

# On Source Anonymity in Wireless 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

## 1 Introduction

Preserving source location privacy is becoming one of the most interesting problems in wireless sensor networks. In a variety of real life applications, such as the deployment of sensor nodes in battlefields, the locations of events monitored by the network are required to remain anonymous. Given the knowledge of the network topology, however, an adversary can expose the locations of such events by determining the individual nodes reporting them.

When 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 private 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 [1], it was shown that routing-based techniques, although can provide source anonymity against local adversaries (those monitoring parts of the network), leak source information to global adversaries (those able to monitor the traffic of the entire network). To protect against such adversaries, nodes are programmed to transmit *fake messages* even if there are no real events to be reported, so that real transmissions can be embedded within the fake ones [1]. Scheduling fake transmissions according to a certain probabilistic distribution, however, gives adversaries the opportunity to distinguish between real and fake transmissions via statistical analysis, if real events are transmitted as they arrive (assuming the arrival distribution of real events is unknown a priori).

One way to overcome such statistical distinguishability between real and fake transmissions is to transmit real events instead of the next scheduled fake transmission. For instance, 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 procedure, obviously, provides source anonymity since an adversary monitoring

a node will observe one transmission every minute. Hence, assuming the semantic security of the underlying encryption, global adversaries have no means of distinguishing between fake and real transmissions.

The above obvious solution, however, has a major drawback: real events must be delayed until the next scheduled fake transmission. When real events have time-sensitive information, such latency might be unacceptable. Reducing the latency 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 (for example, they maybe deployed in an unreachable or hostile environment). Consequently, a more frequent scheduling algorithm can exhaust nodes' batteries rather quickly, rendering sensor nodes useless. Therefore, practical solutions to the source anonymity problem in wireless sensor networks are designed under two main constraints: minimizing latency and maximizing the lifetime of sensors' batteries.

## 2 Current State-of-the-Art Designs

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. However, instead of delaying real events until the next scheduled fake transmission, real events are transmitted as soon as possible under the following condition: the distribution of the entire message transmissions (fake and real) of each node is "statistically" similar to the transmission of pure fake messages. That is, 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.

**Example [2].** Nodes are designed to transmit fake messages according to an exponential distribution with mean  $\mu$ . Further, nodes store a sliding window of times between consecutive transmissions (inter-transmission times), say  $X_1, X_2, \dots, X_k$ , where  $X_i$  is the exponential random variable representing the inter-transmission time between the  $i^{th}$  and the  $i+1^{st}$  transmissions and  $k$  is the length of the sliding window. Now, when a real event occurs, its inter-

transmission time, denoted by  $X_{k+1}$ , is defined to be the *smallest* value such that the sequence  $X_2, X_3, \dots, X_{k+1}$  passes the Anderson-Darling (A-D) statistical goodness of fit test for an exponential sequence with mean  $\mu$ . However, continuing in this fashion will skew the ensemble mean away from  $\mu$  since nodes always favor shorter inter-transmission times to transmit real events. To adjust the mean, the next inter-transmission time following a real event,  $X_{k+2}$  in this example, is set to  $\mu + \epsilon$ , where  $\epsilon$  is determined so that the sequence of inter-transmission times in the sliding window satisfies the A-D goodness of fit test for an exponential sequence of random variables with mean  $\mu$ . Therefore, an adversary observing the sequence of inter-transmission times will observe a sequence that is statistically indistinguishable from an iid sequence of exponential random variables with mean  $\mu$  [2].

### 3 Correlation Analysis

As can be seen from the above example, inter-transmission times within a sliding window are designed to be indistinguishable from the scheduled fake transmission. Obviously, however, the adversary is not limited to examining a single sliding window. The important question we raise here is: if the adversary is examining more than a single sliding window, can she do better?

Before we answer the above question, we first discuss the possible consequences of distinguishing between multiple sliding windows. *Note that we follow the typical assumption that the adversary is always able to infer the distribution of fake transmissions* (this can be done, for instance, by capturing a sensor node). Consider now an adversary comparing two sliding windows, one of which is statistically similar to the known distribution of pure fake transmissions and the other one is statistically different. The sliding window that is distinguishable from the pure fake transmission window is likely to contain real event transmissions. This observation can reveal private source information to adversaries. To see this, consider the example of a sensor network deployed in a battlefield. The observation of a change in the statistical behavior of a certain node's transmissions during a window of time can translate to the existence of real events in the proximity of this node. Therefore, we argue that, although the adversary might not be able to distinguish between individual real and fake transmissions within a sliding window, location privacy can still be breached by an adversary able to distinguish between two sliding windows.

Consider now the current state of the art in designing anonymous sensor networks proposed in [2]. Recall that the transmission time of a real event is faster than the originally scheduled fake transmission (to minimize latency). Recall further that the transmission following the real event is purposely delayed to adjust the ensemble mean. This implies that, on the average, the inter-transmission time

between a real event and its preceding fake transmission is shorter than  $\mu$ ; and the inter-transmission time between a real event and its successive transmission is longer than  $\mu$ . (For simplicity, we call this pattern for inter-transmission times before and after a real event the "correlation pattern".) That is, while inter-transmission times are iid in sliding windows with pure fake transmissions, they are correlated (by design) in sliding windows containing real transmissions. In particular, time windows containing the transmissions of real events should have more correlation patterns than time windows with pure fake transmissions.

To examine our hypothesis, we performed a simulation analysis of the solution appeared in [2] as follows. We ran 10,000 independent trials. Each trial consists of two windows; one consists of pure fake transmissions scheduled according to exponential iid random variables with mean 20 seconds; and the other one consists of a mixed of fake and real transmissions. Real events arrive according to a Poisson process with mean 1/20 seconds. Upon the arrival of a real event, its transmission time is determined by running the A-D algorithm. (Note that these are the exact same parameters used in [2]).

After running the above experiment, we performed the following correlation test. In each trial, we count the number of correlation patterns for each sliding window. In 6,818 trials, the windows with real transmissions have *more* correlation patterns than pure fake windows; in 2,076 trials, the windows with real transmissions have *less* correlation patterns than pure fake windows; and in 1,106 trials the two windows have equal correlation patterns. That is, an adversary basing her decision solely on the number of correlation patterns will be successful in determining windows with real events 74% of the time.

### 4 Conclusion and Future Work

The adversary's inability to distinguish between individual fake and real transmissions is currently used to model anonymity in wireless sensor networks. We showed that this model does not capture a source of information leakage that can undermine the privacy of sources reported by the sensor networks. Namely, the adversary's ability to distinguish intervals that contain real event transmissions. Future work includes the development of a stronger statistical framework that can properly model source anonymity in wireless sensor networks. The next step is then to design systems that can be shown to provide anonymity when analyzed under such a stronger model.

### References

- [1] K. Mehta, D. Liu, and M. Wright, "Location Privacy in Sensor Networks Against a Global Eavesdropper," in *ICNP 2007. IEEE International Conference on Network Protocols.*, 2007.
- [2] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards Statistically Strong Source Anonymity for Sensor Networks," *INFOCOM 2008. The 27th IEEE Conference on Computer Communications.*, 2008.