NSL

UNIVERSITY OF WASHINGTON

Source Anonymity in Sensor Networks

Bertinoro PhD. Summer School, July 2009

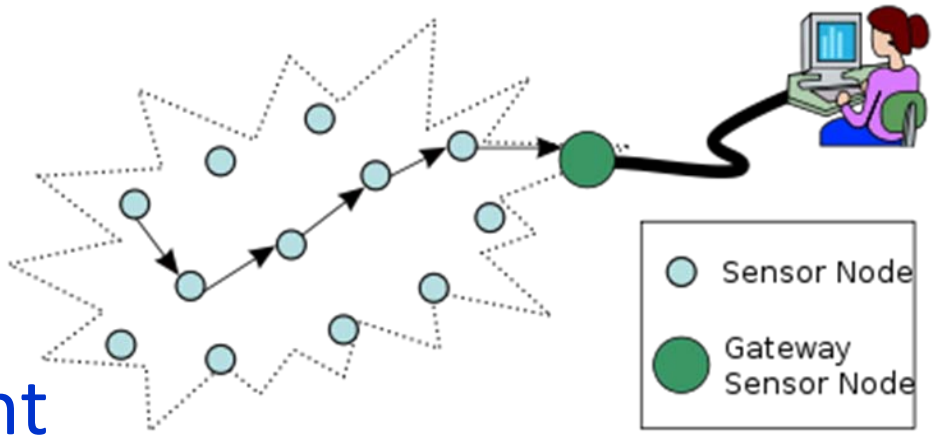
Radha Poovendran
Network Security Lab
Electrical Engineering Department
University of Washington, Seattle, WA
http://www.ee.washington.edu/research/nsl

Outline

- Problem description
- Current solutions
- New model and analysis of current solution
- New approach for designing anonymous systems
- Comparisons
- Conclusions and future work

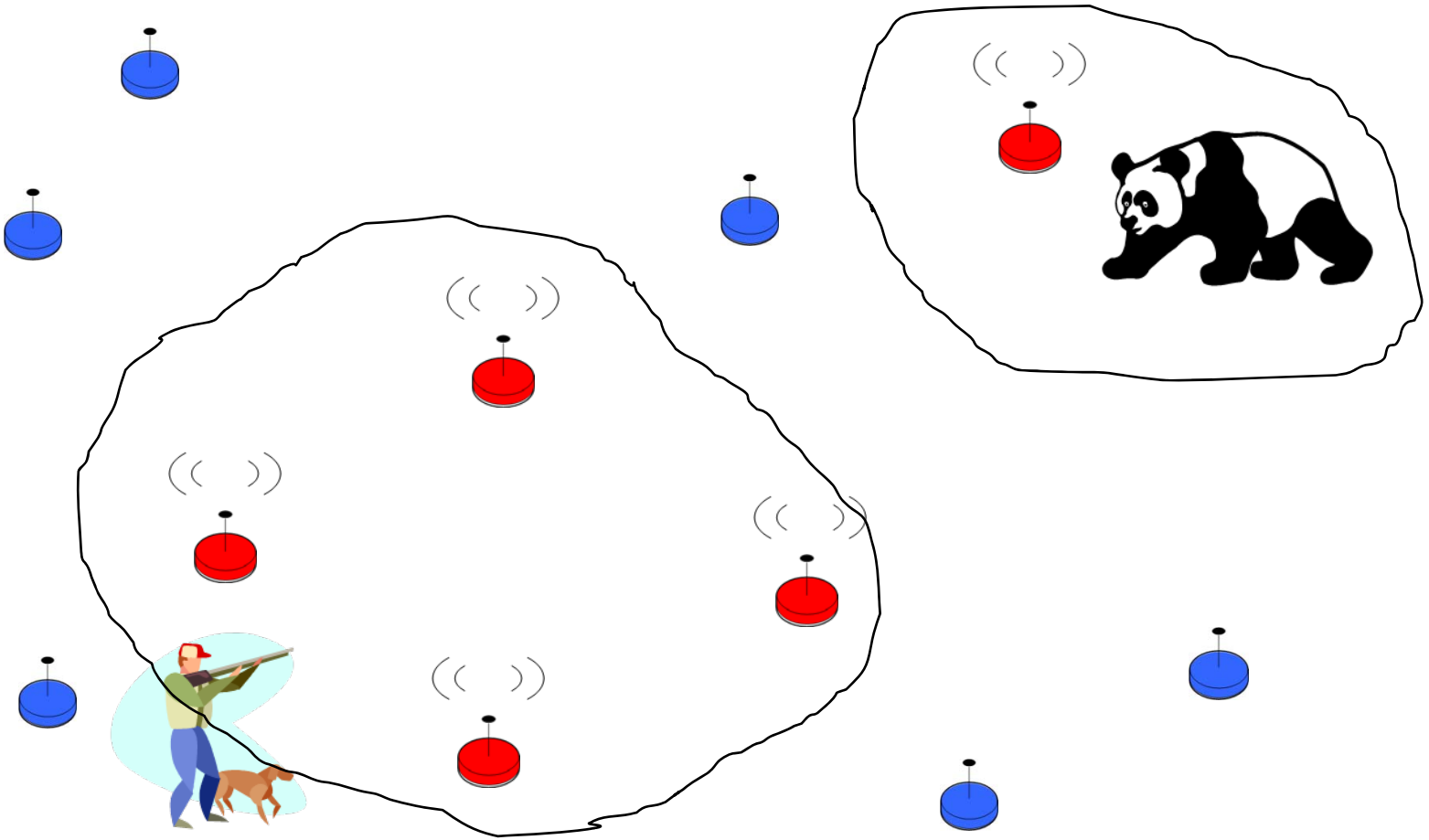
Sensor Networks

- Collection of small devices (sensor nodes)
- Nodes collect and report data
- Nodes have different designs and properties to suit different applications



A Multihop Wireless Sensor Network

Motivation (the Panda-Hunter Game)

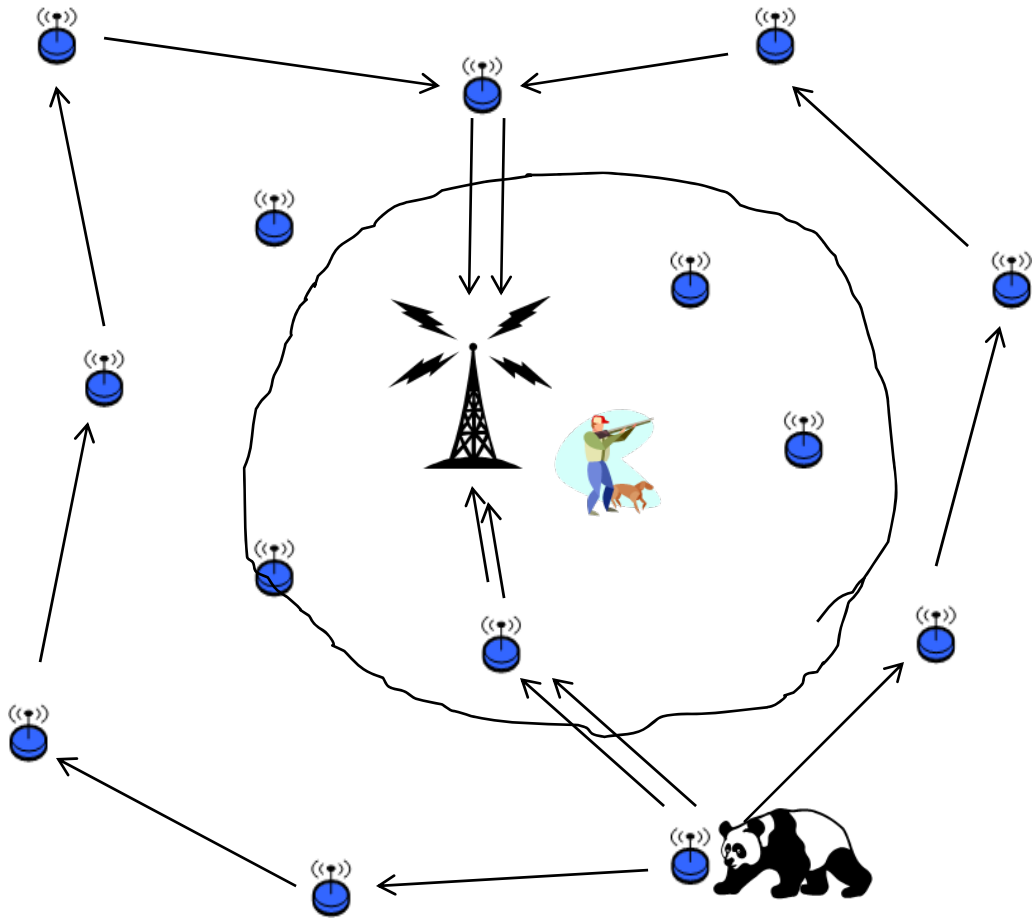


Source Anonymity

- Nodes report sensed events as they arrive
- Adversaries can listen to the same wireless channel used to report data
- Given the location of the node, the location of the event triggering the node can be approximated
- Encryption cannot help! the mere existence of message broadcast is indicative of the occurrence of some events

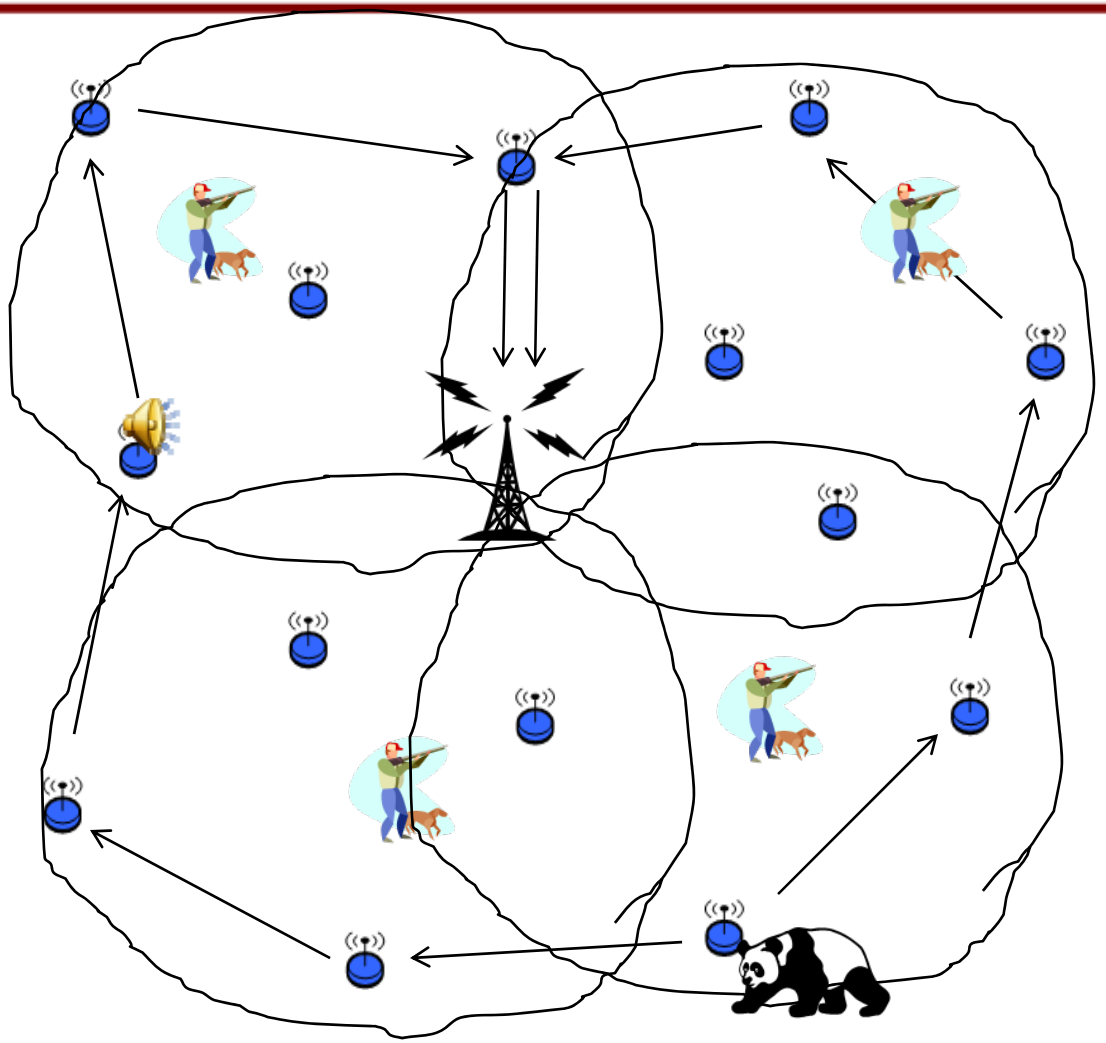
Local Adversary

- Monitor part of the network
- Typically close to the base station
- Routing based approaches have been shown to be effective



Global Adversary

- Monitor the entire network
- Typically through collaboration of adversaries
- Routing based approaches have been shown to be ineffective



Event Triggered Broadcast

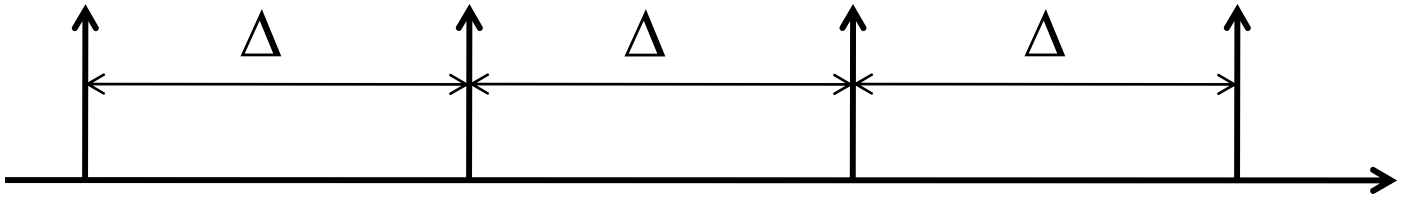
- When nodes only transmit messages upon the occurrence of a real event, the event can be detected by global adversaries
- Program nodes to continuously transmit fake messages
- Real events can be embedded within the fake message transmissions

Example

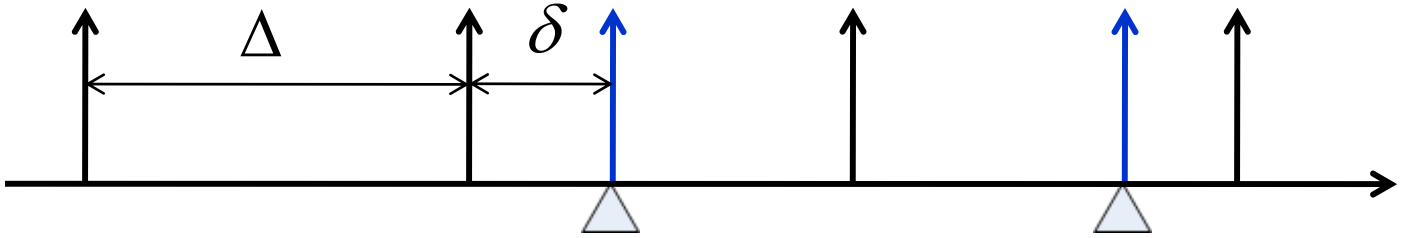
Fake message schedule

Example

Fake message

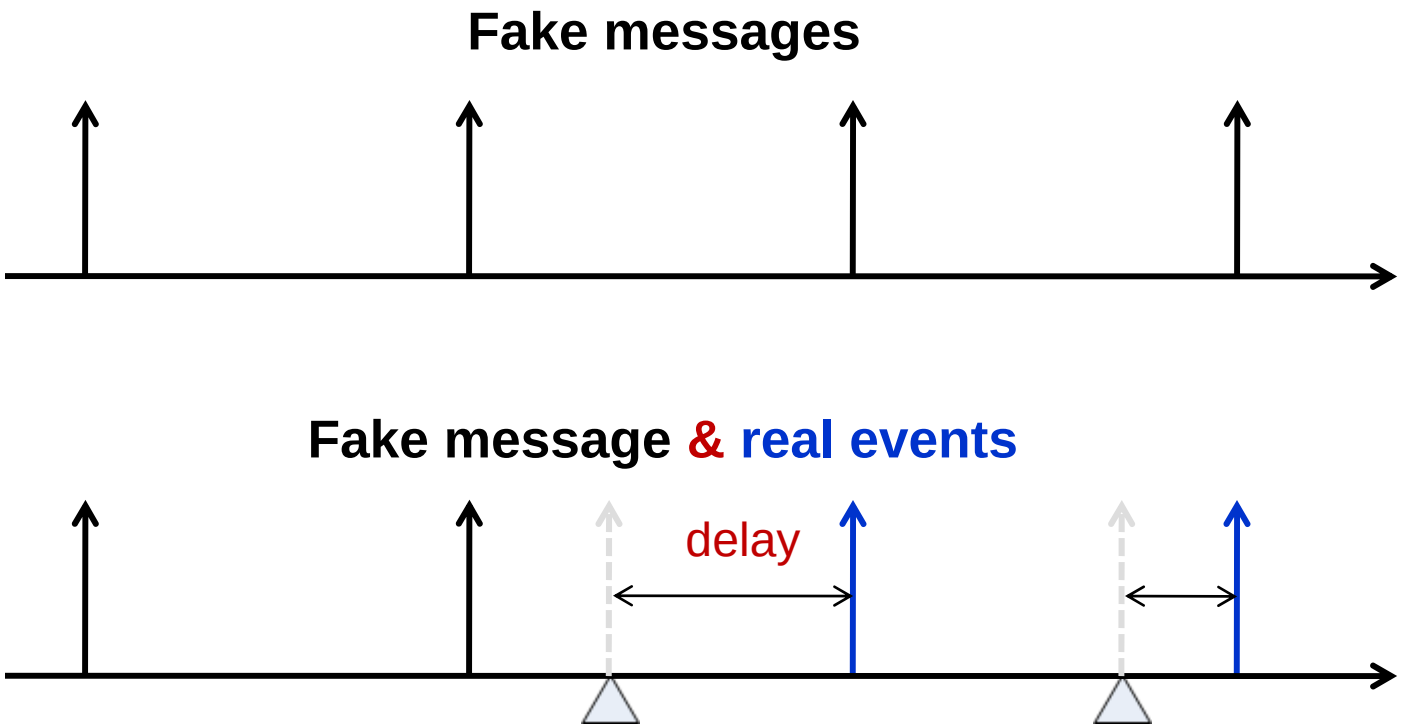


Fake message & real events



How can we remove this statistical distinguishability?

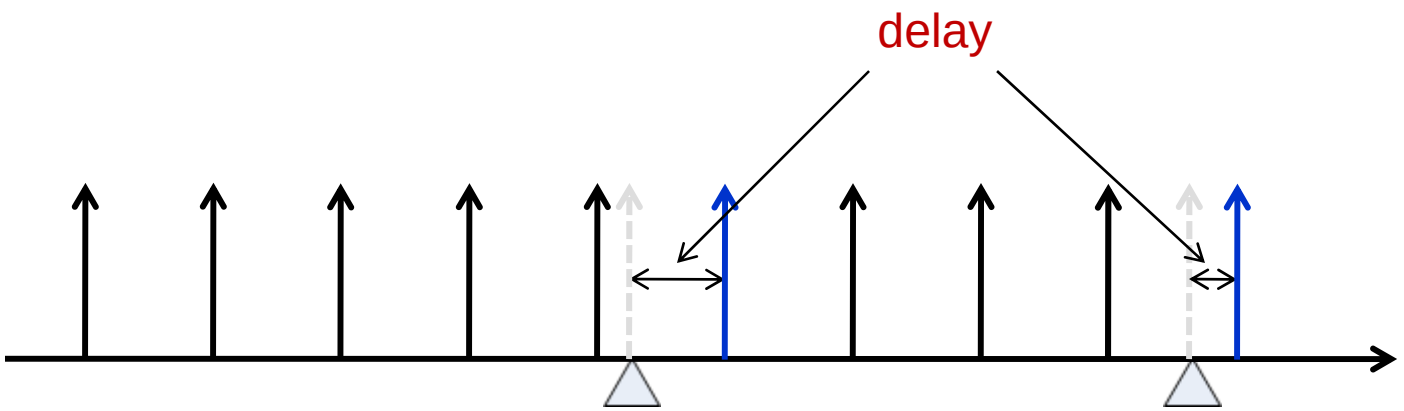
Naive Solution: Delay and Transmit



How can we reduce latency?

Another Naïve Solution: Increase The Fake Event Transmission Rate

Fake messages & real events



More frequent transmissions can exhaust sensors' batteries rather quickly!

Problem Statement

Report real events
with minimum delay,
while maintaining source anonymity,
without exhausting sensors' batteries.

Statistical Goodness of Fit Tests

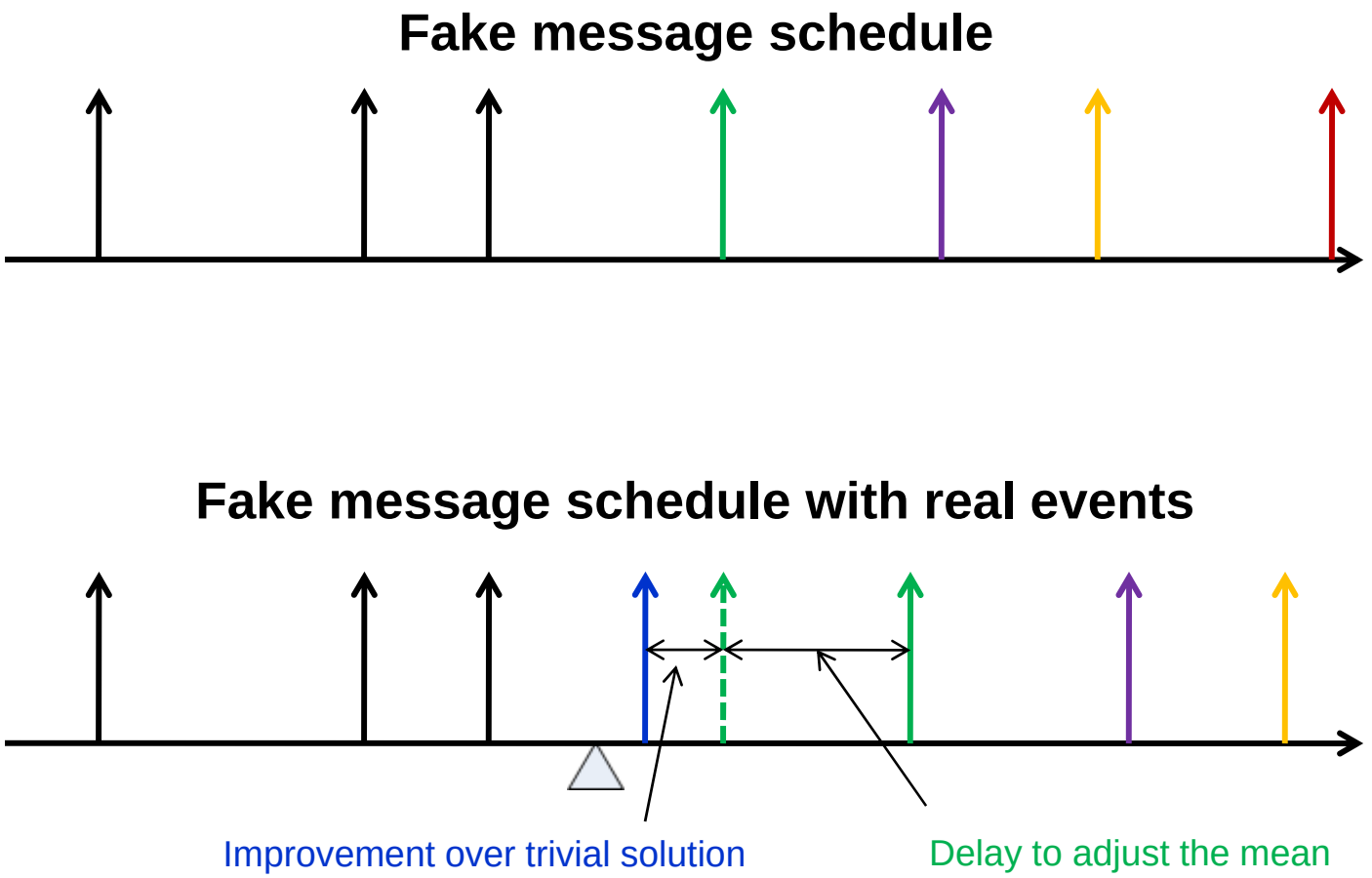
- Given a sequence of data samples, a statistical goodness of fit test determines whether or not the data samples follow a desired probabilistic distribution
- Examples:
  - Anderson-Darling test (A-D test)
  - Kolmogorov-Smirnov test (K–S test)
  - Jarque-Bera test (J-B test)

Current State of the Art

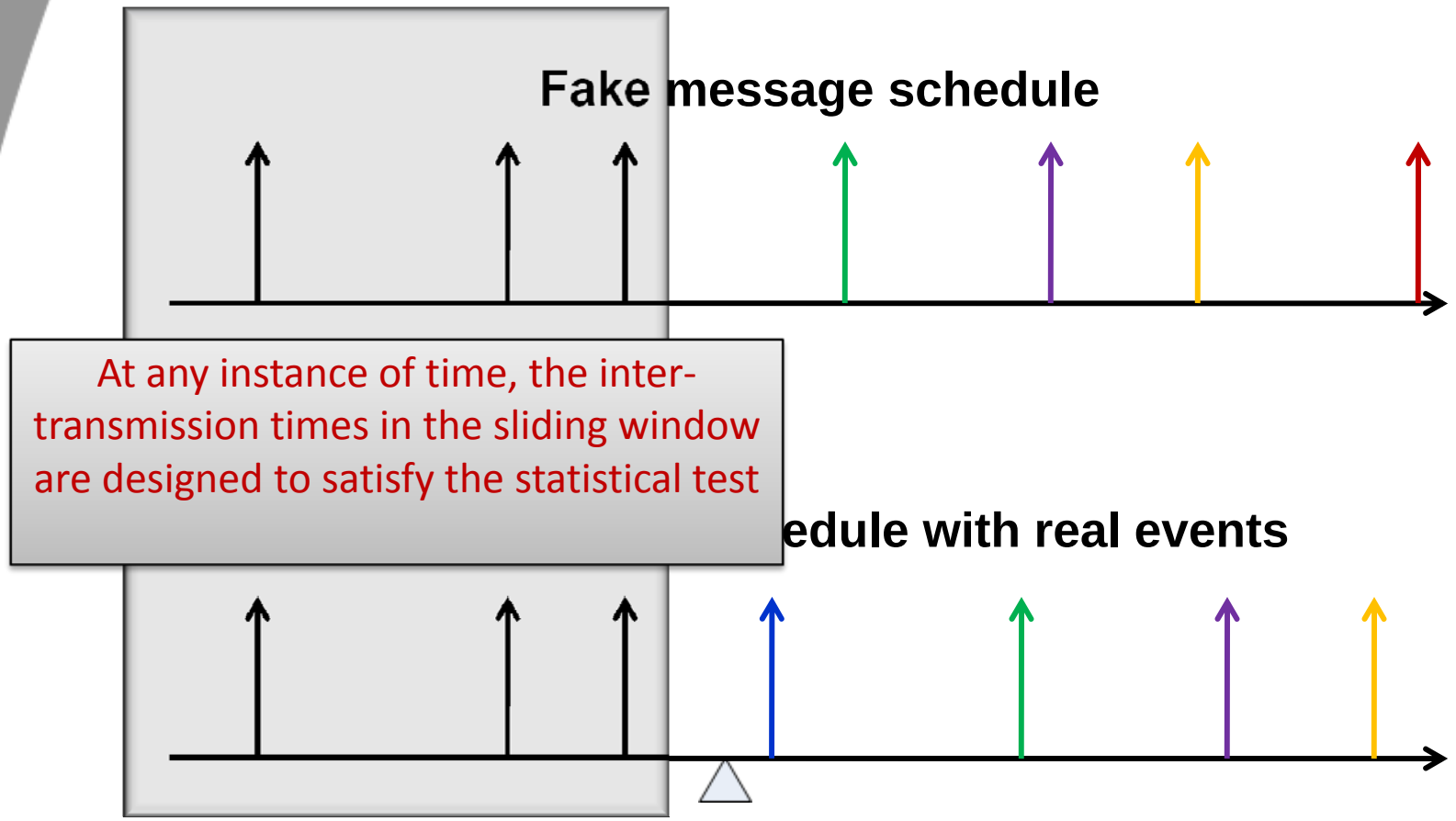
- Nodes are programmed to transmit fake messages according to pre-defined distribution
- Keep a sliding window of inter-transmission times
- When a real event occurs, its transmission time is computed as the minimum time so that the inter-transmission times in the sliding window passes statistical goodness of fit tests
- Continuing sending real events this way will skew the mean
- To fix the mean, the transmission following a real one is delayed

Inter-transmission times in the sliding window always satisfy the used statistical goodness of fit tests for the desired distribution

Example


Fake message schedule
Fake message schedule with real events
Improvement over trivial solution
Delay to adjust the mean

Example


Fake message schedule
At any instance of time, the inter-transmission times in the sliding window are designed to satisfy the statistical test
edule with real events

Example

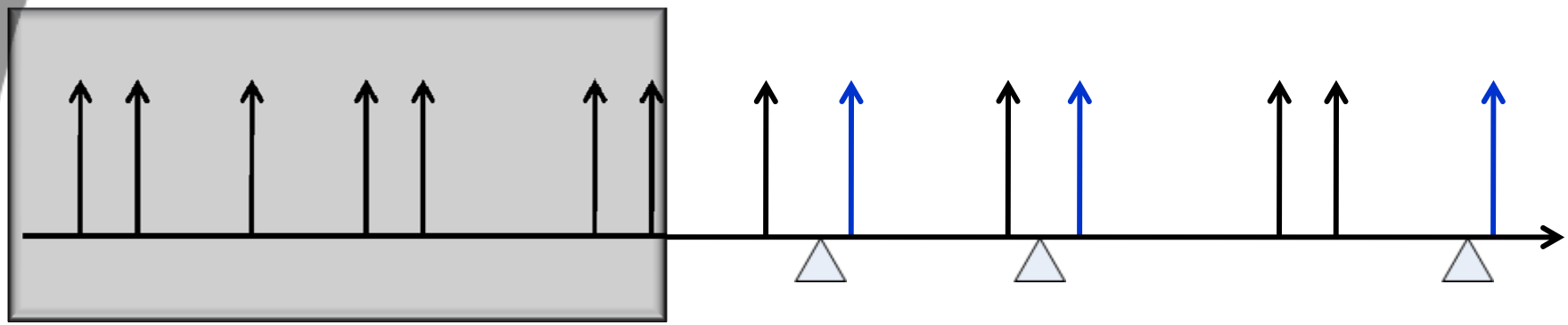
- Shao et al. [1] have introduced the idea of using statistical goodness of fit test to transmit real events
- To determine when to transmit a real event, they run the Anderson-Darling test (A-D test)
- The transmission after a real one is delayed
- The delay is determined to adjust the mean and to satisfy the A-D test for the sliding window
- To analyze the scheme, the Kolmogorov-Smirnov test (K–S test) is used to test the data
- The inter-transmission times in the sliding window were found to also pass the K-S test

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

But adversary is intelligent

- An adversary testing a sliding window has been shown to be unable to distinguish between real and fake transmission
- However, the adversary is not restricted to only examining the sliding window
- Question: can the adversary do better?

Smarter Adversary



Instead of examining a sliding window, can the adversary do better when comparing two time windows?

If the adversary can determine which window contains real transmissions, source location can be revealed

New Model

- Instead of modeling anonymity by the adversary's ability to examine a sliding window, model anonymity by the adversary's ability to distinguish between two windows (one with real events and the other one without)
- Given two time intervals, one with real events (call it the real interval "I_R") and one without (call it the fake interval "I_F"), define the adversaries confidence in distinguishing between them as:

$$\varepsilon = 2Pr(I_g=I_R) - 1,$$

where I_g denotes the adversary's guess of the real interval

NSL

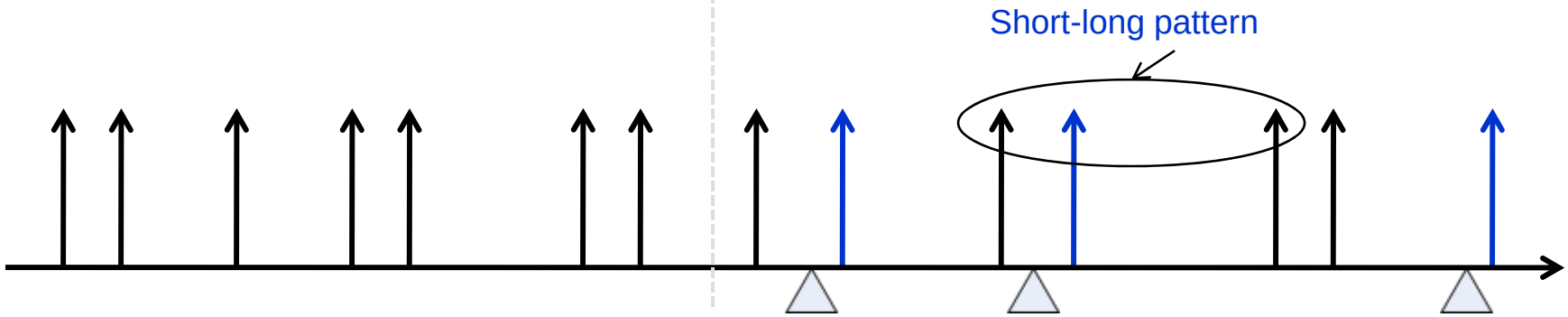
ε-Anonymity

- Definition: given two intervals, a fake one I_F and a real one I_R, a sensor network is said to be ε-Anonymous if it satisfies two conditions
  1) in different stages of each interval, inter-transmission times are indistinguishable
  2) the adversary's confidence in distinguishing between the two intervals is at most ε

Analysis of the Current State of the Art

- Observations:
  - The transmission of a real event is sooner than the transmission of the scheduled fake message
  - The transmission following a real one is delayed to adjust the mean
- Conclusion:
  - In the window containing real events, transmission times are correlated
  - In the window without real events, transmission times are independent

Short-long Patterns



Short-long patterns: the time between a real event and its previous transmission is usually shorter than average, while the time between a real transmission and the following one is usually longer than average

Mathematical Analysis

- X_i :the random variable representing the time between the i^{th} and the $i\text{-}1^{st}$ transmissions, & $E[X_i]=\mu$

Short-long patterns are most likely to occur in the sliding window with real events

- In fake intervals:

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

- In real intervals:

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

$> \mu$ $= \mu$

Therefore, real intervals are expected to have more short-long patterns than fake intervals

Experimental Analysis

- Setup
  - Fake messages are iid exponentials with mean 20s
  - Real events arrive according to a Poisson process with mean 1/20
  - Run 10,000 independent trials
  - Each trial consists of two intervals, a fake one and a real one
  - When real events arrive, their transmission time is determined by the Anderson-Darling test
  - Each trial starts with a “worm up” period where 200 iid exponentials are drawn to constitute a backlog for the A-D test

Experimental Analysis

- Strategy
  - Recall that it was shown mathematically that real intervals are expected to have more short-long patterns than fake intervals
  - For each trial, the adversary count the number of short-long patterns in each interval
  - The interval with more short-long patterns is chosen as the real interval
  - If both intervals have the same number of short-long patterns, the adversary decides randomly

Experimental Analysis

- Results
  - Out of the 10,000 trials the following is obtained
    - 6,818 trials have more short-long patterns in **real intervals**
    - 2,076 trials have more short-long patterns in **fake intervals**
    - 1,106 trials have the same number of short-long patterns in both intervals
  - Anonymity interpretations
    - With the described strategy, the probability of correctly identifying real intervals is 0.7371
    - Therefore, the adversary's confidence in distinguishing between real and fake intervals is ε=0.474

Improved Solutions

- Recall that inter-transmission times in fake intervals are iid's, while correlated in real intervals
- Observe that the definition of ε-anonymity does not require fake intervals to be iid
- Try to introduce the same correlation in real intervals

Experimental Analysis

- Setup
  - Fake messages are iid exponentials with mean 20s
  - Real events arrive according to a Poisson process with mean 1/20
  - Run 10,000 independent trials
  - Each trial consists of two intervals, a fake one and a real one
  - When real events arrive, their transmission time is determined by the Anderson-Darling test
  - Each trial starts with a "worm up" period where 200 iid exponentials are drawn to constitute a backlog for the A-D test
  - During fake intervals, ***dummy events*** are generated with normal distribution of mean 10s and variance 150 (to simulate real events)
  - Dummy events are transmitted according to the A-D test

Experimental Analysis

- Strategy
  - For each trial, the adversary count the number of short-long patterns
  - The interval with more short-long patterns is chosen as the real interval
  - If both intervals have the same number of short-long patterns, the adversary decides randomly

Experimental Analysis

- Results
  - Out of the 10,000 trials the results are summarized in the following table

	$I_R > I_F$	$I_R < I_F$	$I_R = I_F$	ϵ
Basic approach	$6,818$	$2,076$	$1,106$	0.474
Improved approach	$4,566$	$4,272$	$1,162$	0.029
Trivial solution	$4,385$	$4,318$	$1,297$	0.007

The trivial solution is the theoretically anonymous approach of transmitting real events instead of the next scheduled fake transmission

NSL

Acknowledgements

- Graduate Student: Basel Aloimar
- www.ee.washington.edu/research/nsl