Simultaneous and sequential detection of multiple interacting change points

Long Nguyen
Department of Statistics
University of Michigan

Joint work with Ram Rajagopal (Stanford University)

Introduction

- Statistical inference in the context of spatially distributed data processed and analyzed by decentralized systems
  - sensor networks, social networks, the Web

- Two interacting aspects
  - how to exploit the spatial dependence in data
  - how to deal with decentralized communication and computation

Introduction

- Statistical inference in the context of spatially distributed data processed and analyzed by decentralized systems
  - sensor networks, social networks, the Web

- Two interacting aspects
  - how to exploit the spatial dependence in data
  - how to deal with decentralized communication and computation

- Extensive literature dealing with each of these two aspects separately by different communities

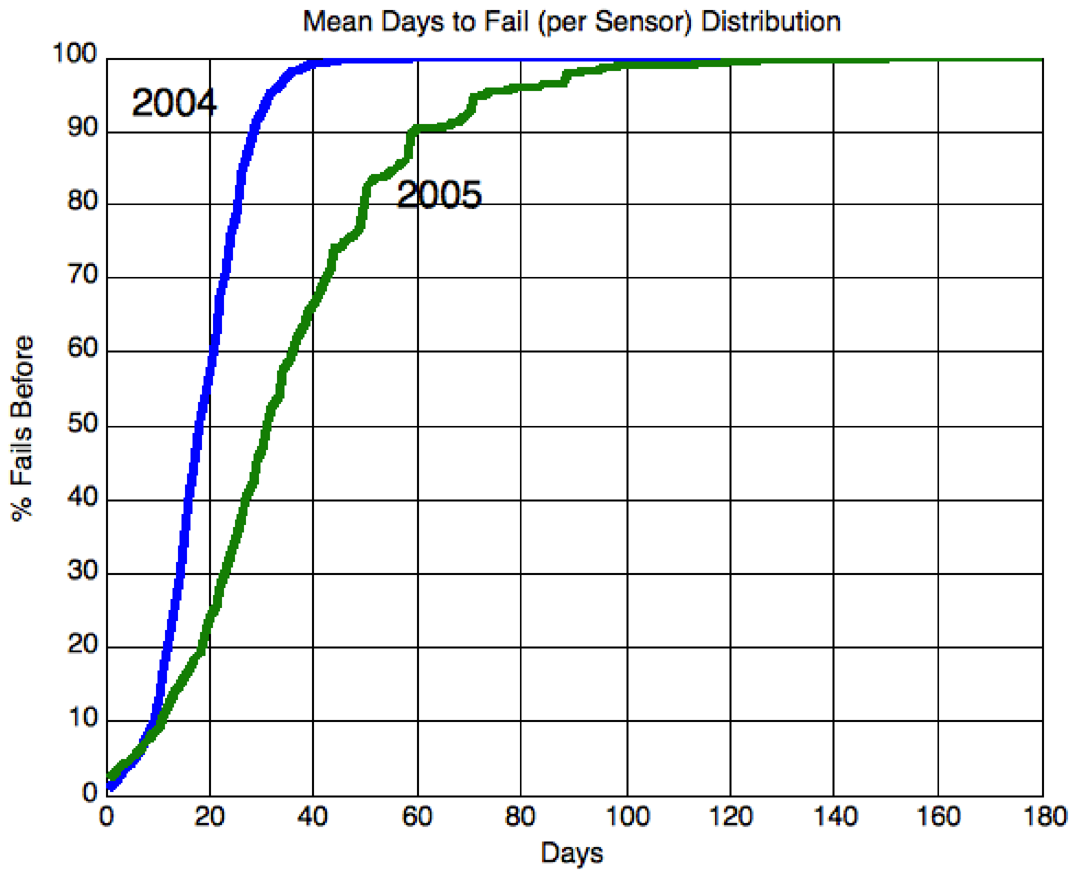
- Many applications call for handling both aspects in near "real-time" data processing and analysis

Example – Sensor network for traffic monitoring

Problem: detecting sensor failures for all sensors in the network

- data: sequence of sensor measurements of traffic volume
- sequential detection rule for change (failure) point, one for each sensor

“Mean days to failure”



- as many as 40% sensors fail a given day
- need to detect failed sensors as early as possible
- separating sensor failure from events of interest is difficult

Talk outline

- statistical formulation for detection of *multiple* change points in a network setting
  - classical sequential analysis
  - graphical models
- sequential and "real-time" message-passing detection algorithms
  - decision procedures with limited data and computation
- asymptotic theory of the tradeoffs between statistical efficiency vs. computation/communication efficiency

Sequential detection for single change point

- sensor u collects sequence of data $X_n(u)$ for $n = 1, 2, \ldots$
- λ_u change point variable for sensor u
- data are i.i.d. according to f_0 before the change point; and iid f_1 after

Sequential detection for single change point

- sensor u collects sequence of data $X_n(u)$ for $n = 1, 2, \ldots$
- λ_u change point variable for sensor u
- data are i.i.d. according to f_0 before the change point; and iid f_1 after
- a sequential change point detection procedure is a stopping time τ_u, i.e., $\{\tau_u \leq n\} \sim \sigma(X_1(u), \ldots, X_n(u))$

Sequential detection for single change point

- sensor u collects sequence of data $X_n(u)$ for $n = 1, 2, \ldots$
- λ_u change point variable for sensor u
- data are i.i.d. according to f_0 before the change point; and iid f_1 after
- a sequential change point detection procedure is a stopping time τ_u, i.e., $\{\tau_u \leq n\} \sim \sigma(X_1(u), \ldots, X_n(u))$

- Neyman-Pearson criterion:
  - constraint on false alarm error

$$PFA(\tau_u(X)) = P(\tau_u < \lambda_u) \leq \alpha \text{ for some small } \alpha$$

  - minimum detection delay

$$\mathbb{E}[(\tau_u - \lambda_u) | \tau_u \geq \lambda_u].$$

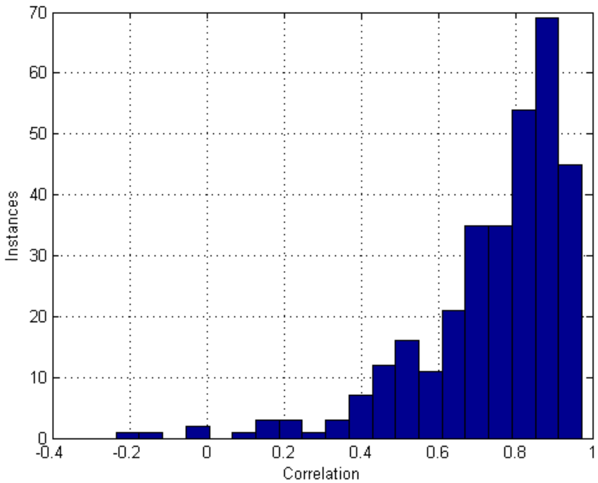
Beyond a single change point

- we have multiple change points, one for each sensor

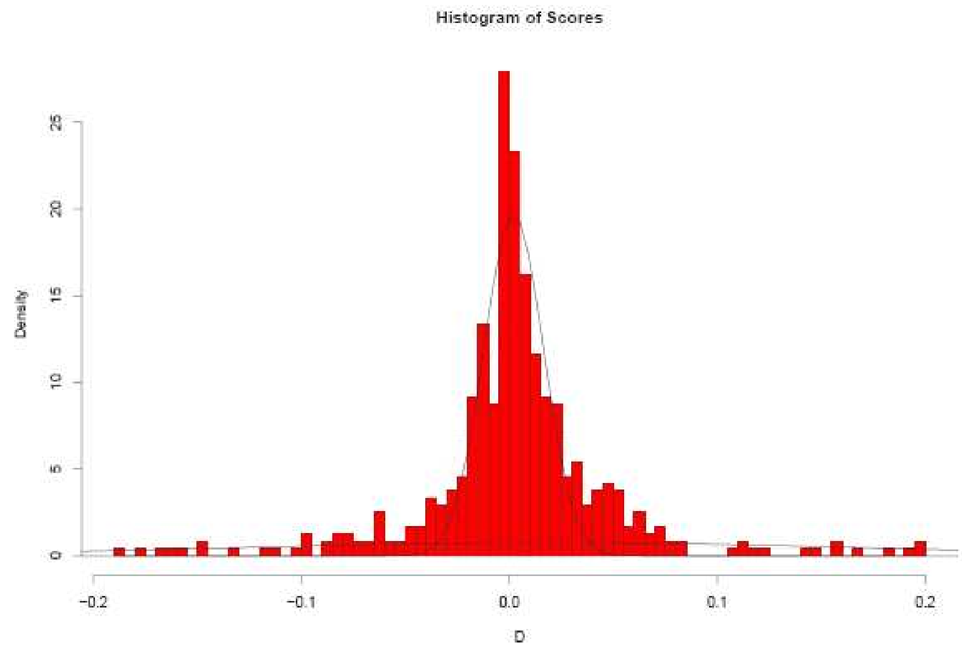
- we could apply the single change point method to each sensor independently, but this is not a good idea
  - measurements from a single sensor are very noisy
  - failed sensors may still produce plausible measurement values

- borrowing information from neighboring sensors may be useful
  - due to spatial dependence of measurements
  - but data sharing limited to neighboring sensors
  - data sharing via a message-passing mechanism

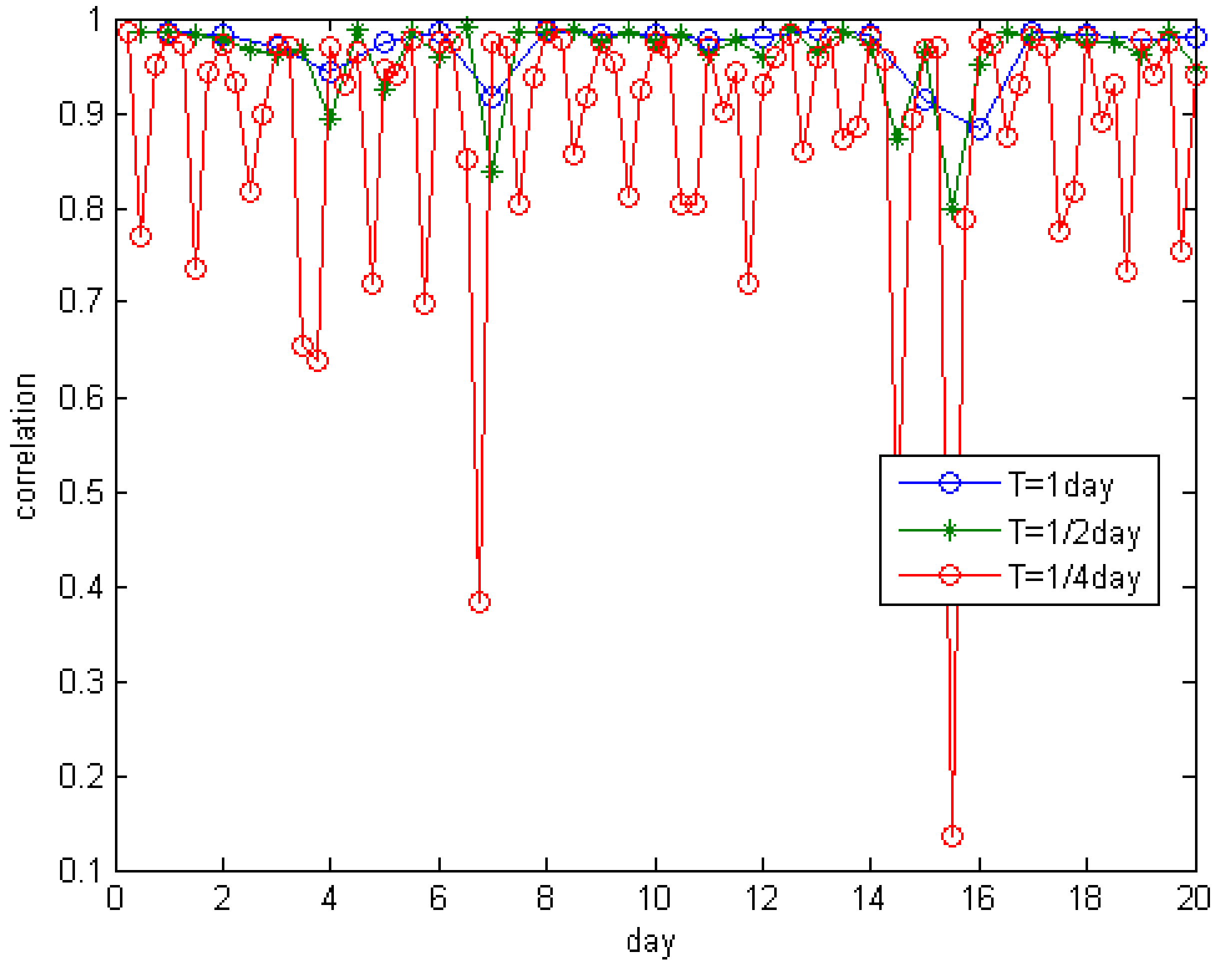
Sample correlation with neighbors



Correlation with good sensors



Correlation with failed sensors



Correlation statistics have been successfully utilized in practice, although not in a sequential and decentralized setting (Kwon and Rice, 2003)

A formulation for multiple change points

- m sensors labeled by $U = \{u_1, \ldots, u_m\}$
- given a graph $G = (U, E)$ that specifies the the connections among $u \in U$
- each sensor u fails at time λ_u
  - λ_u is endowed with (independent) prior distribution π_u

A formulation for multiple change points

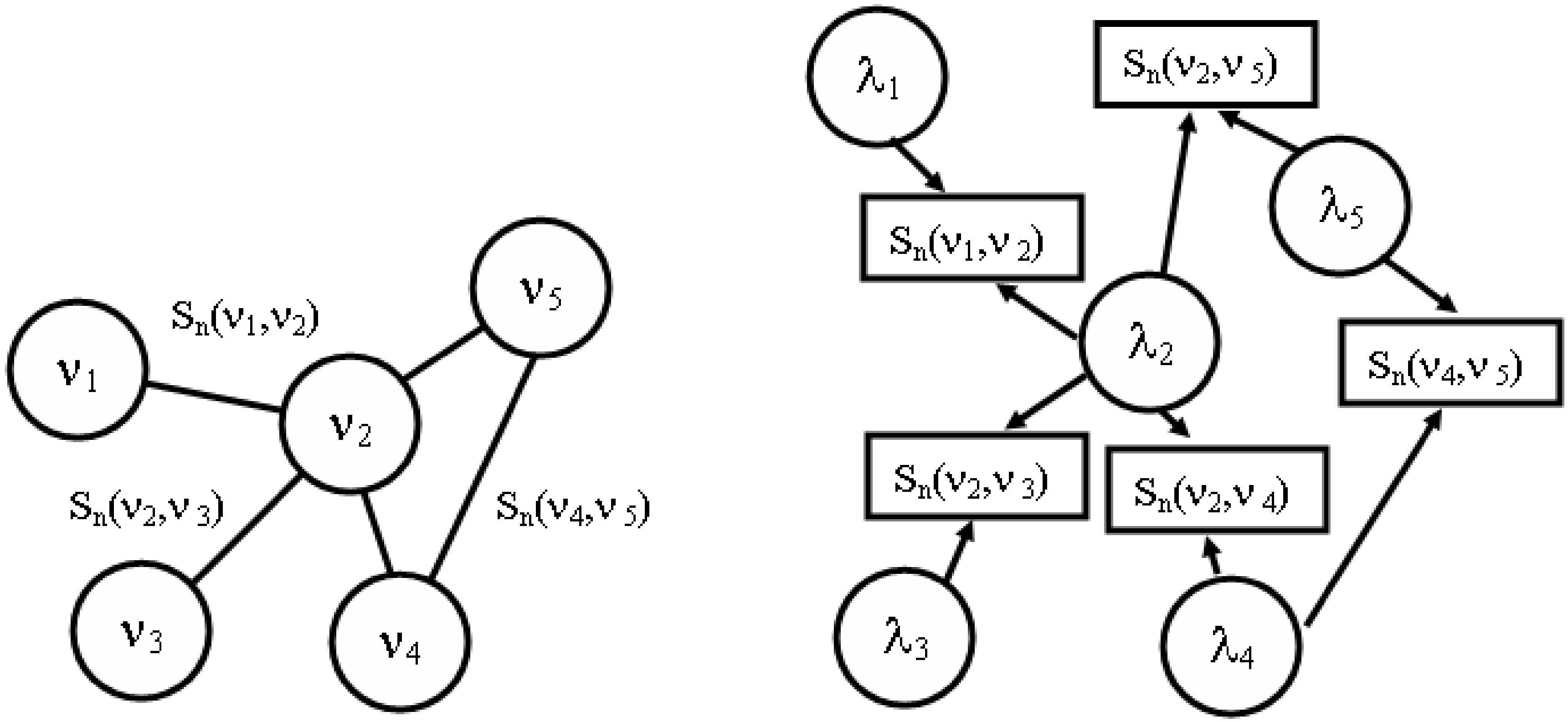
- m sensors labeled by $U = \{u_1, \ldots, u_m\}$
- given a graph $G = (U, E)$ that specifies the the connections among $u \in U$
- each sensor u fails at time λ_u
  - λ_u is endowed with (independent) prior distribution π_u
- there is private data sequence $X_n(u)$ for sensor u
  - private data sequence changes its distribution after λ_u

A formulation for multiple change points

- m sensors labeled by $U = \{u_1, \ldots, u_m\}$
- given a graph $G = (U, E)$ that specifies the the connections among $u \in U$
- each sensor u fails at time λ_u
  - λ_u is endowed with (independent) prior distribution π_u
- there is private data sequence $X_n(u)$ for sensor u
  - private data sequence changes its distribution after λ_u
- there is shared data sequence $(Z_n(u, v))_n$ for each *neighboring pair* of sensors u and v:

$$
\begin{aligned}
Z_n(u, v) \quad &\overset{iid}{\sim} \quad f_0(\cdot|u, v), \quad \text{for } n < \min(\lambda_u, \lambda_v) \\
&\overset{iid}{\sim} \quad f_1(\cdot|u, v), \quad \text{for } n \geq \min(\lambda_u, \lambda_v)
\end{aligned}
$$

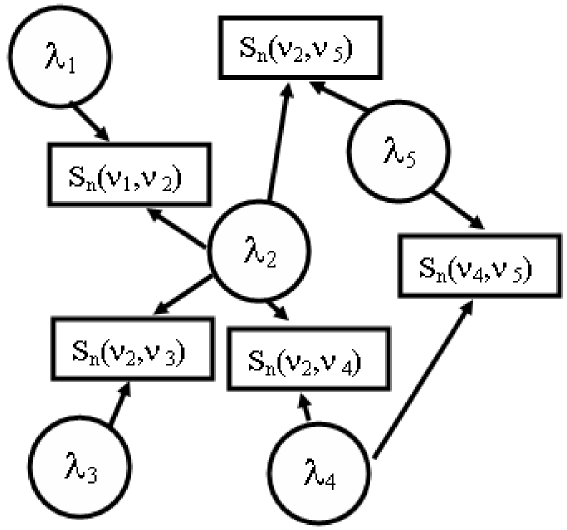
Graphical model of change points



(a) Topology of sensor network (b) Graphical model of random variables

- Conditionally on the shared data sequences, change point variables are no longer independent

Localized stopping times



- Data constraint. Each sensor has access to only shared data with its neighbors
- Definition. Stopping rule for u, denoted by τ_u, is a *localized stopping time*, which depends on measurements of u and its neighbors:
  - for any $t > 0$:

$$\{\tau_u \leq t\} \in \sigma\Big(\{X_n(u), Z_n(u,v) | n \leq t, v \in N(u)\}\Big)$$

Performance metrics

- false alarm rate

$$PFA(\tau_u) = \mathbb{P}(\tau_u \leq \lambda_u).$$

- expected failure detection delay

$$D(\tau_u) = \mathbb{E}[\tau_u - \lambda_u | \tau_u \geq \lambda_u].$$

- Problem: for each sensor u, find a localized stopping time τ_u

$$\min_{\tau_u} D(\tau_u) \text{ such that } PFA(\tau_u) \leq \alpha.$$

Review of results for single change point detection

- optimal sequential rule is a stopping rule by thresholding the posterior of λ_u under some conditions: (Shiryaev, 1978)

$$\tau_u(X) = \inf\{n : \Lambda_n \geq 1 - \alpha\},$$

where

$$\Lambda_n = \mathbb{P}(\lambda_u \leq n | X_1(u), \ldots, X_n(u)).$$

- well-established asymptotic properties (Tartakovsky & Veeravalli, 2006):
  - false alarm:

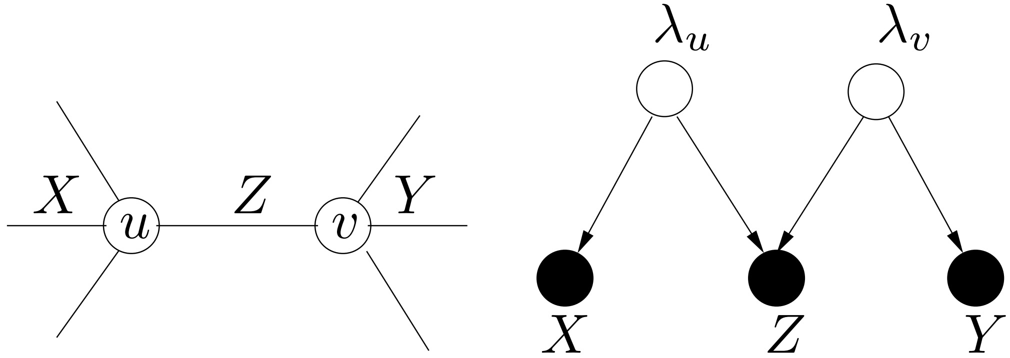
$$PFA(\tau_u(X)) \leq \alpha.$$

  - detection delay:

$$D(\tau_u(X)) = \frac{|\log \alpha|}{q(X) + d}\Big(1 + o(1)\Big) \text{ as } \alpha \to 0.$$

  - here $q(X) = KL(f_1(X)||f_0(X))$, the Kullback-Leibler information, d some constant

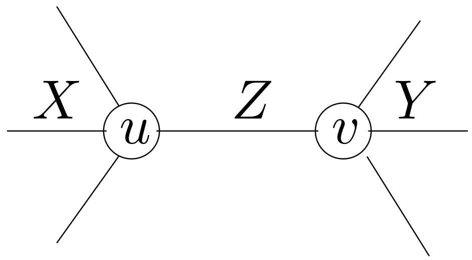
Two sensor case: An initial idea

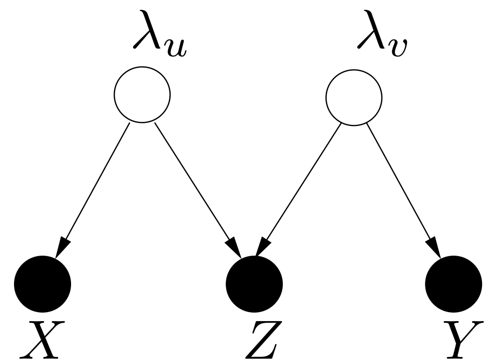


- Idea: use both private data $X_1, \ldots, X_n$ and shared data $Z_1, \ldots, Z_n$:

$$\tau_u(X, Z) = \inf\{n : \mathbb{P}(\lambda_u \leq n | (X_1, Z_1), \ldots, (X_n, Z_n)) \geq 1 - \alpha\}.$$

Two sensor case: An initial idea





- Idea: use both private data $X_1, \ldots, X_n$ and shared data $Z_1, \ldots, Z_n$:

$$\tau_u(X, Z) = \inf\{n : \mathbb{P}(\lambda_u \leq n | (X_1, Z_1), \ldots, (X_n, Z_n)) \geq 1 - \alpha\}.$$

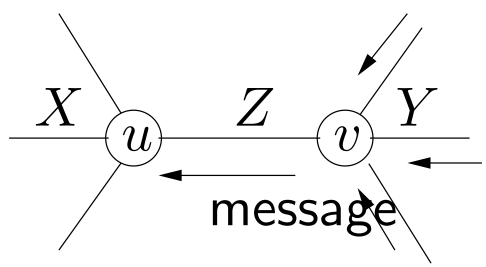
- Theorem 1: The false alarm for $\tau_u(X, Z)$ is bounded from above by α, while expected delay takes the form:

$$D(\tau_u(X, Z)) = \frac{|\log \alpha|}{q(X) + d}\left(1 + o(1)\right) \text{ as } \alpha \to 0.$$

  - Z not helpful in improving the delay (at least in the asymptotics!)
  - this suggests to use information from Y as well (to predict λ_u)

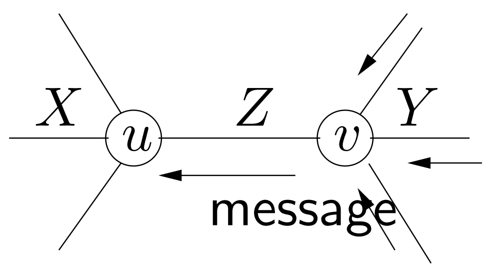
Localized stopping rule with message exchange

- Modified Idea:
  - u should use information given by shared data Z *only if* its neighbor v has not changed (failed) ...
  - but u does not know whether v has changed or not, so ...
  - instead of deciding this by itself, u will wait for v to tell it



Localized stopping rule with message exchange

- Modified Idea:
  - u should use information given by shared data Z *only if* its neighbor v has not changed (failed) ...
  - but u does not know whether v has changed or not, so ...
  - instead of deciding this by itself, u will wait for v to tell it



Stopping rule for u ultimately hinges also information given by data sequence Y, passed to u indirectly via neighbor sensor v

Localized stopping rule with information exchange

- Algorithmic Protocol:
  - each sensor uses all data shared with neighbors that have not declared to change (fail)
  - if a sensor v stops according to its stopping rule, v broadcasts this information to all its neighbors, who promptly drop v from the list of their respective neighbors

Localized stopping rule with information exchange

- Algorithmic Protocol:
  - each sensor uses all data shared with neighbors that have not declared to change (fail)
  - if a sensor v stops according to its stopping rule, v broadcasts this information to all its neighbors, who promptly drop v from the list of their respective neighbors
- Formally, for two sensors:
  - stopping rule for u, using only X: $\tau_u(X)$
  - stopping rule for u, using both X and Z: $\tau_u(X, Z)$
  - similarly, for sensor v: $\tau_v(Y)$ and $\tau_v(Y, Z)$
  - then, the overall stopping rule for u is:

$$\bar{\tau}_u(X, Y, Z) = \begin{cases} \tau_u(X, Z) & \text{if } \tau_u(X, Z) \leq \tau_v(Y, Z) \\ \max(\tau_u(X), \tau_v(Y, Z)) & \text{otherwise} \end{cases}$$

Asymptotic expression of detection delay

(Rajagopal, Nguyen, Ergen & Varaiya, 2010)

Theorem 2: Expected detection delay for u takes the form:

$$D(\bar{\tau}_u) = D_1\delta_\alpha + D_2(1-\delta_\alpha) \text{ as } \alpha \to 0.$$

- here,

$$D_1 = D(\tau_u(X)) = \frac{|\log \alpha|}{q(X)+d}\Big(1+o(1)\Big),$$

$$D_2 = \frac{|\log \alpha|}{q(X)+q(Z)+d}\Big(1+o(1)\Big) \lesssim D_1.$$

- δ_α is the probability that u's neighbor declares "fail" before u.

- clearly, for sufficiently small α there holds: $D(\bar{\tau}_u) < D(\tau_u(X))$. Under additional conditions, this delay is asymptotically optimal.

Upper bound for false alarm rate

Theorem 3: False alarm rate for τ_u satisfies:

$$PFA(\bar{\tau}_u) \leq \alpha + \xi(\bar{\tau}_u).$$

- $\xi(\bar{\tau}_u)$ is termed error-coupling probability: probability that u thinks v has not changed, while in fact, v already has:

$$\xi(\bar{\tau}_u) = \mathbb{P}(\bar{\tau}_u \leq \bar{\tau}_v, \lambda_v \leq \bar{\tau}_u \leq \lambda_u).$$

Upper bound for false alarm rate

Theorem 3: False alarm rate for τ_u satisfies:

$$PFA(\bar{\tau}_u) \leq \alpha + \xi(\bar{\tau}_u).$$

- $\xi(\bar{\tau}_u)$ is termed error-coupling probability: probability that u thinks v has not changed, while in fact, v already has:

$$\xi(\bar{\tau}_u) = \mathbb{P}(\bar{\tau}_u \leq \bar{\tau}_v, \lambda_v \leq \bar{\tau}_u \leq \lambda_u).$$

- Moreover, $\xi(\bar{\tau}_u) \to 0$ at a rate that is faster than α^p for some constant $p > 0$.

- $p > 1$ under conditions that the Kullback-Leibler information given by shared data Z are sufficiently dominated by that of private data X and Y.

Power rate of error-coupling probability

- Define $b = q_0(X) - q_1(Z) + d$ and the rate

$$r_a^* = \frac{1}{w^*} \frac{[\min\{q_0(X), q_1(Z)\} + q_1(Y)]^2}{\max\{\sigma_0^2(X), \sigma_1^2(Z)\} + \sigma_1^2(Y)},$$

where

$$w^* = \sqrt{\frac{\sigma_1^2(X) + \sigma_1^2(Z)}{\max\{\sigma_0^2(X), \sigma_1^2(Z)\} + \sigma_1^2(Y)}}[\min\{q_0(X), q_1(Z)\} + q_1(Y)] - b,$$

constants $\sigma_0^2(X)$, $\sigma_1^2(Z)$ and $\sigma_1^2(Y)$ are variances of the likelihood ratios.

- Then

$$\lim_{\alpha \to 0} \frac{\log \xi(\bar{\tau}_u)}{\log \alpha} \geq p, \text{where}$$

(a) if $b_1 \leq 0$ then $p = r_a^*$;

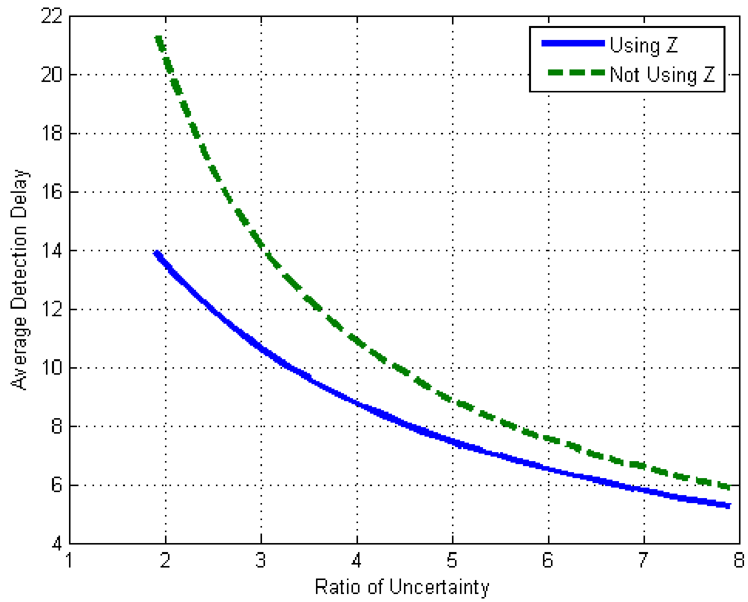
(b) if $b_1 > 0$ then $p = \max(r_a^*, r_b^*)$, where $r_b^* = \frac{4b}{\sigma_1^2(X)+\sigma_1^2(Z)}$.

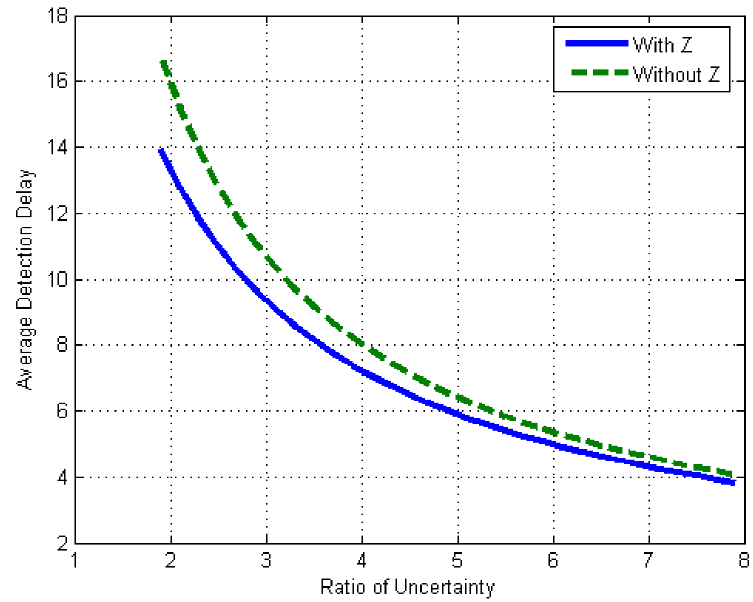
Simulation set-up

- $f_0(X) = \mathcal{N}(1, \sigma^2(X))$; $f_1(X) = \mathcal{N}(0, \sigma^2(X))$
- $f_0(Y) = \mathcal{N}(1, \sigma^2(Y))$; $f_1(Y) = \mathcal{N}(0, \sigma^2(Y))$
- $f_0(Z) = \mathcal{N}(1, \sigma^2(Z))$; $f_1(Z) = \mathcal{N}(0, \sigma^2(Z))$
- Change points λ_1 and λ_2 are endowed with geometric priors and simulated accordingly

Benefits of message-passing with shared data/information

Two-sensor network:





left: evaluated by simulations

right: predicted by Theorem 2

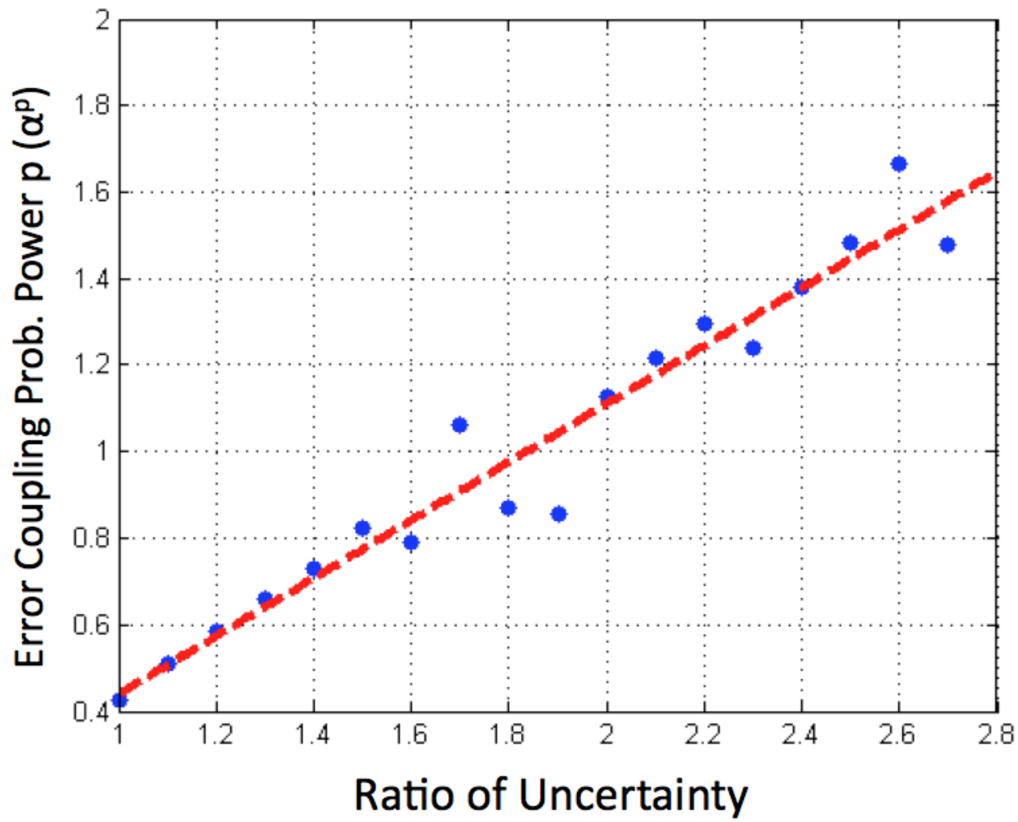
X-axis: Ratio of uncertainty $\sigma^2(Z)/\sigma^2(X)$

Y-axis: Detection delay time

There is extra loss in terms of false alarm probability:

$$PFA(\bar{\tau}_u) \leq \alpha + \alpha^p.$$

where $p > 1$ if $\sigma_Z^2/\sigma_X^2 > 3$ (by Theorem 2).

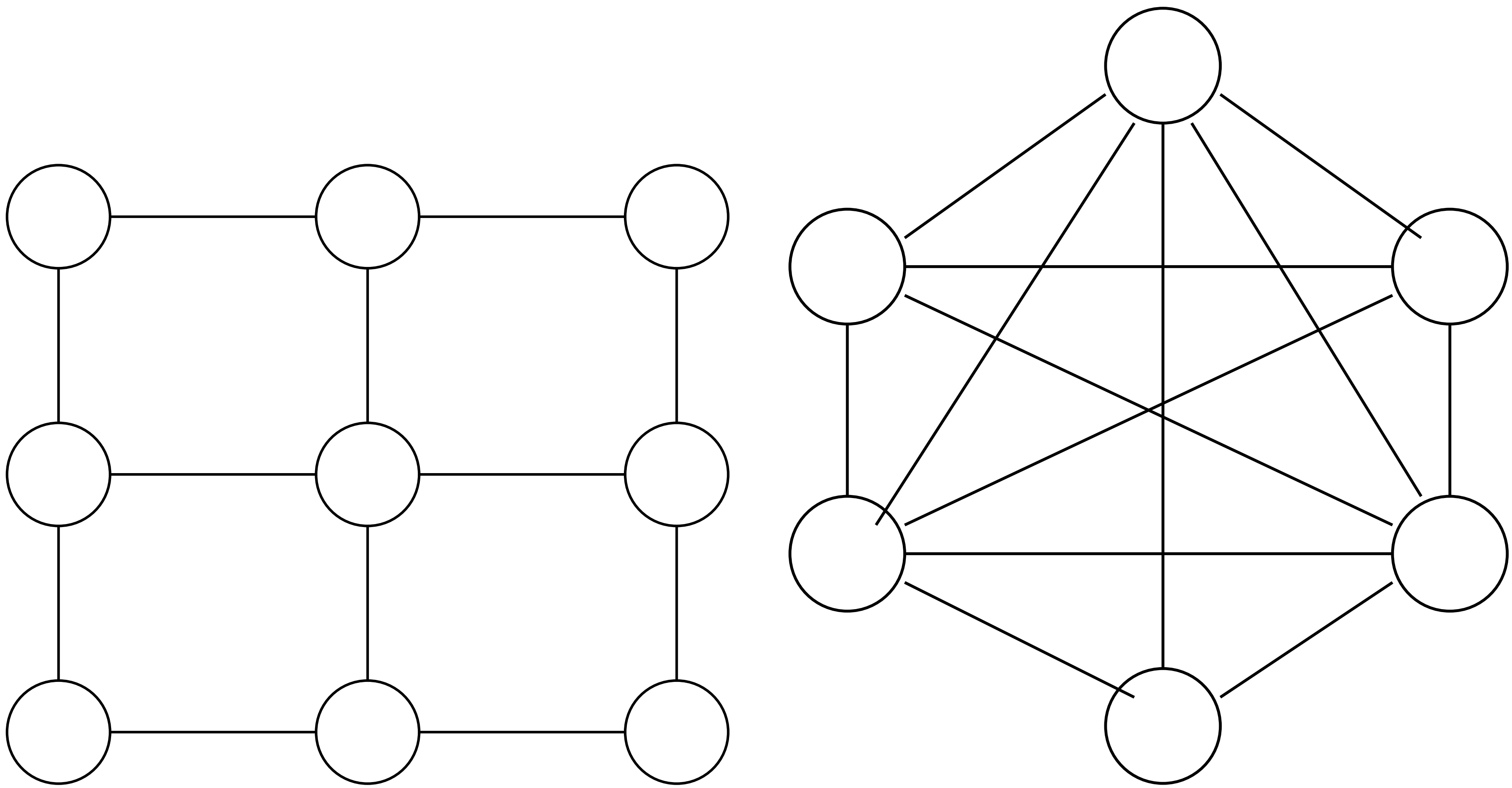


By simulation $p > 1$ if $\sigma_Z^2/\sigma_X^2 > 1.8$.

Network with many sensors

- our algorithmic protocol is readily applicable to network with arbitrarily number of sensors and arbitrary topology
- The Algorithmic Protocol:
  - each sensor uses all data shared with neighbors that have not declared to change (fail)
  - if a sensor v stops according to its stopping rule, v broadcasts this information to all its neighbors, who promptly drop v from the list of their respective neighbors
- asymptotic theory for the false alarm probability remains open
  - comparison of stopping times is intricate

Examples of network topologies

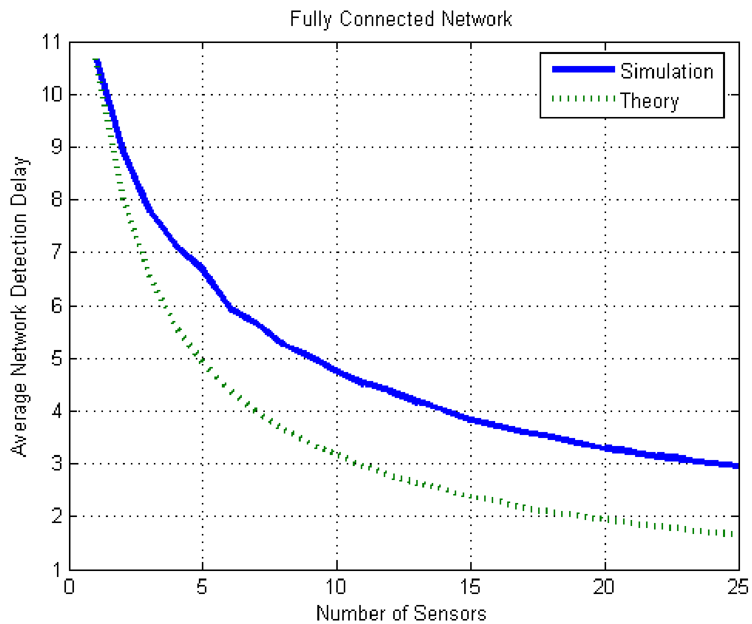


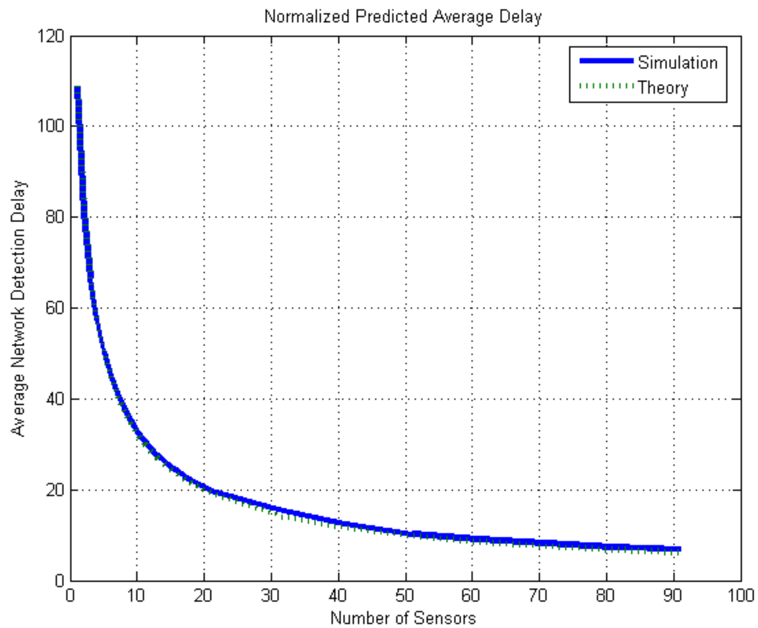
(a) Grid network

(b) Fully connected network

Number of sensors vs Detection delay time

Fully connected network:



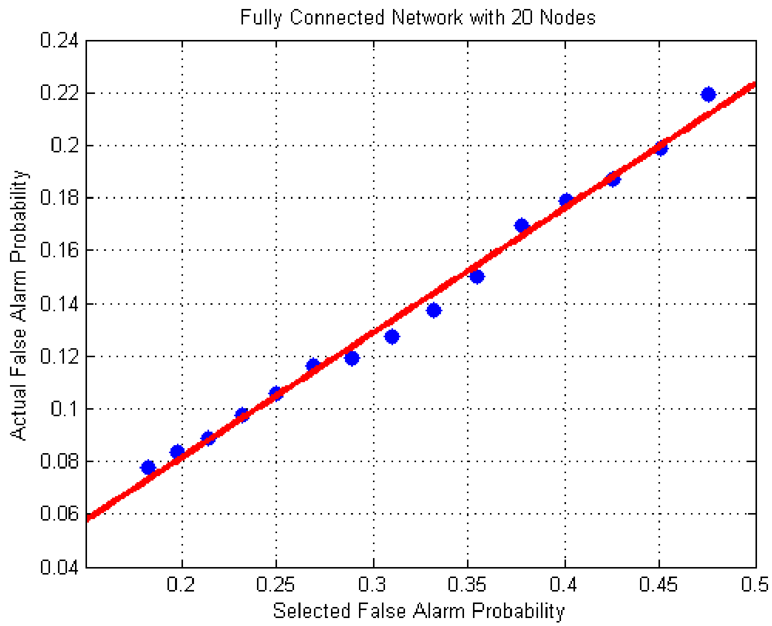


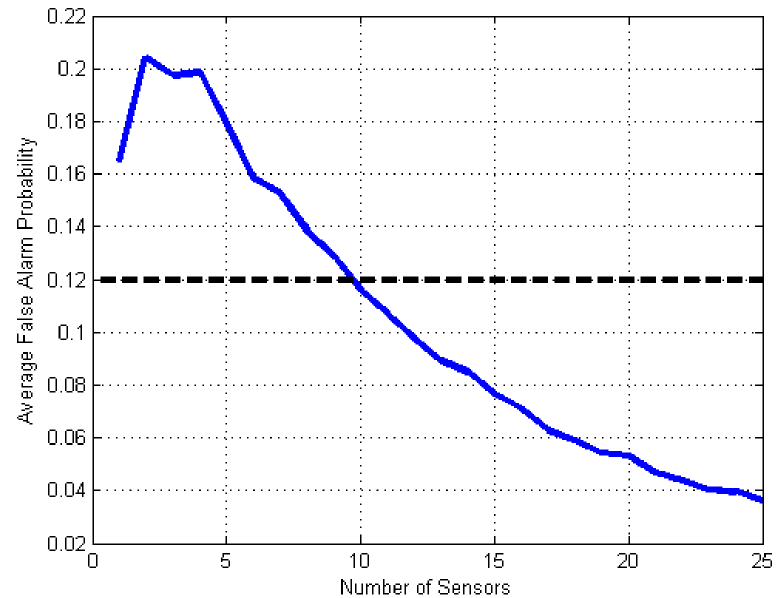
left: $\alpha = .1$

right: $\alpha = 10^{-4}$ (theory predicts well!)

False alarm rates

Fully connected network



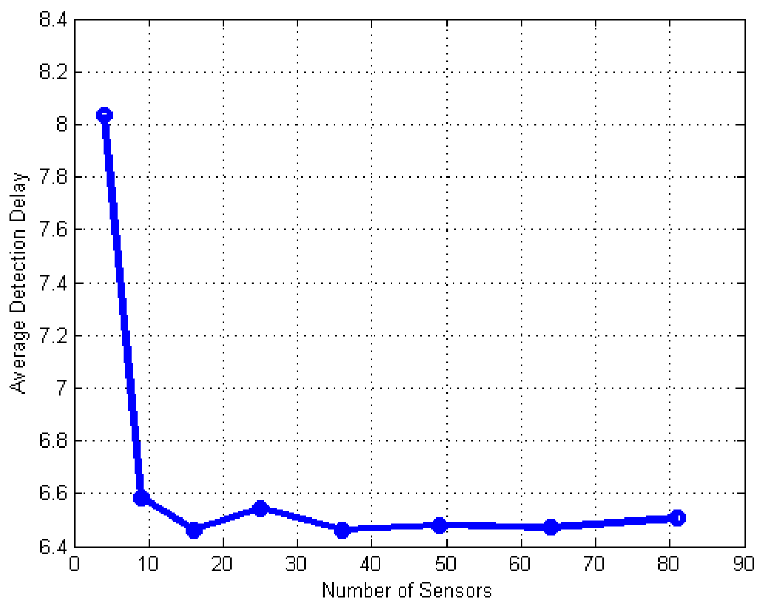


simulated false alarm rate vs. actual rate

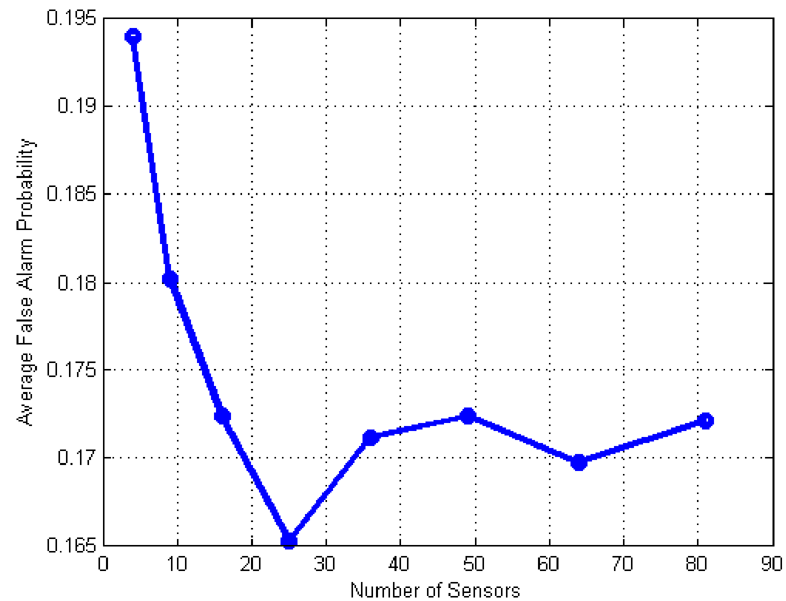
number of sensors vs. actual rate

Effects of network topology

Grid network (each sensor has fixed number of neighbors)



num. of sensors vs. detection delay



num. of sensors vs. actual FA rate

Summary

- decentralized sequential detection of multiple change points
  - application to detection failures in a sensor network

- new statistical formulation drawing from classical ideas:
  - sequential analysis
  - probabilistic graphical models

- introduced a "message-passing" sequential detection algorithm, exploiting the benefit of "network information"

- asymptotic theories for analyzing false alarm rates and detection delay

- Acknowledgement: Ram Rajagopal (Stanford University), Sinem Coleri Ergen (Koc University) and Pravin Varaiya (UC Berkeley)

- for more detail, see
  - R. Rajagopal, X. Nguyen, S.C. Ergen and P. Varaiya. *Simultaneous sequential detection of multiple interacting faults*. http://arxiv.org/abs/1012.1258