state of the art in designing anonymous sensor networks works as follows. In the absence of real events, nodes are programmed to transmit independent identically distributed (iid) fake messages according to a certain distribution with a certain rate. However, unlike the trivial solution, real events are transmitted as soon as possible (earlier than the next pre-scheduled fake transmissions) under the following condition: the distribution of the entire message transmissions (fake and real) of each node is "statistically" similar to the transmission of only fake messages. (Statistical similarity is achieved via the use of statistical goodness of fit tests that determine if a sequence of data samples follows a certain probabilistic distribution.) Consequently, to a global adversary monitoring the network, the time between any two transmissions (real or fake) will follow the same distribution of fake messages only. The current consensus is that this approach provides dependable solutions for the source anonymity problem in wireless sensor networks [5]–[9].

In this paper, we take a closer look at the current state-of-the-art in designing anonymous sensor networks. The driving motive behind this work is the key observation that, although an adversary might not be able to distinguish between real and fake transmissions, there still exists a source of information leakage that can affect the security of such designs. The inability to detect the source of information leakage in the current approach is not a result of false statements claimed in previous proposals; the lack of a formal framework that properly models anonymity in wireless sensor networks is the main reason for the inability to detect such a vulnerability. The main purpose of this work is to provide such a framework.

A. Our Contributions

We summarize our contributions by the following points.

- We detect a source of information leakage in the current designs that can undermine their anonymity.
- We introduce the new notion of interval indistinguishability to analyze anonymity in wireless sensor networks. The new notion is stronger than existing notions and captures the source of information leakage that is undetectable by existing notions.
- We propose a quantitative measure to evaluate anonymity in sensor networks.
- We analyze, both analytically and via simulation, the current state-of-the-art in designing anonymous sensor networks and quantify the amount of information leakage when the current approach is analyzed under the proposed model.
- Based on our model, we discuss a new direction to enhance the anonymity of the current state-of-the-art.

II. Modeling Anonymity

In this section we introduce our anonymity model for wireless sensor networks. Intuitively, anonymity should be measured by the amount of information about sources' locations an adversary can infer 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.

A. Network and Adversarial Models

We assume that communications take place in a network of energy constrained sensor nodes. That is, nodes are assumed to be powered with unchargeable batteries, thus, conserving nodes energy is a design requirement. Nodes are also equipped with a semantically secure encryption algorithm, so that adversaries are unable to distinguish between real and fake transmission by means of cryptographic tests. When a node detects an event, it places information about the event in a message and broadcasts an encrypted version of the message.

Our adversary is similar to the one considered in [4], [5], in that it is *external*, *passive*, and *global*. By external, we mean that the adversary does not control any of the nodes in the network and also has no control over the real event process. By passive, we mean that the adversary is capable of eavesdropping on the network, active attacks are not considered. By global, we mean that the adversary can simultaneously monitor the activity of all nodes in the network.

As opposed to a global adversary, a local adversary is only capable of eavesdropping over a small area, typically the area surrounding the base station, and attempts to determine the source of traffic by examining the packet routing information or trying to follow the packets back to their source. Protocols that attempt to disguise the source of traffic through routing, while highly secure against local adversaries, do not defend against global adversaries [4].

We also assume that the adversary is capable of storing a large amount of message traffic data and performing complex statistical tests. Furthermore, the adversary is assumed to know the distribution of fake message transmissions. The only information unknown to the adversary is the timing when real events occur.

B. Event Indistinguishability (EI)

Currently, anonymity in sensor networks is modeled by the adversary's ability to distinguish between individual real and fake transmissions by means of statistical tests. That is, given a series of nodes' transmissions, the adversary should not be able to distinguish, with significant confidence, which transmission carries real information and which transmissions is fake.

Consider, however, an adversary observing the sensor network over multiple time intervals, without being able to distinguish between individual fake and real nodes' transmissions. Assume, further, that during a certain time interval the adversary is able to notice a change in the statistical behavior of transmission times of a certain node in the network. This distinguishable change in transmission behavior can be indicative of the existence of real activities reported by that node, even though the adversary was unable to distinguish between individual transmissions.

For example, consider a sensor network deployed in a battlefield. For a certain time interval, there were no activities in the vicinity of a sensor node the enemy is monitoring. Hence, by design, the node has been transmitting fake messages for the duration of that time interval. Assume now that a moving platoon is in the vicinity of this node and the node started to report location information about the moving platoon. It is unnecessary to distinguish between individual transmissions to infer the existence of the moving platoon. That is, the ability to distinguish between the time interval when no real activities are reported and the time interval when the platoon is in the vicinity of the sensor node is sufficient to infer private location information.

Consequently, in many applications, modeling source anonymity in sensor networks by the adversary's ability to distinguish between *individual transmissions* is insufficient to guarantee location privacy. This fact calls for a stronger model to properly address source anonymity in sensor networks. Before we proceed, we formally define the currently adopted notion to model anonymity, namely, event indistinguishability.

Definition 1 (Event Indistinguishability - 'EI'*): Events reported by sensor nodes are said to be indistinguishable if the inter-transmission times between them cannot be distinguished with significant confidence.*

C. Interval Indistinguishability (II)

The main goal of source location privacy is to hide the existence of real events. This implies that, an adversary observing a sensor node during different time intervals, in which some of the intervals include the transmission of real events and the others do not, must not be able to determine with significant confidence which of the intervals contain real traffic. This leads to the notion of interval indistinguishability that will be essential for our anonymity formalization.

Definition 2 (Interval Indistinguishability - 'II'*): Let I_F denotes a time interval with only fake event transmissions (call it the "fake interval"), and I_R denotes a time interval with real event transmissions (call it the "real interval"). 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.*

To model interval indistinguishability, we propose the following game between a challenger $\mathcal{C}$ (the system designer) and an adversary $\mathcal{A}$.

Game 1 (Modeling Interval Indistinguishability):

1) *$\mathcal{C}$ draws a bit $b \in \{0, 1\}$ uniformly at random.*
2) *$\mathcal{C}$ chooses two intervals I_0 and I_1, in which I_b is a real interval and the other one is fake.*
3) *$\mathcal{C}$ gives I_0 and I_1 to $\mathcal{A}$.*
4) *$\mathcal{A}$ makes any statistical test of her choice on I_0 and I_1 and outputs a bit b'.*
5) *If $b' = b$, $\mathcal{A}$ wins the game.*

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 observe two time intervals, for example. If the two time intervals are distinguishable, then it is likely that one of them is a real interval and the other is fake. Moreover, since nodes are not tamper-resistant in most application, an adversary capturing a node in the network can discover the distribution of fake intervals. Even if nodes are tamper-resistant, 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. Therefore, Game 1 is suitable to analyze practical systems.

With Definition 2 and Game 1, we aim to find a security measure that can formally quantify the anonymity of different designs. Let σ denote an adversarial strategy for attacking the system. Let $\Pr[b' = b]_\sigma$ be the adversary's probability of winning Game 1 using strategy σ. We quantify the anonymity of a sensor network against the strategy σ by

$$\Lambda_\sigma := 1 - 2\Big(\Pr[b' = b]_\sigma - 0.5\Big). \tag{1}$$

In the best case scenario, the adversary's strategy is a pure random guess, leading to $\Pr[b' = b]_\sigma = 1/2$ and $\Lambda_\sigma = 1$ (absolute anonymity). In the worst case, the adversary will have a strategy with $\Pr[b' = b]_\sigma = 1$ leading to $\Lambda_\sigma = 0$ (no anonymity). Any intelligent strategy will result in a probability of winning the game belonging to the interval $[0.5, 1]$, leading to an anonymity measure in the interval $[0, 1]$.

Let Σ be the set of all possible adversarial strategies to attack the system. Then, the anonymity of the system is:

$$\Lambda := \min_{\sigma \in \Sigma} \Lambda_\sigma. \tag{2}$$

Given Definitions 1 and 2, the relation between event indistinguishability (EI) and interval indistinguishability (II) is stated as follows.

TABLE I
A LIST OF USED TERMS AND NOTATIONS.

$\boldsymbol{E_i}$	The random variable representing the event reported in the i^{th} transmission
$\boldsymbol{X_i}$	The random variable representing the inter-transmission time between the i^{th} and the $i+1^{st}$ transmissions
I_F	A fake interval: an interval consisting of fake events only
I_R	A real interval: an interval containing some real event transmissions
short inter-transmission times	inter-transmission times that are shorter than the mean of the pre-defined distribution
long inter-transmission times	inter-transmission times that are longer than the mean of the pre-defined distribution
short-long pattern	a short inter-transmission time followed by a long inter-transmission time

Lemma 1: II $\Rightarrow$ EI.

Proof: Assume a system satisfying interval indistinguishability but does not satisfy event indistinguishability. Then, real and fake transmissions are distinguishable, by assumption. Therefore, 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, and the lemma follows. ■

To show that the proposed notion is stronger than the current one, it remains to show that event indistinguishability does not imply interval indistinguishability (II$\not\Leftarrow$ EI). Section III proves this fact by providing a counter example.

With the above definition of interval indistinguishability, we introduce the notion of Λ-anonymity in sensor networks.

Definition 3 (Λ-anonymity): A wireless sensor network is said to be Λ-anonymous if it satisfies two conditions

1) *the beginning and the end of an interval cannot be distinguishable,*
2) *the anonymity of the system, as defined in equation (2), is at least Λ.*

The first condition in Definition 3 is necessary to ensure that there is no distinguishable transition region between intervals. If such a transition exists, it can lead to anonymity breach.

III. Analysis of EI-based Approaches

In this section we analyze, using our proposed model, systems that were shown to be secure under event indistinguishability; i.e., EI-based systems. First, we provide theoretical analysis showing that real and fake intervals can be statistically distinguishable. Then, we simulate an existing scheme to show that the theoretical analysis is also practical. We start by describing the current state-of-the-art, as first proposed in [5].

A. EI-based Approaches

Nodes are designed to transmit fake messages according to a pre-specified distribution. Furthermore, nodes store a sliding window of times between consecutive transmissions, say $\boldsymbol{X_i}, \boldsymbol{X_{i+1}}, \cdots, \boldsymbol{X_{k+i-1}}$, where $\boldsymbol{X_j}$ is the random variable representing the time between the j^{th} and the $j+1^{st}$ transmissions and k is the length of the sliding window. When a real event occurs, its transmission time, represented by $\boldsymbol{X_{k+i}}$, is defined to be the smallest value such that the sequence $\boldsymbol{X_i}, \cdots, \boldsymbol{X_{k+i}}$ passes some statistical goodness of fit tests. That is, an adversary observing the sequence of inter-transmission times will observe a sequence that is statistically indistinguishable from an iid sequence of random variables with the pre-specified distribution of fake transmissions.

However, by continuing in this fashion, the mean will skew since nodes always favor shorter times to transmit real events. To adjust the mean, the next transmission following a real one, $\boldsymbol{X_{k+i+1}}$ in this example, will be purposely delayed. Again, the delay is determined so that the sequence in the sliding window satisfies some statistical goodness of fit test. Consequently, as shown in [5], an adversary observing the sensor node cannot differentiate between real and fake transmissions.

B. Theoretical Interval Distinguishability

As discussed in Section II, when an adversary can distinguish between real and fake intervals, source location can be exposed, even if the adversary cannot distinguish between individual transmissions. In what follows, we give theoretical analysis of interval indistinguishability in EI-based systems.[1]

Let $\boldsymbol{X_i}$ be the random variable representing the time between the i^{th} and the $i+1^{st}$ transmissions and let $E[\boldsymbol{X_i}] = \mu$. We now examine two intervals, a fake interval and a real one.

1) Fake Interval (I_F): In fake intervals, inter-transmission times are iid random variables. That is, the $\boldsymbol{X_i}$'s are iid's with mean μ. Therefore, during any fake interval I_F, for any $\boldsymbol{X_{i-1}}, \boldsymbol{X_i} \in I_F$,

$$E\big[\boldsymbol{X_i} \mid \boldsymbol{X_{i-1}} < \mu\big] = \mu. \tag{3}$$

2) Real Interval (I_R): Let $\boldsymbol{E_i}$ be the random variable representing the event reported in the i^{th} transmission. Then, $\boldsymbol{E_i}$ can take the values R and F, where R denotes a real event and F denotes a fake one. Since in general scenarios the distribution of inter-arrival times of real events can be varying and unknown beforehand, we will assume that $\boldsymbol{E_i}$ can take the values R and F with arbitrary probabilities.

Recall that the time between the transmission of a real event and its preceding fake one is usually shorter than the mean μ by design (to reduce latency). Recall further that the time between the transmission of a real event and its successive one is usually longer than μ by design (to adjust the ensemble mean). That is, during any real interval I_R, for any $\boldsymbol{X_{i-1}}, \boldsymbol{X_i} \in I_R$,

$$E\big[\boldsymbol{X_i} \mid \boldsymbol{X_{i-1}} < \mu, \boldsymbol{E_i} = R\big] > \mu, \tag{4}$$

and,

$$E\big[\boldsymbol{X_i} \mid \boldsymbol{X_{i-1}} < \mu, \boldsymbol{E_i} = F\big] = \mu, \tag{5}$$

[1]The vulnerability of [5] to correlation attacks has been informally discussed in a fast abstract at the DSN 2010 conference [10].

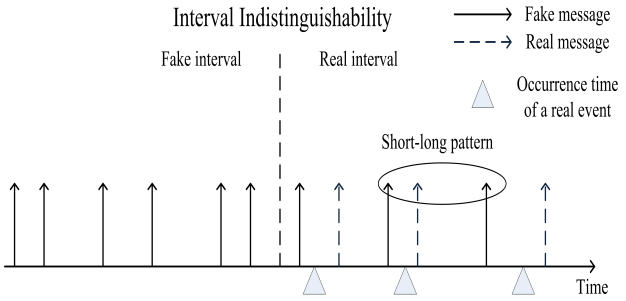

Fig. 3. An illustration of interval distinguishability in the current approach.

by design. Using equations (4) and (5) we get,

$$\begin{aligned} &E\big[\boldsymbol{X_i} \mid \boldsymbol{X_{i-1}} < \mu\big] \\ &\quad = E\big[\boldsymbol{X_i} \mid \boldsymbol{X_{i-1}} < \mu, \boldsymbol{E_i} = R\big] \cdot \Pr[\boldsymbol{E_i} = R] \\ &\qquad + E\big[\boldsymbol{X_i} \mid \boldsymbol{X_{i-1}} < \mu, \boldsymbol{E_i} = F\big] \cdot \Pr[\boldsymbol{E_i} = F] \qquad (6) \\ &\quad > \mu \cdot \Pr[\boldsymbol{E_i} = R] + \mu \cdot \Pr[\boldsymbol{E_i} = F] \qquad (7) \\ &\quad = \mu. \qquad (8) \end{aligned}$$

Therefore, by equations (3) and (8), shorter inter-transmission times followed by longer inter-transmission times are most likely to occur in real intervals than fake intervals. This suggests the following strategy to distinguish between fake and real intervals: given two time intervals I_0 and I_1, in which one of them is real and the other one is fake, the adversary counts the number of short followed by long inter-transmission times, simply called *short-long patterns* for the remainder of the paper. (An inter-transmission time is said to be short if its length is shorter than the mean μ, and is said to be long if it is longer than μ.) The interval that has more counts of short-long patterns is the real interval. Figure 3 illustrates the pattern of short-long inter-transmission times.

We emphasize that the notion of short-long patterns is merely a way of representing the class of correlation tests to distinguish between real and fake intervals. That is, one can use other correlation tests to distinguish between real and fake intervals in EI-based approaches. Short-long correlation patterns are chosen since they are intuitive and easy to analyze. After all, all that is needed to show that event indistinguishability is insufficient to model anonymity is a single successful attack that cannot be captured under event indistinguishability.

C. Case Study

So far, it has been shown that real and fake intervals are theoretically distinguishable via correlation tests (for example, the number of short-long patterns). However, it remains to show that the discrepancy in the number of short-long correlation patterns is large enough to substantially improve the adversary's probability of distinguishing between real and fake intervals, since this cannot be inferred from the analysis.

In what follows, we study the scheme appeared in [5], an instance of the EI-based approachs, and evaluate its anonymity using the proposed model. In the scheme of [5], the Anderson-Darling (A-D) goodness of fit test is used to determine the time for transmitting real events. Similarly, the A-D test is also used to implement the mean recovery algorithm. The authors of [5] used different statistical tests, such as the Kolmogorov-Smirnov (K-S) test, to show that their design satisfies event indistinguishability.[2]

1) Experimental Setup: For a reliable analysis of [5], we use the same parameters appeared in their paper. The inter-transmission times between fake transmissions are iid exponentials with mean 20 seconds. Real events arrive according to a Poisson Arrival process with mean $1/20$. The two parameters of the A-D test are the significance level of the test and the allowed deviation from the mean which are set to 0.05 and 0.1, respectively.

The experiment was run for $10,000$ independent trials. Each trial consists of two intervals, a real one I_R and a fake one I_F. Every trial starts with a "warmup" period, where 200 iid exponential random variables with mean 20 are drawn to constitute a backlog to be used in the A-D goodness of fit test. Then real events start arriving and they are transmitted according to procedure described earlier (please refer to [5] for detailed description and algorithms). Each real interval consists of 50 real events. After the 50^{th} real event has been transmitted, the fake interval starts for the same amount of time the real interval lasted.

2) Simulation Results: After running the above experiment for $10,000$ independent trials, and comparing the number of short-long patterns in fake and real intervals for each trial, the following results were found. Out of the $10,000$ trials, real intervals have *more* short-long patterns than fake intervals in $6,818$ trials; real intervals have *less* short-long patterns than fake intervals in $2,076$ trials; and real intervals have the same number of short-long patterns as fake intervals in $1,106$ trials.

3) Λ-anonymity Interpretation: Recall that, by equations (3) and (8), a short-long pattern is most likely to occur in real intervals than fake ones. Consequently, real intervals are likely to have more short-long patterns than fake intervals. Indeed, our simulation results agree with equations (3) and (8).

Consider Game 1 for analyzing interval indistinguishability. Given two intervals I_0 and I_1 at which one of them is real and one is fake, let the adversary's strategy for deciding which is which is as follows. Count the number of short-long patterns in each interval. If both intervals have the same number of short-long patterns, the adversary decides randomly. If one interval has more short-long patterns than the other, the adversary chooses it as the real interval. With this strategy, given the simulation results provided above, the adversary's probability of correctly identifying real intervals, without resorting to complicated statistical tests, is 0.737. That is, the anonymity of the system is at most $\Lambda = 0.526$.

IV. New Direction for Improved Anonymity

In this section, we discuss a promising direction to improve anonymity against correlation attacks, and give an example illustrating how to apply it to the scheme of [5].

[2]The A-D and K-S tests are popular goodness of fit tests that, given a sequence of data samples and a desired degree of accuracy, determine whether the samples follow a certain probabilistic distribution with certain parameters, within the specified degree of accuracy.

TABLE II

A QUANTITATIVE COMPARISON OF THE THREE SCHEMES, THE EI-BASED APPROACH OF [5], OUR II-BASED MODIFICATION OF [5], AND THE TRIVIAL SOLUTION OF SENDING REAL EVENTS INSTEAD OF THEIR SUCCESSIVE SCHEDULED FAKE TRANSMISSIONS. $I_R > I_F$ DENOTES MORE SHORT-LONG PATTERNS IN REAL INTERVALS, $I_R < I_F$ DENOTES MORE SHORT-LONG PATTERNS IN FAKE INTERVALS, WHILE $I_R = I_F$ DENOTES EQUAL SHORT-LONG PATTERNS IN REAL AND FAKE INTERVALS. THE SIMULATION RESULTS ARE OBTAINED FROM $10,000$ INDEPENDENT TRIALS.

	$I_R > I_F$	$I_R < I_F$	$I_R = I_F$	Anonymity bound
EI-based approach	$6,818$	$2,076$	$1,106$	0.526
Our II-based approach	$4,566$	$4,272$	$1,162$	0.971
Trivial solution	$4,385$	$4,318$	$1,297$	0.993

As can be seen from the analysis in Section III, inter-transmission times during fake intervals are iid's, while inter-transmission times during real intervals are neither independent nor identically distributed. In theory, the only way to guarantee that a sequence of random variables is statistically indistinguishable from a given iid sequence is to generate it as an iid sequence with the same distribution.

The notion of interval indistinguishability, suggests a different approach for the design of anonymous sensor networks. Observe that Definition 2 of interval indistinguishability does not impose any requirements, such as iid, on the distribution of inter-transmission times during fake intervals. Therefore, designing fake intervals with the distribution that is easiest to emulate during real intervals is the most logical solution. This idea opens the door for more solutions as it gives more flexibility for system designers.

To improve the anonymity of EI-based approaches against correlation attacks, we suggest introducing the same correlation of inter-transmission times during real intervals to inter-transmission times during fake intervals. In the scheme of [5], consider the generation of "dummy events" during fake intervals that are to be handled as if they are real events. That is, dummy events are generated independently of fake messages and, upon their arrival, their transmission times are determined according to the used statistical goodness of fit test. The purpose of this procedure is to introduce the same correlation into fake intervals.

To test our approach, we ran the same simulation of Section III with one major difference. To make fake intervals possess the same correlation of real intervals, we introduced *dummy events* in fake intervals. Dummy events were generated according to iid Gaussian inter-arrival times with mean 10 seconds and a variance of 150. Note the distinction between fake messages and dummy events. Fake messages are the ones transmitted to hide the existence of real transmissions, while dummy events are the ones generated, during fake intervals only, to resemble the existence of real events. Note also that the inter-arrival distribution of dummy events is purposely different than the inter-arrival distribution of real events to count for the general case of unknown distribution of real events inter-arrivals. The A-D test is used to determine the transmission times of dummy events, just as if they were real events.

Table II summarizes our simulation results. Observe the improvement in anonymity against correlation attacks in our modified version. We emphasize, however, that this is not meant to be a complete solution for anonymous systems. Its only purpose is to exemplify the idea of introducing the same correlation patterns to fake intervals in the scheme of [5]. In fact, with this approach, we believe interval indistinguishability can be achieved without resorting to computationally cumbersome statistical tests. This will be the main focus of future research.

V. CONCLUSION AND FUTURE WORK

In this paper, we provided a statistical framework for modeling, analyzing, and evaluating anonymity in sensor networks. We introduced the notion of interval indistinguishability, proved that it implies the currently adopted model (event indistinguishability), and showed that it captures the source of information leakage that was not captured by event indistinguishability (correlation tests). We analyzed an EI-based approach, which was shown to provide anonymity under event indistinguishability, and quantified its information leakage when analyzed under our proposed model. Finally, we proposed a modification to existing solutions to improve their anonymity against correlation attacks.

Future extensions to this work include taking advantage of the key point that fake intervals are not restricted to have iid inter-transmission times to design an efficient system that satisfies the notion of interval indistinguishability, without resorting to computationally cumbersome statistical tests.

REFERENCES

[1] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing Source-Location Privacy in Sensor Network Routing," *ICDCS 2005*.

[2] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *SASN*, 2004.

[3] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks," in *IPDPS*, 2006.

[4] K. Mehta, D. Liu, and M. Wright, "Location Privacy in Sensor Networks Against a Global Eavesdropper," in *ICNP*, 2007.

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

[6] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," in *WiSec*, 2008.

[7] N. Li, N. Zhang, S. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: A state-of-the-art survey," *Ad Hoc Networks*, 2009.

[8] H. Wang, B. Sheng, and Q. Li, "Privacy-aware routing in sensor networks," *Computer Networks*, 2009.

[9] M. Shao, W. Hu, S. Zhu, G. Cao, S. Krishnamurthy, and T. La Porta, "Cross-layer Enhanced Source Location Privacy in Sensor Networks," *SECON*, 2009.

[10] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "On Source Anonymity in Wireless Sensor Networks," DSN, 2010, (Fast Abstract).