Temporal Privacy in Wireless Sensor Networks

Pandurang Kamat, Wenyuan Xu, Wade Trappe and Yanyong Zhang
WINLAB
Rutgers University

Temporal Privacy in Sensor Networks

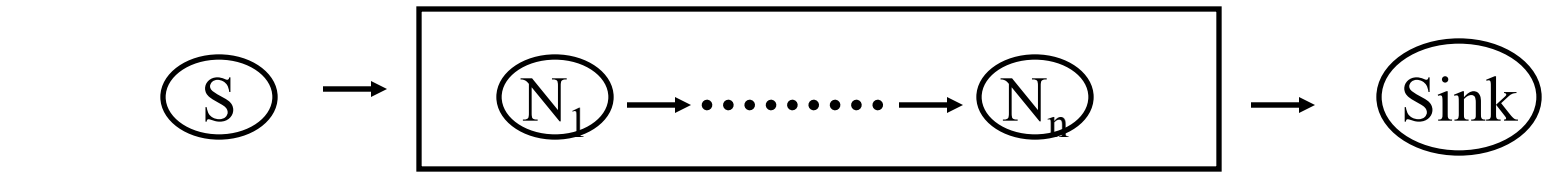
- Although the content of sensor messages describing “events of interest” may be encrypted to provide confidentiality, the context surrounding these events may also be sensitive and therefore should be protected from eavesdroppers

- An adversary armed with knowledge of the network deployment, routing algorithms, and the base-station (data sink) location can infer the temporal patterns of interesting events by merely monitoring the arrival of packets at the sink, thereby allowing the adversary to remotely track the spatio-temporal evolution of a sensed event

Formulating Temporal Privacy

- Temporal pattern of events in sensor networks may be inferred by an adversary observing the traffic at the data-sink.

- Introduce random delay at intermediate nodes to mask time of origin
  - Amortizes buffer requirements across all intermediate nodes.
  - Each node uses a different delay distribution.



X
Msg. distribution at source

Y
Cumulative delay

Z
Message arrival at sink

Information adversary learns by observing Z is defined by the mutual information :

$I(X;Z) = h(X) - h(X|Z) = h(Z) - h(Z|X) = h(Z) - h(X + Y|X) = \mathbf{h(Z) - h(Y)}$

Formulating Temporal Privacy contd..

- The entropy-power inequality gives a lower bound on this privacy measure:

$$I(X;Z) \geq \frac{1}{2\ln 2}\left(2^{2h(X)} + 2^{2h(Y)}\right) - h(Y).$$

- The challenge of temporal privacy is to choose a delay distribution f_Y to minimize inferred information about X.

$$\min_{f_Y(y)} I(X;Z) = h(X+Y) - h(Y),$$

- For a series of messages let X^n be the message origination times, Y^n represent the series of delays introduced in the intermediate nodes and Z^n be the adversary observation times for the messages.

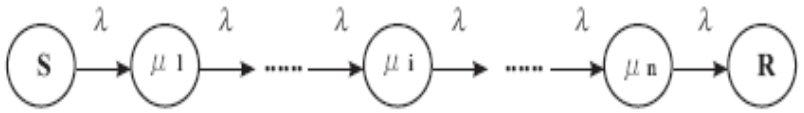
- Assuming $\{X_j\}$ to be Poisson traffic with rate λ (inter-arrival times 1/λ) and choosing an exponential delay distribution $\{Y_j\}$ with mean 1/μ, we derive :

$$I(X^n, Z^n) \leq \sum_{j=1}^{n} \ln\left(1 + \frac{j\mu}{\lambda}\right)$$

- Exponential distribution yields maximal entropy among non-negative distributions.

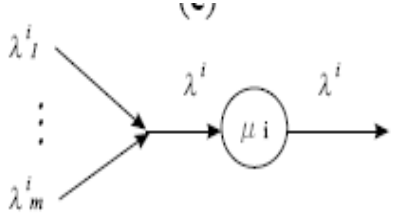
Queuing Analysis

- We model the nodes in the network as M/M/k/k queues where each node can buffer upto k packets. The probability that a new packet finds the buffers full is given by Erlang's loss formula



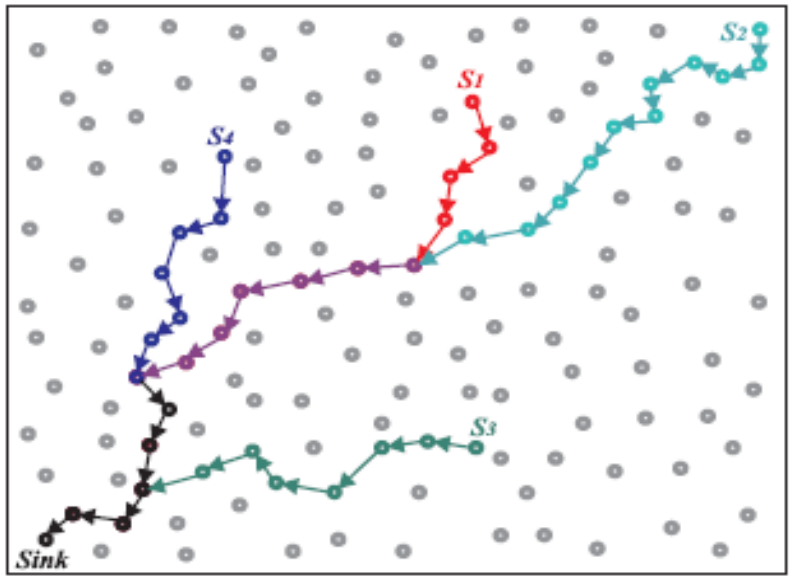
$$E(\rho, k) = \frac{\rho^k}{k!} p_0 = \frac{\frac{\rho^k}{k!}}{\sum_{i=0}^{k} \frac{\rho^i}{i!}}$$

where $\rho = \lambda / \mu$.



Network & Adversary Models

- Network Model
  - Delay tolerant application
  - Encrypted payload
  - Cleartext headers



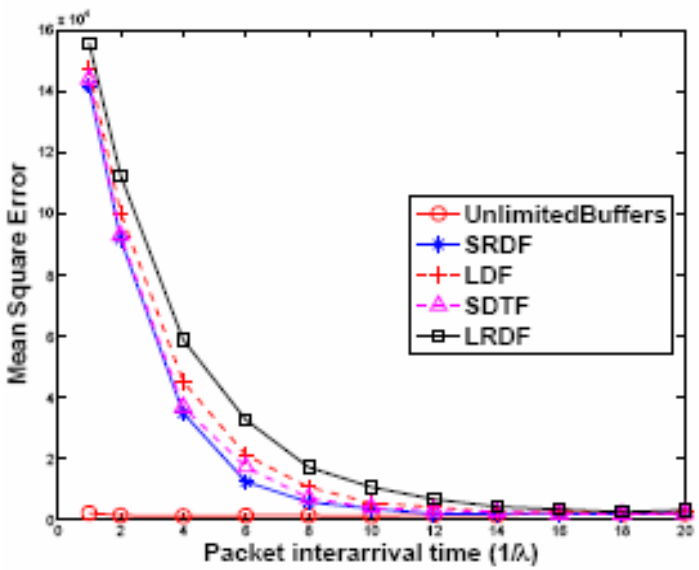
- Adversary Model
  - Protocol aware
  - Able to eavesdrop
  - Deployment aware
  - Non interfering
  - Can infer the hop-count between source and sink
  - Knows the delay distributions on the nodes

Measuring Temporal Privacy

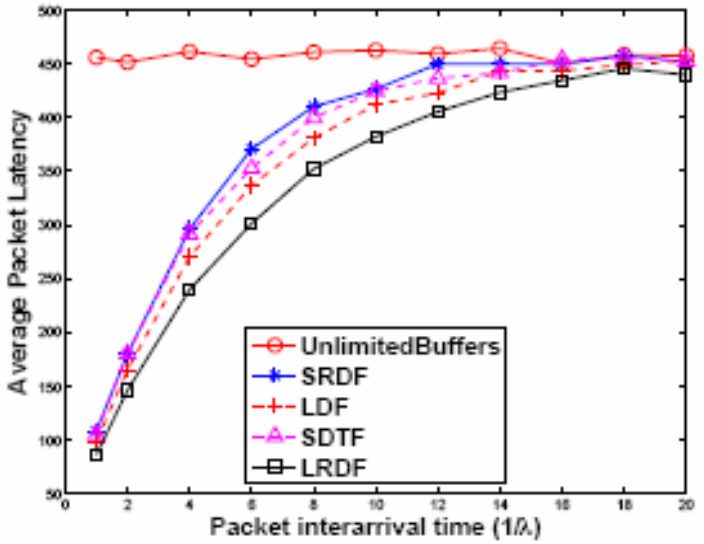
- Adversary observes time of arrival z of a packet.
- Adversary estimate of origin time of packet : x' = z – y'
- Privacy metric for 'm' observed packets, is the mean square error in the estimate: $\sum(x_i' - x_i)^2/m$
- Higher estimation error for the adversary means better temporal privacy for the network.
- This metric has a direct relationship with the information theoretic metric of mutual information.

RCAD: Rate Controlled Adaptive Delaying

- Due to limited buffers on sensor nodes, we need to adjust delay distributions based on incoming traffic rate and buffer availability. The main idea behind RCAD is to preempt buffered packets (send out before time), using one of the following strategies when the buffer is full.
  - Longest Delayed First (LDF)
  - Longest Remaining Delay First (LRDF)
  - Shortest Delay Time First (SDTF)
  - Shortest Remaining Delay First (SRDF)



(a) Mean square error



(b) Delivery latency

Improved Model for Delay Distributions

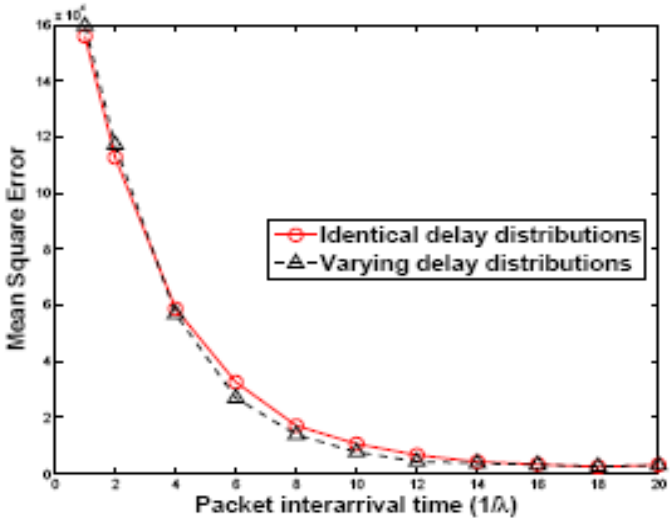
- Because of the peculiar traffic patterns in sensor networks the nodes near the sink experience higher concentration of traffic.
- This results in their buffer getting filled faster, causing more preemption decisions to be made ➔ wasted processing cycles.
- So we vary delays on nodes based on their distance from the sink.
- Closer nodes ➔ shorter delays
- Number of buffered packets at a node can be estimated as: $\overline{N_i} = \rho_i = \frac{\lambda_i}{\mu_i}$

- Consider a h hop flow, with node h being the source and node 1 the last node before sink.
- Goal: Keep number of buffered packets constant across all h hops while maintaining a target cumulative delay of D.
- Given D each node can derive its average delay time as:
  - Where γ is the Euler-Mascheroni constant and
  - ψ is the digamma function.

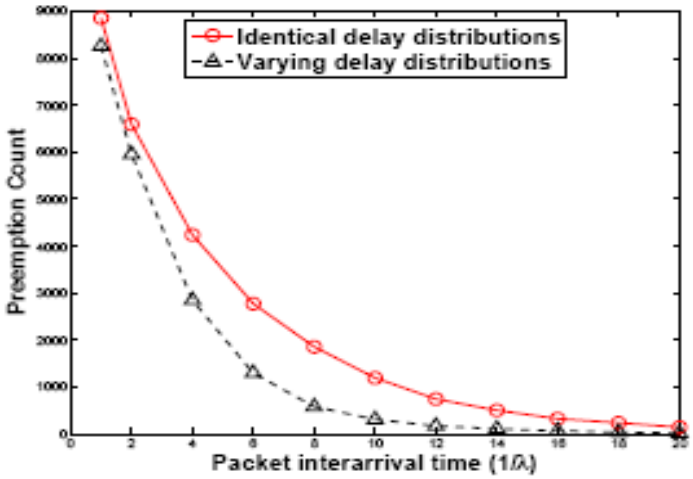
$$1/\mu_i = \frac{D}{(i+1)(\gamma + \psi(h+1))}.$$

Reducing preemptions with variable delay

- Fewer preemptions ➔ more computational efficiency ➔ without any privacy reduction.



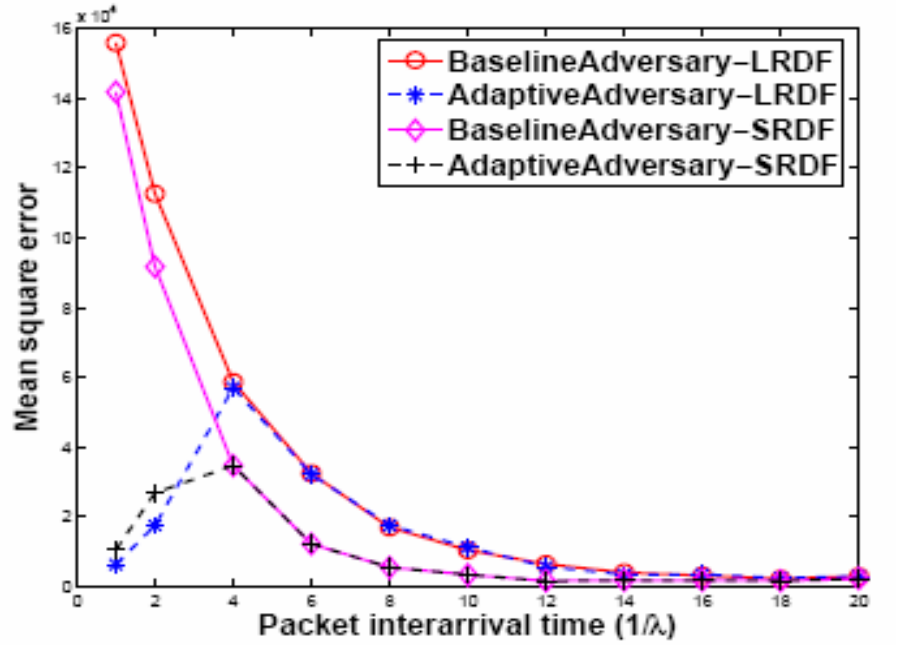
(a) Mean square error



(c) Number of Preemptions

Adaptive Adversary Model

- At higher traffic rates the adversary estimation error is high
  - More traffic ➔ higher preemptions ➔ effective cumulative delay distribution is significantly different from the original one.
- At higher traffic rates the adversary estimation error is high
  - More traffic ➔ higher preemptions ➔ effective cumulative delay diverges further from the original.
- Given: Total n sources, h is the hop count for source S, k is the size of the buffers, λ is the adversary observed rate of packet arrivals.



The adaptive adversary estimate algorithm for average network delay uses the Erlang's loss formula to estimate preemption probability and adapt, during high volume traffic to account for RCAD preemption strategies.

Average adversary network delay estimate:
$Y = h / \mu$, if loss_probability < threshold
$Y = hk / \lambda$, otherwise

Related Work

- [Chan-Perrig 2003] Address content privacy in sensor networks.

- [Gruteser et.al 2003] Privacy in location services using sensors
  - Data cloaking using k-anonymity model
  - Constant-rate and size traffic to counter traffic analysis
  - Applied to location services in office-like environments.
  - Cannot be applied to large-scale sensor networks like the ones we address.

- [Deng et. al. 2004] Anti-traffic analysis strategies to protect data-sink.
  - Constant-data rate suggested to mask real traffic headed to sink.
  - No privacy metric defined nor any information leakage or lack thereof after applying their algorithm is measured.

- [Deng et.al.2005] Anti-traffic analysis strategies to protect data-sink.
  - Use the idea of random walk similar to ours to protect location of base-station
  - Fractal propagation as a routing strategy
  - Injection of fake traffic to create high-communication-activity-areas.

Conclusions

- We have proposed an information theoretic formulation of temporal privacy in wireless sensor networks
- We presented an adaptive delaying technique that provides temporal privacy while alleviating congestion in the proximity of the sink.
- Temporal privacy and buffer utilization were shown to be conflicting objectives.
- To manage this tradeoff we introduced Rate Controlled Adaptive Delaying (RCAD) strategies for packet preemptions.
- We also studied a enhanced adversary model and showed that RCAD can still provide temporal privacy against it.

Questions ?