Cost-Performance Tradeoff in Multi-hop Aggregation for Statistical Inference

Anima Anandkumar[1] Lang Tong[1] Ananthram Swami[2]
Anthony Ephremides[3]

[1]ECE Dept., Cornell University, Ithaca, NY.

[2]Army Research Laboratory, Adelphi MD.

[3]EE Dept., University of Maryland College Park, MD.

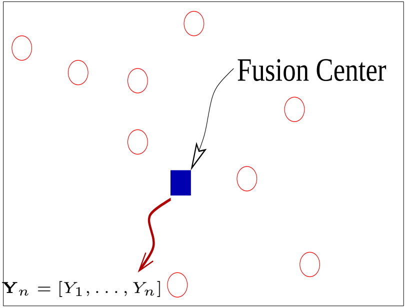
IEEE ISIT 2008
07/07/2008

Supported by Army Research Laboratory CTA.

Distributed Statistical Inference

Sensor Network Applications: Statistical Inference

- Detection, e.g., Target, Pollutant
- Estimation, e.g., Temperature, Pressure



Classical Distributed Inference

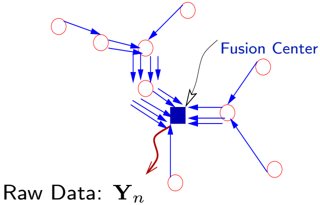
- Sensors: take measurements
- Fusion Center: Final decision

Design Issues

- Sensor Quantization
- Fusion Center Rule

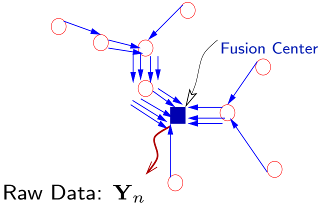
Non-Scalable Nature of Layered Architecture

Routing: Shortest Path

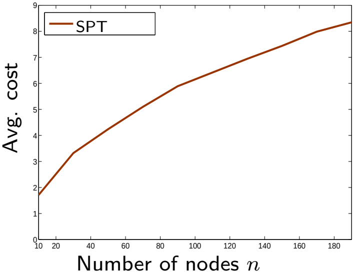


Non-Scalable Nature of Layered Architecture

Routing: Shortest Path

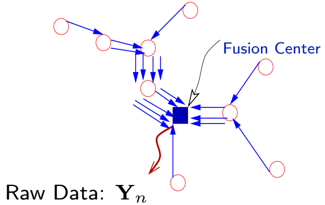


Routing Cost along link (i, j)= dist$(i, j)^2$

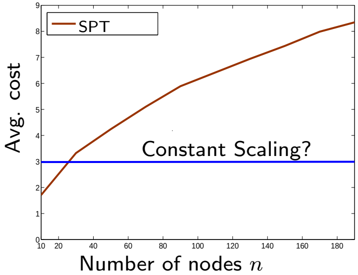


Non-Scalable Nature of Layered Architecture

Routing: Shortest Path

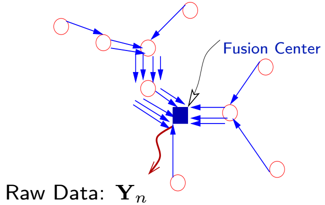


Routing Cost along link $(i, j)=$ dist$(i, j)^2$



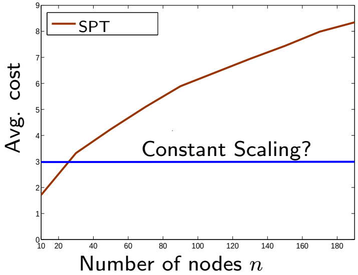
Non-Scalable Nature of Layered Architecture

Routing: Shortest Path



In-Network Data Fusion

Routing Cost along link $(i,j)=$ dist$(i,j)^2$



Sufficient Statistic For Inference

Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistic For Inference

Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistics for Mean Estimation $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$

$\sum_i Y_i$ sufficient to estimate θ: no performance loss

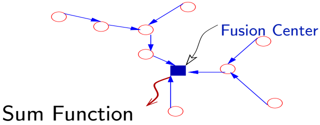
Sufficient Statistic For Inference

Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistics for Mean Estimation $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$

$\sum_i Y_i$ sufficient to estimate θ: no performance loss



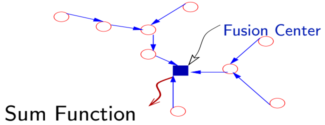
Sufficient Statistic For Inference

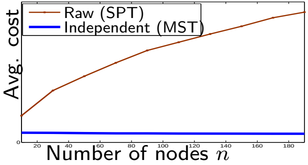
Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistics for Mean Estimation $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$

$\sum_i Y_i$ sufficient to estimate θ: no performance loss





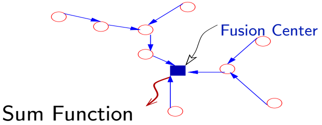
Sufficient Statistic For Inference

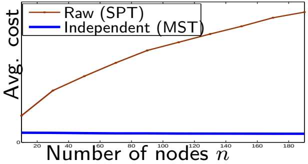
Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistics for Mean Estimation $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$

$\sum_i Y_i$ sufficient to estimate θ: no performance loss





Binary Hypothesis Test: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f(y; \mathcal{H}_0)$ or $f(y; \mathcal{H}_1)$

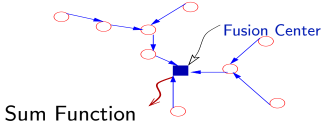
Sufficient Statistic For Inference

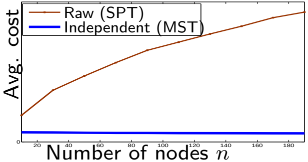
Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistics for Mean Estimation $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$

$\sum_i Y_i$ sufficient to estimate θ: no performance loss





Binary Hypothesis Test: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f(y; \mathcal{H}_0)$ or $f(y; \mathcal{H}_1)$

- $[\sum_i \log f(Y_i; \mathcal{H}_0), \sum_i \log f(Y_i; \mathcal{H}_1)]$ sufficient to decide hypothesis

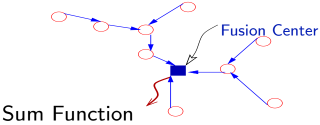
Sufficient Statistic For Inference

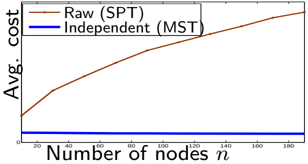
Sufficient Statistic For Inference: No Performance Loss

- Function of raw data: same inference performance at fusion center
- Reduction in dimensionality compared to raw data

Sufficient Statistics for Mean Estimation $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$

$\sum_i Y_i$ sufficient to estimate θ: no performance loss





Binary Hypothesis Test: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f(y; \mathcal{H}_0)$ or $f(y; \mathcal{H}_1)$

- $[\sum_i \log f(Y_i; \mathcal{H}_0), \sum_i \log f(Y_i; \mathcal{H}_1)]$ sufficient to decide hypothesis
- LLR$= \sum_i \log \frac{f(Y_i; \mathcal{H}_0)}{f(Y_i; \mathcal{H}_1)}$ minimally sufficient to decide hypothesis

Minimal Sufficient Statistic

Definition

A statistic is minimally sufficient if every other sufficient statistic is a function of it: Maximum dimensionality reduction

Minimal Sufficient Statistic

Definition

A statistic is minimally sufficient if every other sufficient statistic is a function of it: Maximum dimensionality reduction

Minimal Sufficient Statistic for Binary Hypothesis Testing (Dynkin 61)

$$\text{Log Likelihood Ratio: } \mathsf{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V; \mathcal{H}_0)}{f(\mathbf{Y}_V; \mathcal{H}_1)}$$

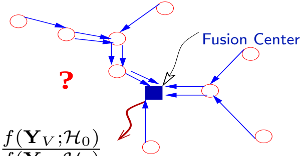
Minimal Sufficient Statistic

Definition

A statistic is minimally sufficient if every other sufficient statistic is a function of it: Maximum dimensionality reduction

Minimal Sufficient Statistic for Binary Hypothesis Testing (Dynkin 61)

$$\text{Log Likelihood Ratio: } \mathsf{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V;\mathcal{H}_0)}{f(\mathbf{Y}_V;\mathcal{H}_1)}$$



$$\mathsf{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V;\mathcal{H}_0)}{f(\mathbf{Y}_V;\mathcal{H}_1)}$$

Extent of Processing?

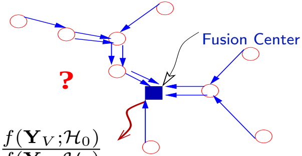
Minimal Sufficient Statistic

Definition

A statistic is minimally sufficient if every other sufficient statistic is a function of it: Maximum dimensionality reduction

Minimal Sufficient Statistic for Binary Hypothesis Testing (Dynkin 61)

$$\text{Log Likelihood Ratio: } \mathsf{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V; \mathcal{H}_0)}{f(\mathbf{Y}_V; \mathcal{H}_1)}$$



$$\mathsf{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V; \mathcal{H}_0)}{f(\mathbf{Y}_V; \mathcal{H}_1)}$$

Extent of Processing?

Fusion Scheme?

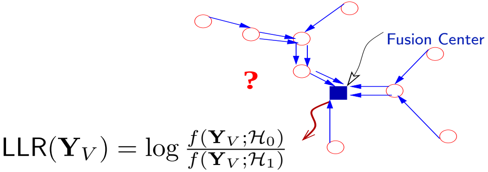
Minimal Sufficient Statistic

Definition

A statistic is minimally sufficient if every other sufficient statistic is a function of it: Maximum dimensionality reduction

Minimal Sufficient Statistic for Binary Hypothesis Testing (Dynkin 61)

Log Likelihood Ratio: $\mathrm{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V; \mathcal{H}_0)}{f(\mathbf{Y}_V; \mathcal{H}_1)}$



Extent of Processing?

Fusion Scheme?

Advantages: Cross-Layer Design

- Raw data not needed at fusion center (sufficient statistic)
- Goal: to achieve optimal inference performance

Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \dfrac{f(\mathbf{Y}_{V_s}; \mathcal{H}_0)}{f(\mathbf{Y}_{V_s}; \mathcal{H}_1)}$

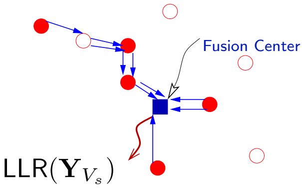
Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \dfrac{f(\mathbf{Y}_{V_s}; \mathcal{H}_0)}{f(\mathbf{Y}_{V_s}; \mathcal{H}_1)}$



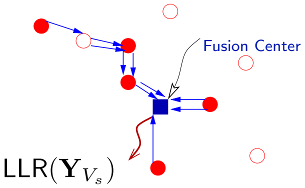
Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \dfrac{f(\mathbf{Y}_{V_s}; \mathcal{H}_0)}{f(\mathbf{Y}_{V_s}; \mathcal{H}_1)}$



Cost & Performance Not Decentralized

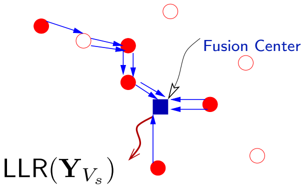
Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \frac{f(\mathbf{Y}_{V_s}; \mathcal{H}_0)}{f(\mathbf{Y}_{V_s}; \mathcal{H}_1)}$



Cost & Performance Not Decentralized

- Multi-Hop Routing & Fusion

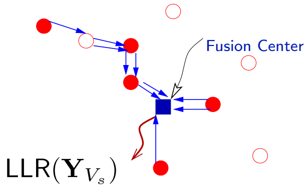
Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \frac{f(\mathbf{Y}_{V_s}; \mathcal{H}_0)}{f(\mathbf{Y}_{V_s}; \mathcal{H}_1)}$



Cost & Performance Not Decentralized

- Multi-Hop Routing & Fusion
- Presence of Correlation

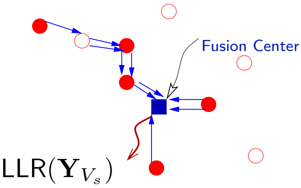
Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \dfrac{f(\mathbf{Y}_{V_s};\mathcal{H}_0)}{f(\mathbf{Y}_{V_s};\mathcal{H}_1)}$



Cost & Performance Not Decentralized

- Multi-Hop Routing & Fusion
- Presence of Correlation
- Computation of Marginal LLR: NP-hard

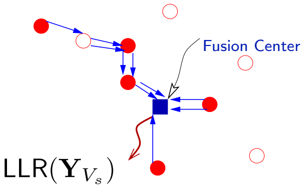
Optimal Node Selection For Tradeoff

Cost-Performance Tradeoff

- Cost $\equiv$ Total Cost of Multi-Hop Routing with Fusion
- Performance $\equiv$ Neyman-Pearson mis-detection error probability

Minimal Sufficient Statistic for a Node Subset $V_s \subset V$

Marginal Log-Likelihood Ratio: $\text{LLR}(\mathbf{Y}_{V_s}) = \log \dfrac{f(\mathbf{Y}_{V_s}; \mathcal{H}_0)}{f(\mathbf{Y}_{V_s}; \mathcal{H}_1)}$



Cost & Performance Not Decentralized

- Multi-Hop Routing & Fusion
- Presence of Correlation
- Computation of Marginal LLR: NP-hard

Brute Force: $2^{|V|}$ Possible Subsets

Outline

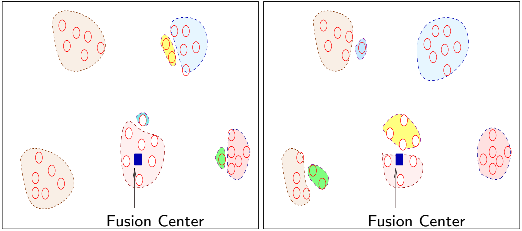
Binary Hypothesis Testing

Null Hypothesis: $\mathcal{C}_0$

$$\mathcal{H}_0 : \prod_{c \in \mathcal{C}_0} f(\mathbf{Y}_c; \mathcal{H}_0)$$

Alternative: $\mathcal{C}_1$

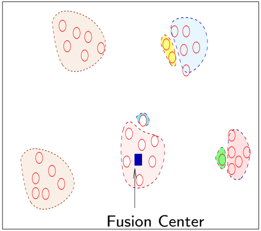
$$\mathcal{H}_1 : \prod_{c \in \mathcal{C}_1} f(\mathbf{Y}_c; \mathcal{H}_1)$$



Binary Hypothesis Testing

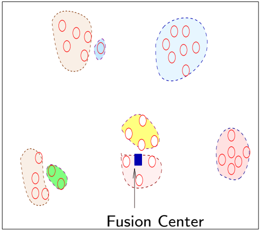
Null Hypothesis: $\mathcal{C}_0$

$$\mathcal{H}_0 : \prod_{c \in \mathcal{C}_0} f(\mathbf{Y}_c; \mathcal{H}_0)$$



Alternative: $\mathcal{C}_1$

$$\mathcal{H}_1 : \prod_{c \in \mathcal{C}_1} f(\mathbf{Y}_c; \mathcal{H}_1)$$



Spatial Correlation Model: Large Inter-Node, Small Intra-Node Distance

Conditionally independent node clusters

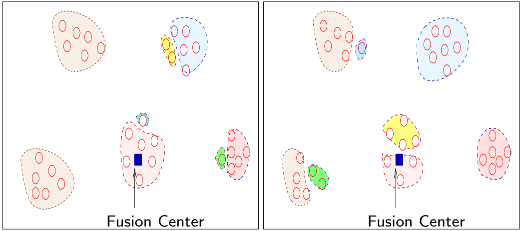
Binary Hypothesis Testing

Null Hypothesis: $\mathcal{C}_0$

$$\mathcal{H}_0 : \prod_{c \in \mathcal{C}_0} f(\mathbf{Y}_c; \mathcal{H}_0)$$

Alternative: $\mathcal{C}_1$

$$\mathcal{H}_1 : \prod_{c \in \mathcal{C}_1} f(\mathbf{Y}_c; \mathcal{H}_1)$$



Spatial Correlation Model: Large Inter-Node, Small Intra-Node Distance

Conditionally independent node clusters

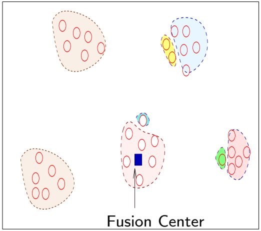
Minimal Sufficient Statistic for Binary Hypothesis Testing

$$\mathrm{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V; \mathcal{H}_0)}{f(\mathbf{Y}_V; \mathcal{H}_1)} = \sum_{\substack{c \in \mathcal{C} \\ \mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1}} \log \frac{f(\mathbf{Y}_c; \mathcal{H}_0)}{f(\mathbf{Y}_c; \mathcal{H}_1)}$$

Binary Hypothesis Testing

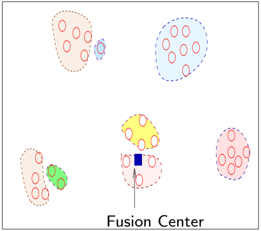
Null Hypothesis: $\mathcal{C}_0$

$$\mathcal{H}_0 : \prod_{c \in \mathcal{C}_0} f(\mathbf{Y}_c; \mathcal{H}_0)$$



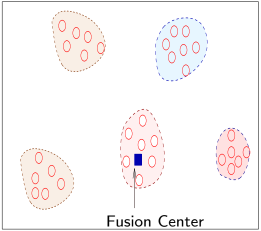
Alternative: $\mathcal{C}_1$

$$\mathcal{H}_1 : \prod_{c \in \mathcal{C}_1} f(\mathbf{Y}_c; \mathcal{H}_1)$$



Log-Likelihood Ratio: $\mathcal{C}_0 \cup \mathcal{C}_1$

$$\sum_{c \in \mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1} \log \frac{f(\mathbf{Y}_c; \mathcal{H}_0)}{f(\mathbf{Y}_c; \mathcal{H}_1)}$$



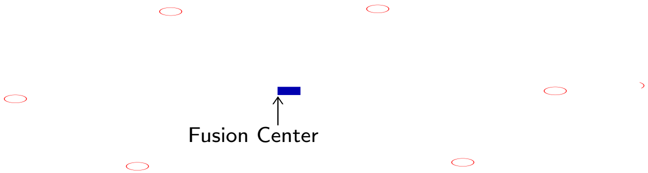
Spatial Correlation Model: Large Inter-Node, Small Intra-Node Distance

Conditionally independent node clusters

Minimal Sufficient Statistic for Binary Hypothesis Testing

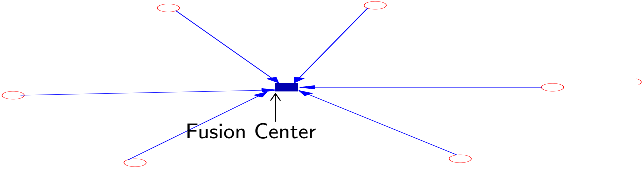
$$\text{LLR}(\mathbf{Y}_V) = \log \frac{f(\mathbf{Y}_V; \mathcal{H}_0)}{f(\mathbf{Y}_V; \mathcal{H}_1)} = \sum_{\substack{c \in \mathcal{C} \\ \mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1}} \log \frac{f(\mathbf{Y}_c; \mathcal{H}_0)}{f(\mathbf{Y}_c; \mathcal{H}_1)}$$

Sensor Selection & Data Fusion



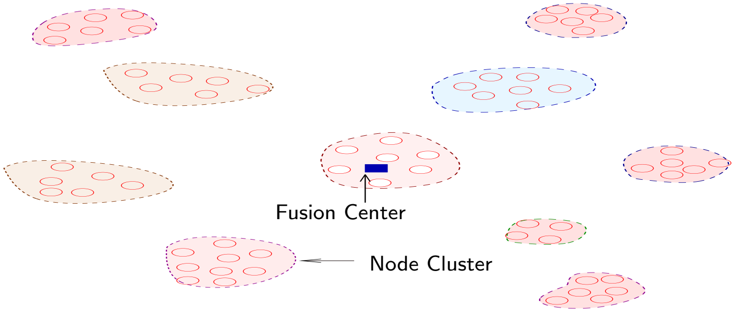
Sensor Selection & Data Fusion

Single Hop with Independent Measurements



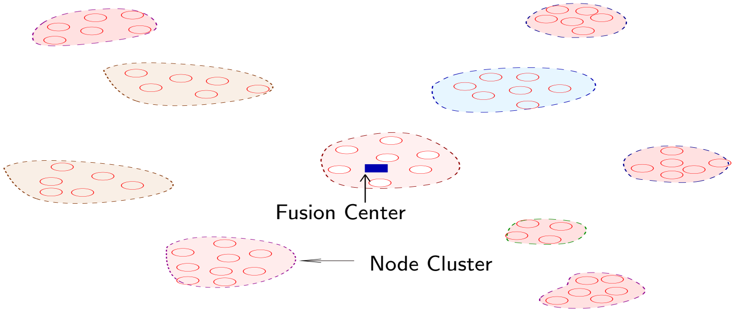
Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements

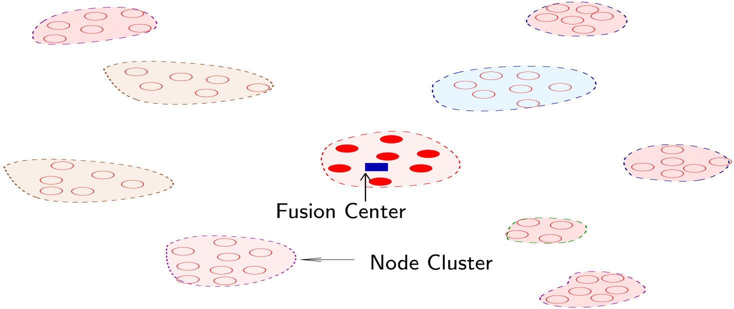


Subset of Node Selection Policies & Fusion Schemes

- Cluster selection : Select all or none of nodes in a clusters

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements

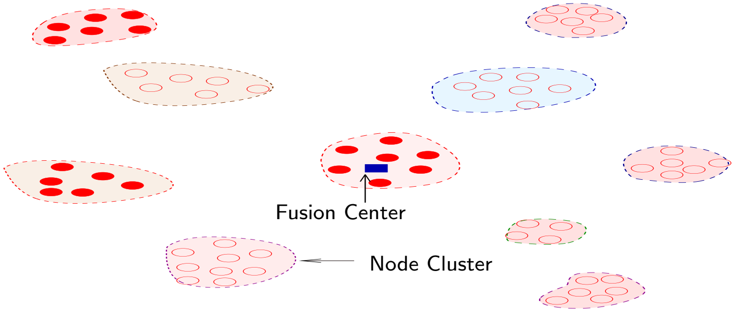


Subset of Node Selection Policies & Fusion Schemes

- Cluster selection : Select all or none of nodes in a clusters

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements

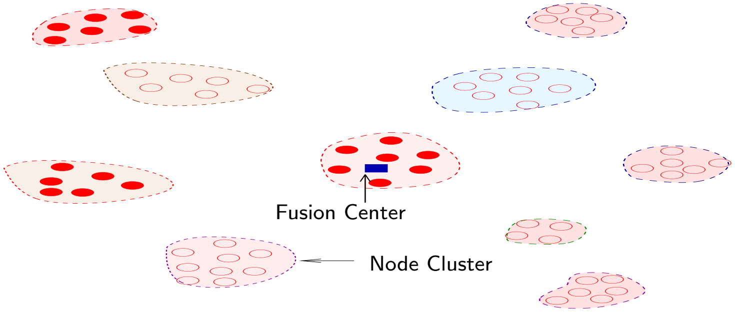


Subset of Node Selection Policies & Fusion Schemes

- Cluster selection : Select all or none of nodes in a clusters

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

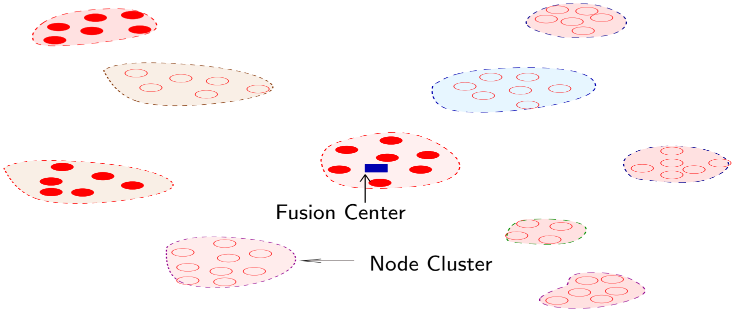
- Cluster selection : Select all or none of nodes in a clusters

Distributed Processing of Cluster : LLR($\mathbf{Y}_c$) for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

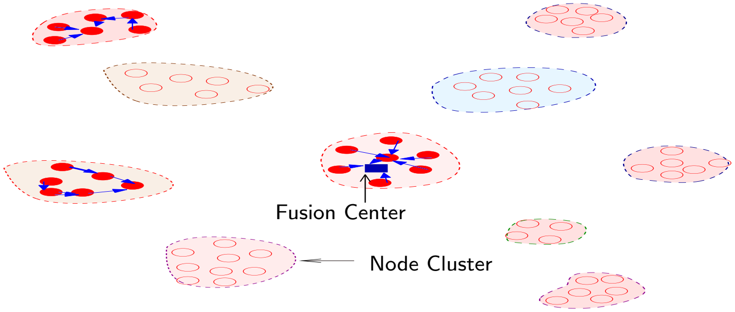
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR$(\mathbf{Y}_c)$ for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

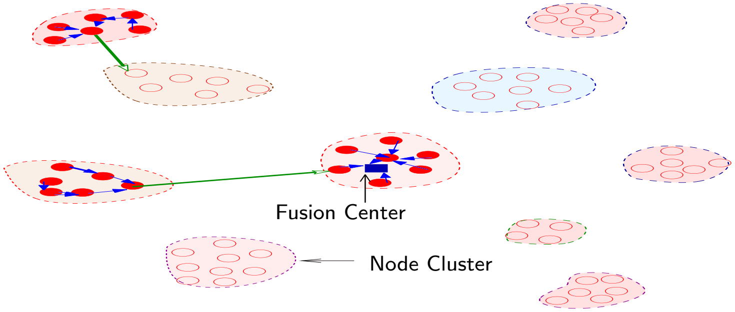
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR$(\mathbf{Y}_c)$ for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

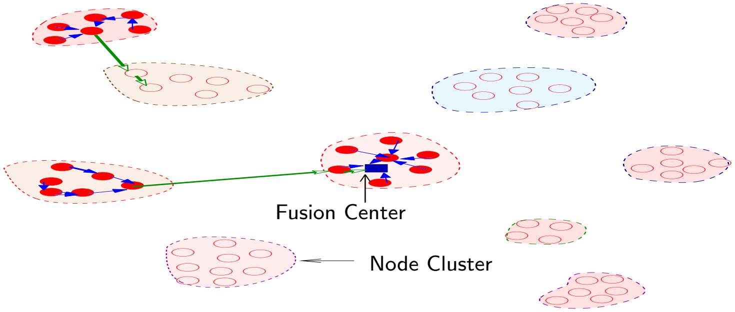
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR$(\mathbf{Y}_c)$ for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

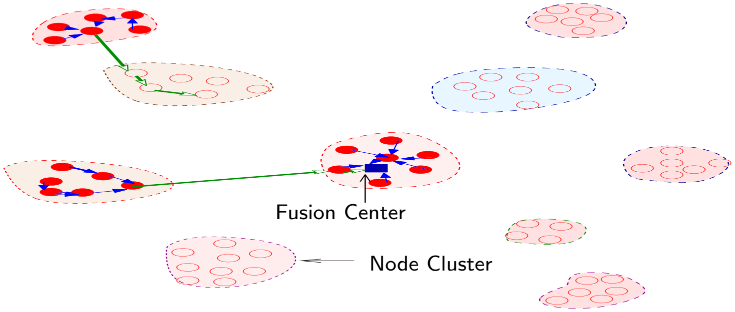
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR($\mathbf{Y}_c$) for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

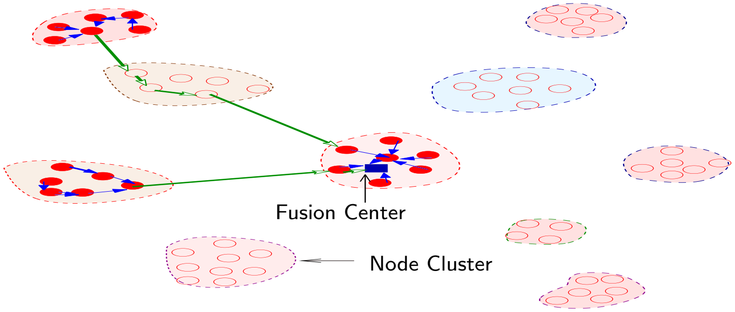
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR$(\mathbf{Y}_c)$ for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

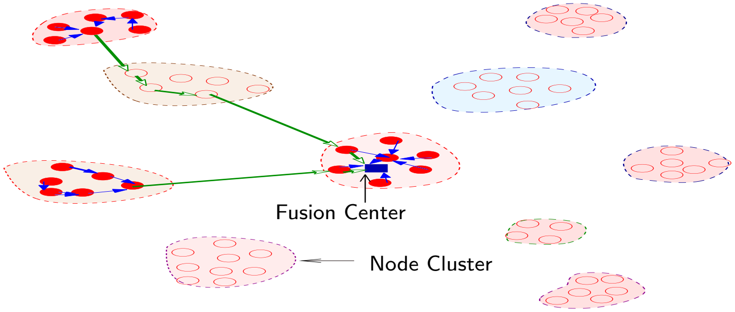
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR$(\mathbf{Y}_c)$ for $c \in \mathcal{C}$

Designate site of computation: Processor of cluster c

Sensor Selection & Data Fusion

Multi Hop with Correlated Measurements



Subset of Node Selection Policies & Fusion Schemes

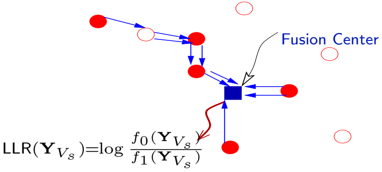
- Cluster selection : Select all or none of nodes in a clusters
- Local processing : processor is a cluster member

Distributed Processing of Cluster : LLR($\mathbf{Y}_c$) for $c \in \mathcal{C}$

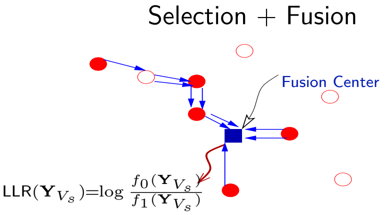
Designate site of computation: Processor of cluster c

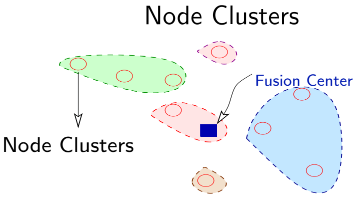
Problem Formulation of In-network Data Fusion

Selection + Fusion

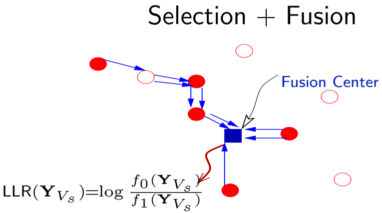


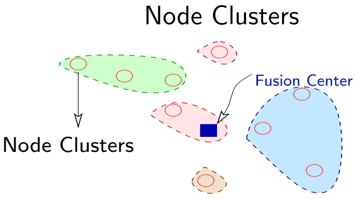
Problem Formulation of In-network Data Fusion





Problem Formulation of In-network Data Fusion





Cost-Performance Tradeoff

Select subset of node clusters for processing s.t. optimal tradeoff between routing costs and fraction of Neyman-Pearson mis-detection probability

$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \mu \log \frac{P_M(V_s)}{P_M(V)} \right], \quad \mu > 0.$$

Summary of Results for Cost-Performance Tradeoff

$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \qquad \mu \log \frac{P_M(V_s)}{P_M(V)} \qquad \right], \quad \mu > 0,$$

Summary of Results for Cost-Performance Tradeoff

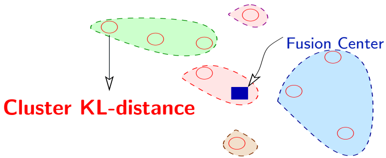
$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \qquad \mu \log \frac{P_M(V_s)}{P_M(V)} \qquad \right], \quad \mu > 0,$$

Neyman Pearson Error Exponent for Node Clusters

$$\mathcal{D} \overset{\Delta}{=} -\,\mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \log P_M(V) = \mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \sum_{c \in \mathcal{C}} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Summary of Results for Cost-Performance Tradeoff

Performance Measure



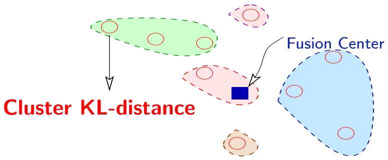
$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \qquad \mu \log \frac{P_M(V_s)}{P_M(V)} \qquad \right], \quad \mu > 0,$$

Neyman Pearson Error Exponent for Node Clusters

$$\mathcal{D} \stackrel{\Delta}{=} -\mathsf{p}\lim_{|V|\to\infty} \frac{1}{|V|} \log P_M(V) = \mathsf{p}\lim_{|V|\to\infty} \frac{1}{|V|} \sum_{c\in\mathcal{C}} D(f(\mathbf{Y}_c;\mathcal{H}_0)||f(\mathbf{Y}_c;\mathcal{H}_1)).$$

Summary of Results for Cost-Performance Tradeoff

Performance Measure



$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \qquad \mu \log \frac{P_M(V_s)}{P_M(V)} \qquad \right], \quad \mu > 0,$$

Neyman Pearson Error Exponent for Node Clusters

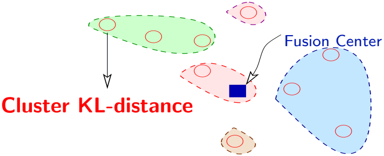
$$\mathcal{D} \stackrel{\Delta}{=} -\,\mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \log P_M(V) = \mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \sum_{c \in \mathcal{C}} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Decentralized Performance Measure: KL-Distance of Cluster

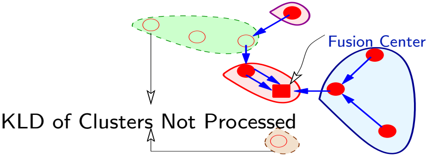
$$\log \frac{P_M(V_s)}{P_M(V)} \approx \sum_{c \in \mathcal{C} \setminus \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Summary of Results for Cost-Performance Tradeoff

Performance Measure



Optimal Tradeoff



$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \qquad \mu \log \frac{P_M(V_s)}{P_M(V)} \qquad \right], \quad \mu > 0,$$

Neyman Pearson Error Exponent for Node Clusters

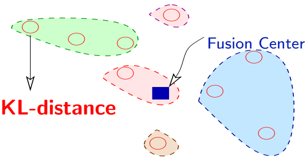
$$\mathcal{D} \overset{\Delta}{=} -\,\mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \log P_M(V) = \mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \sum_{c \in \mathcal{C}} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Decentralized Performance Measure: KL-Distance of Cluster

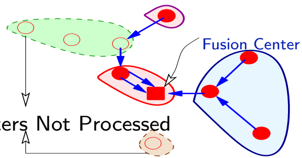
$$\log \frac{P_M(V_s)}{P_M(V)} \approx \sum_{c \in \mathcal{C} \setminus \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Summary of Results for Cost-Performance Tradeoff

Performance Measure



Optimal Tradeoff



$$\min_{V_s \subset V} \left[\mathsf{C}(G(V_s)) + \mu \sum_{c \in \mathcal{C} \setminus \mathcal{C}_s} D(f_0(\mathbf{Y}_c) || f_1(\mathbf{Y}_c)) \right], \quad \mu > 0,$$

Neyman Pearson Error Exponent for Node Clusters

$$\mathcal{D} \stackrel{\Delta}{=} - \mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \log P_M(V) = \mathsf{p} \lim_{|V| \to \infty} \frac{1}{|V|} \sum_{c \in \mathcal{C}} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Decentralized Performance Measure: KL-Distance of Cluster

$$\log \frac{P_M(V_s)}{P_M(V)} \approx \sum_{c \in \mathcal{C} \setminus \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1)).$$

Results for Cost-Performance Tradeoff (cont.)

Results for Cost-Performance Tradeoff (cont.)

Cost-Performance Tradeoff

Select subset of clusters for processing such that optimal sum of routing costs plus KLD of clusters not selected

$$\min_{\mathcal{C}_s \subset \mathcal{C}} [\mathsf{C}(G) + \mu \sum_{c \notin \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))].$$

Results for Cost-Performance Tradeoff (cont.)

Cost-Performance Tradeoff

Select subset of clusters for processing such that optimal sum of routing costs plus KLD of clusters not selected

$$\min_{\mathcal{C}_s \subset \mathcal{C}} [\mathsf{C}(G) + \mu \sum_{c \notin \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))].$$

Prize-Collecting Steiner tree (PCST) Reduction

- PCST expanded graph with scaled cluster KL distance as cluster representative node penalty
- Approximation factor preserving reduction
- Goemans-Williamson approximation algorithm applicable
- Approximation factor of $2 - \dfrac{1}{\#\text{ of profitable clusters} - 1}$

Outline

Neyman Pearson Error Exponent

Neyman-Pearson Detection

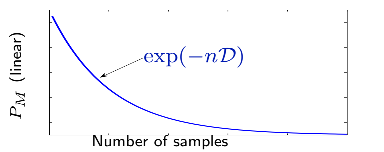
Minimize $P_M = P[\text{Decision} = \mathcal{H}_0 | \mathcal{H}_1]$ s.t. $P_F = P[\text{Decision} = \mathcal{H}_1 | \mathcal{H}_0] \leq \alpha$

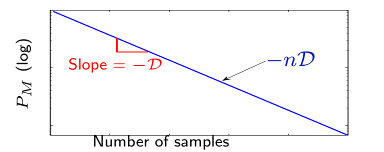
Neyman Pearson Error Exponent

Neyman-Pearson Detection

Minimize $P_M = P[\text{Decision} = \mathcal{H}_0 | \mathcal{H}_1]$ s.t. $P_F = P[\text{Decision} = \mathcal{H}_1 | \mathcal{H}_0] \leq \alpha$

Error Exponent: Rate of Decay of Error Probability





$$P_M \approx e^{-n\mathcal{D}}$$

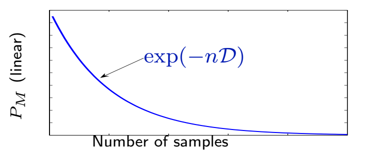
$$\log P_M \approx -n\mathcal{D}$$

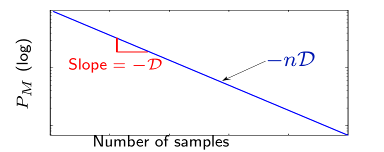
Neyman Pearson Error Exponent

Neyman-Pearson Detection

Minimize $P_M = P[\text{Decision} = \mathcal{H}_0 | \mathcal{H}_1]$ s.t. $P_F = P[\text{Decision} = \mathcal{H}_1 | \mathcal{H}_0] \le \alpha$

Error Exponent: Rate of Decay of Error Probability





$P_M \approx e^{-n\mathcal{D}}$ $\qquad$ $\log P_M \approx -n\mathcal{D}$

NP Error Exponent for Clusters under Uniform Convergence

$$\mathcal{D} = \text{p} \lim_{|V|\to\infty} \frac{1}{|V|} \text{LLR}(\mathbf{Y}_V), \quad \mathbf{Y}_V \sim \mathcal{H}_0 \qquad \text{(Chen '96)}$$

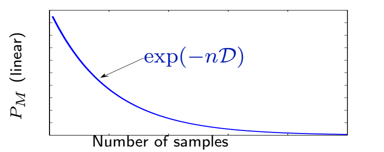
$$\mathcal{D} = \text{p} \lim_{|V|\to\infty} \frac{1}{|V|} \sum_{c\in\mathcal{C}} D(f(\mathbf{Y}_c;\mathcal{H}_0) || f(\mathbf{Y}_c;\mathcal{H}_1))$$

Neyman Pearson Error Exponent

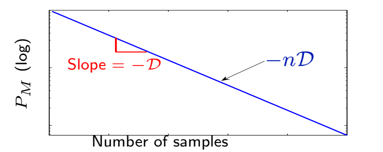
Neyman-Pearson Detection

Minimize $P_M = P[\text{Decision} = \mathcal{H}_0 | \mathcal{H}_1]$ s.t. $P_F = P[\text{Decision} = \mathcal{H}_1 | \mathcal{H}_0] \leq \alpha$

Error Exponent: Rate of Decay of Error Probability



$P_M \approx e^{-n\mathcal{D}}$



$\log P_M \approx -n\mathcal{D}$

NP Error Exponent for Clusters under Uniform Convergence

$$\mathcal{D} = \text{p} \lim_{|V|\to\infty} \frac{1}{|V|} \text{LLR}(\mathbf{Y}_V), \quad \mathbf{Y}_V \sim \mathcal{H}_0 \qquad \text{(Chen '96)}$$

$$\mathcal{D} = \text{p} \lim_{|V|\to\infty} \frac{1}{|V|} \sum_{c \in \mathcal{C}} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))$$

KLD of each cluster is a decentralized performance measure

Cost-Performance Tradeoff for Inference

Problem Statement

Select subset of clusters for processing s.t. optimal tradeoff between routing costs and detection performance

$$\min_{\mathcal{C}_s \subset \mathcal{C}} [\mathsf{C}(G) + \mu \sum_{c \notin \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))]$$

Cost-Performance Tradeoff for Inference

Problem Statement

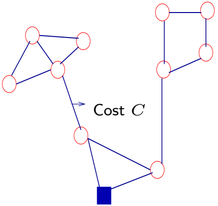
Select subset of clusters for processing s.t. optimal tradeoff between routing costs and detection performance

$$\min_{\mathcal{C}_s \subset \mathcal{C}} [\mathsf{C}(G) + \mu \sum_{c \notin \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))]$$

Network & Communication Model

Connected Network, Bidirectional Links, Unicast Mode

Comm. Graph with Link Costs



Cost-Performance Tradeoff for Inference

Problem Statement

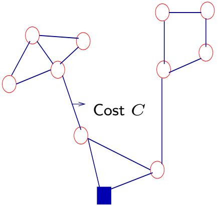
Select subset of clusters for processing s.t. optimal tradeoff between routing costs and detection performance

$$\min_{\mathcal{C}_s \subset \mathcal{C}} [\mathsf{C}(G) + \mu \sum_{c \notin \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))]$$

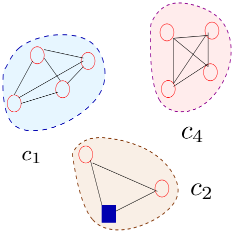
Network & Communication Model

Connected Network, Bidirectional Links, Unicast Mode

Comm. Graph with Link Costs



Node Clusters



Cluster Selection

Cost-Performance Tradeoff for Inference

Problem Statement

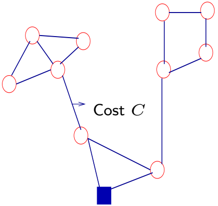
Select subset of clusters for processing s.t. optimal tradeoff between routing costs and detection performance

$$\min_{\mathcal{C}_s \subset \mathcal{C}} [\mathsf{C}(G) + \mu \sum_{c \notin \mathcal{C}_s} D(f(\mathbf{Y}_c; \mathcal{H}_0) || f(\mathbf{Y}_c; \mathcal{H}_1))]$$

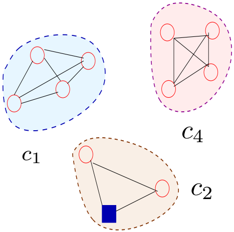
Network & Communication Model

Connected Network, Bidirectional Links, Unicast Mode

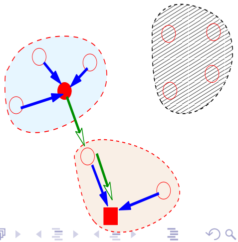
Comm. Graph with Link Costs



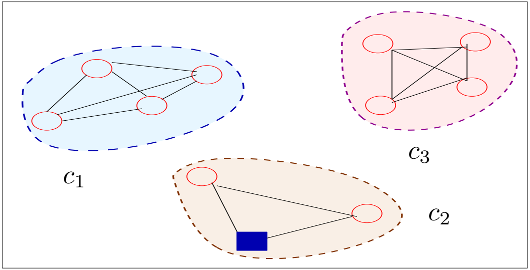
Node Clusters

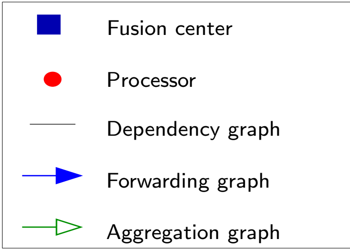


Cluster Selection

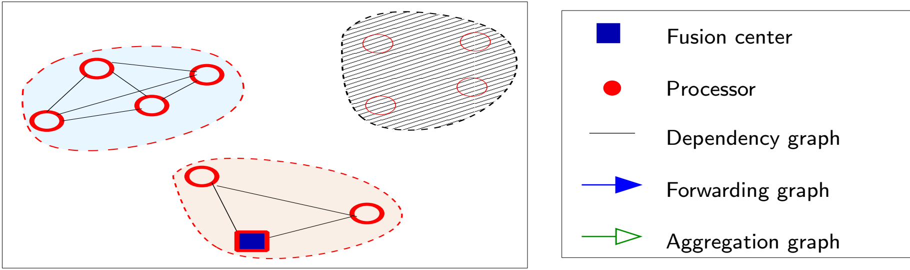


Overview of Fusion Schemes

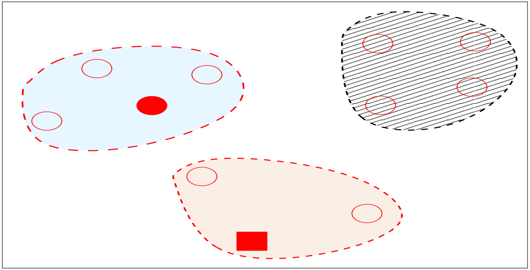


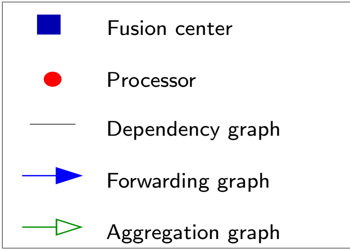


Overview of Fusion Schemes

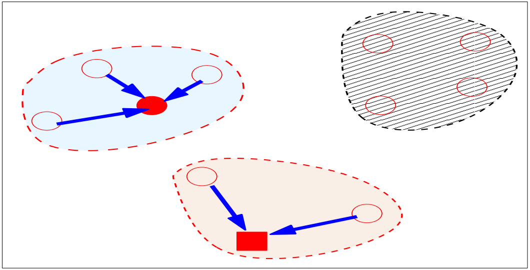


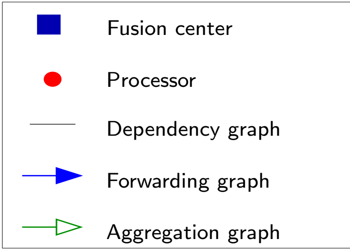
Overview of Fusion Schemes



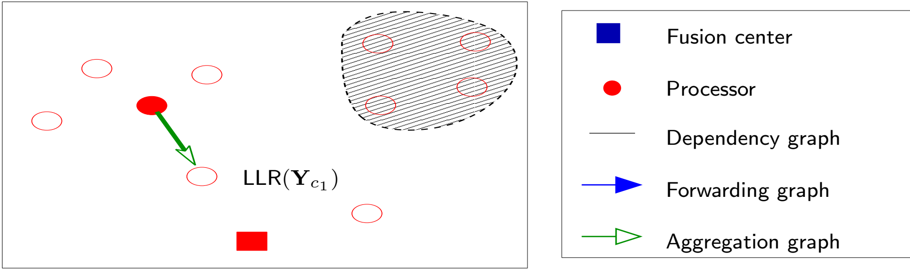


Overview of Fusion Schemes

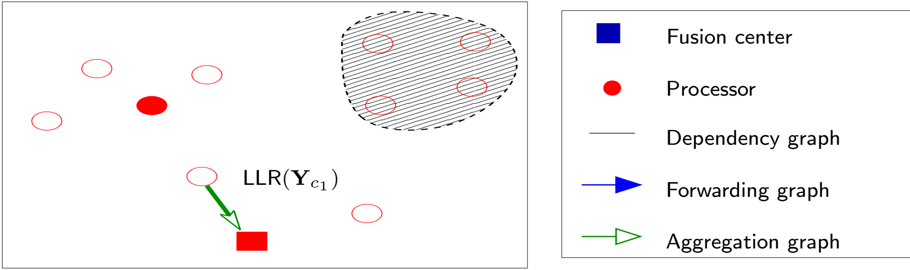




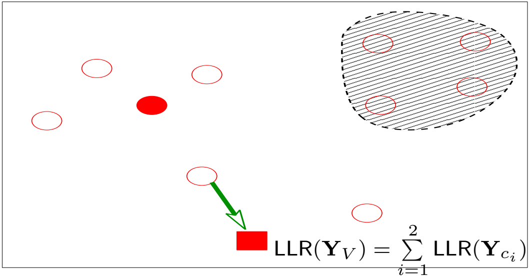
Overview of Fusion Schemes

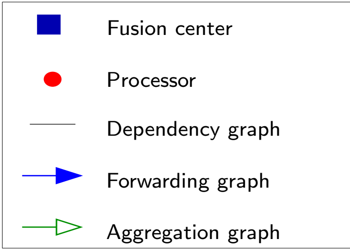


Overview of Fusion Schemes

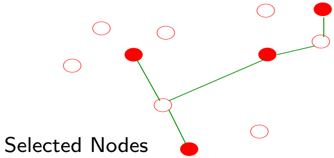


Overview of Fusion Schemes





Prize Collecting Steiner Tree (PCST)



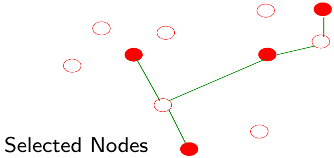
Definition

- Tree with minimum sum edge costs plus node penalties not spanned

$$T_* = \arg \min_{T=(V',E')} \Big[\sum_{e \in E'} c_e + \sum_{i \notin V'} \pi_i \Big].$$

- NP-hard, Goemans-Williamson algorithm has approx. ratio of $2 - \frac{1}{n-1}$

Prize Collecting Steiner Tree (PCST)



Definition

- Tree with minimum sum edge costs plus node penalties not spanned

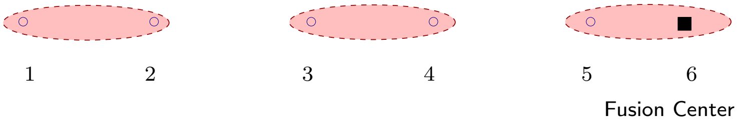
$$T_* = \arg \min_{T=(V',E')} \Big[\sum_{e \in E'} c_e + \sum_{i \notin V'} \pi_i \Big].$$

- NP-hard, Goemans-Williamson algorithm has approx. ratio of $2 - \frac{1}{n-1}$

PCST Reduction

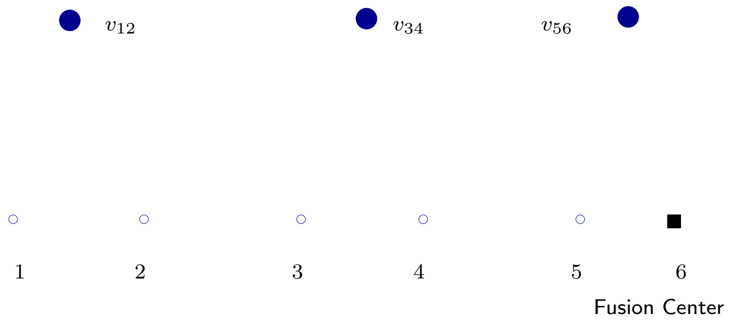
PCST on expanded graph with scaled cluster KLD as penalty

Example : Binary Clusters



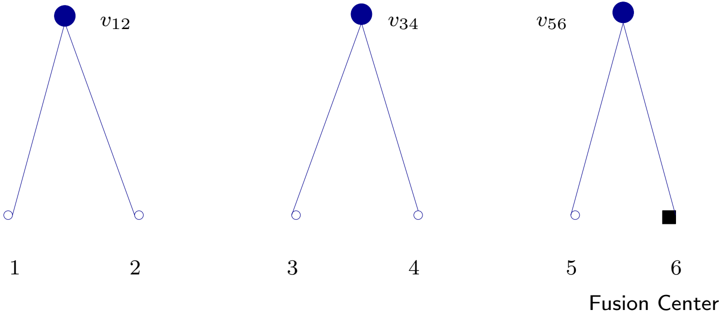
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



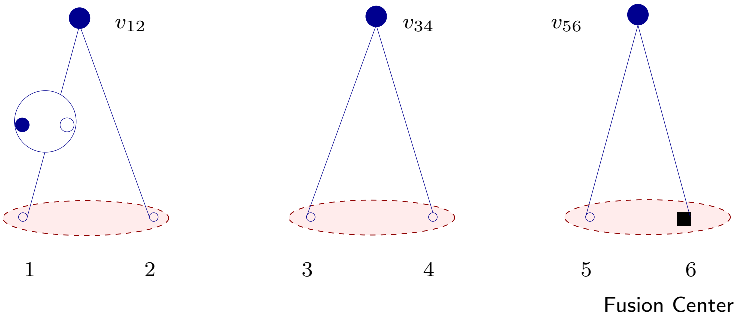
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



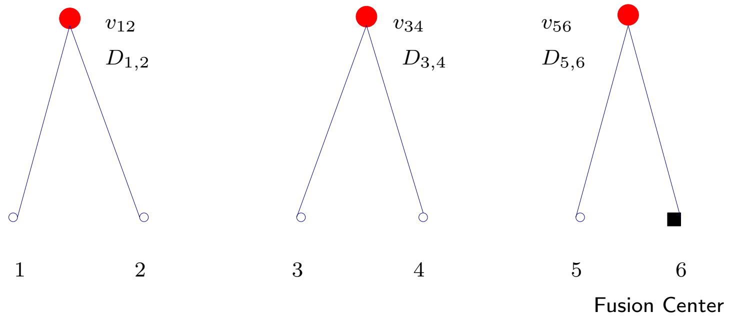
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



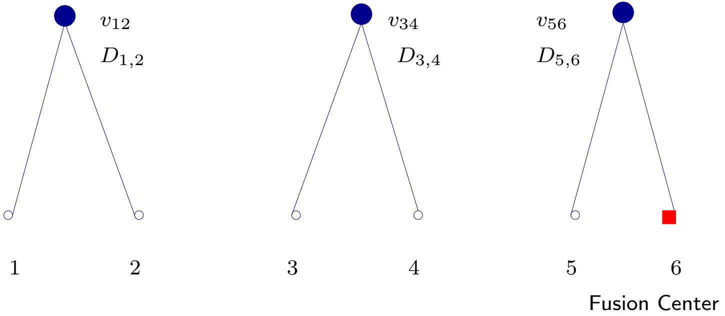
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



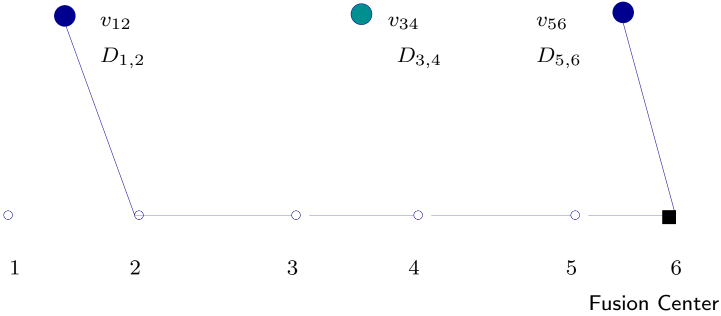
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



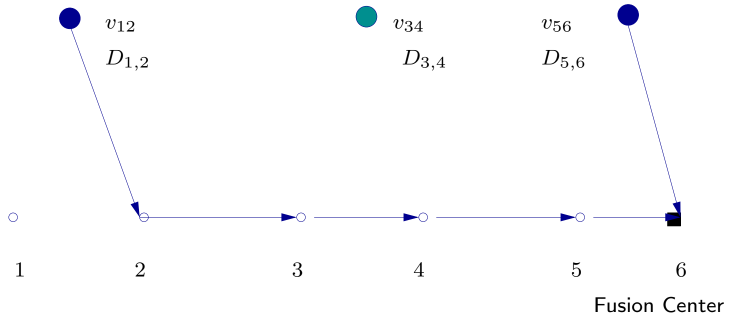
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



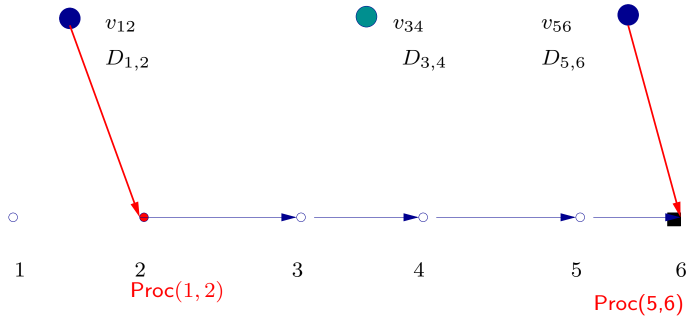
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters

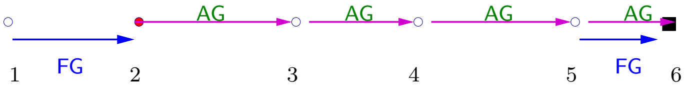


Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters

Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters

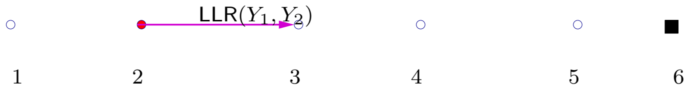


Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters

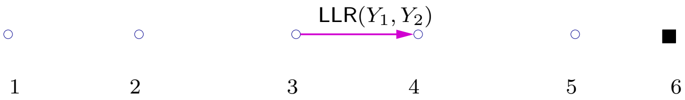
Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



Graph transformation and building prize-collecting Steiner tree.

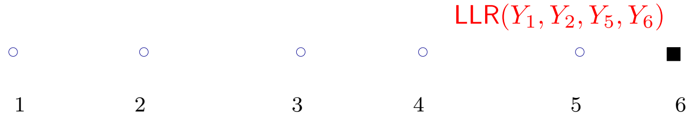
Example : Binary Clusters

Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters

Graph transformation and building prize-collecting Steiner tree.

Example : Binary Clusters



Graph transformation and building prize-collecting Steiner tree.

Outline

Conclusion

Summary of Cluster Based Data Fusion

- Exploit correlation structure for sensor selection and data fusion
- KL-distance of each node cluster as the performance measure
- Prize collecting Steiner tree reduction for cost-performance tradeoff
    - Approximation factor preserving reduction
    - Goemans-Williamson approximation algorithm applicable

Outlook

- Incorporating physical layer issues
    - Effect of interference, Broadcast nature of wireless medium
- General node selection policies

Outline

Results for Minimum Cost Fusion

Minimum Cost Data Fusion for Inference (Infocom '08)

Min total routing costs s.t. likelihood ratio is delivered to fusion center

Results for Minimum Cost Fusion

Minimum Cost Data Fusion for Inference (Infocom '08)

Min total routing costs s.t. likelihood ratio is delivered to fusion center

More General Spatial Correlation Model

Markov random field with dependency graph

AggMST: MST-based Heuristic (Infocom '08)

- Separation of local processor selection and aggregation
- Approximation Ratio of 2 for Nearest-Neighbor Dependency

Results for Minimum Cost Fusion

Minimum Cost Data Fusion for Inference (Infocom '08)

Min total routing costs s.t. likelihood ratio is delivered to fusion center

More General Spatial Correlation Model

Markov random field with dependency graph

AggMST: MST-based Heuristic (Infocom '08)

- Separation of local processor selection and aggregation
- Approximation Ratio of 2 for Nearest-Neighbor Dependency

Constant Average Cost Scaling for AggMST (in progress)

k-NNG and disk dependency in random large constant density networks

Steiner Tree Reduction (Infocom '08)

- Joint design of local processors and aggregation
- Optimal Cost is given by Steiner tree on expanded graph

Results for Minimum Cost Fusion

Minimum Cost Data Fusion for Inference (Infocom '08)

Min total routing costs s.t. likelihood ratio is delivered to fusion center

More General Spatial Correlation Model

Markov random field with dependency graph

AggMST: MST-based Heuristic (Infocom '08)

- Separation of local processor selection and aggregation
- Approximation Ratio of 2 for Nearest-Neighbor Dependency

Constant Average Cost Scaling for AggMST (in progress)

k-NNG and disk dependency in random large constant density networks

Steiner Tree Reduction (Infocom '08)

- Joint design of local processors and aggregation
- Optimal Cost is given by Steiner tree on expanded graph

Optimal Node Density for Inference

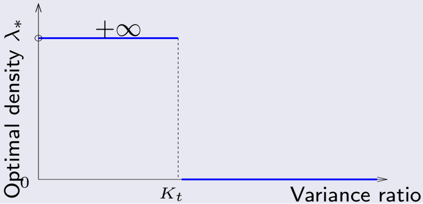
Optimal Node Density of Random Placement

Maximize error exponent subject to average routing cost constraint

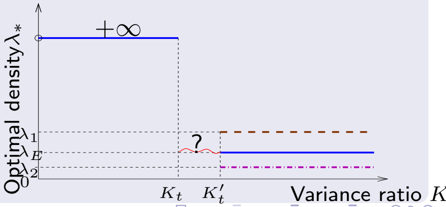
Large random networks (ITsub '06, SP '08)

- Closed-form error exponent and optimal node density
- Threshold effect on optimal node density
- Law of large numbers for graph functionals

No Energy constraint



Feasible Average Energy



References

Minimum Cost Data Fusion

- A. Anandkumar, L. Tong, A. Swami, and A. Ephremides, "*Minimum Cost Data Aggregation with Localized Processing for Statistical Inference*," in Proc. IEEE Infocom 08, April 2008
- A. Anandkumar, A. Ephremides, L. Tong, and A. Swami, "*Minimum Cost Routing with Local Processing for Distributed Statistical Inference*," in handbook on Array Processing and Sensor Networks, (S. Haykin and R. Liu, eds.), 2008.

Aggregation and Error Exponents in Large Random Networks

- A. Anandkumar, L. Tong, and A. Swami, "*Detection of Gauss-Markov random fields with nearest-neighbor dependency*," sub. to IEEE Tran. Information Theory, Jan.07
- A. Anandkumar, L. Tong, and A. Swami, "*Optimal Node Density for Detection in Energy Constrained Random Networks*," accepted to IEEE Tran. Signal Processing, Oct. 2007

Related Work

Correlated Data Gathering (Cristescu et al. 06, Scaglione&Servetto 02)

- Raw data not needed at fusion center
- Only the likelihood function for optimal inference

In-network Function Computation (Giridar & Kumar 06)

- Valid for symmetric functions
- LLR has this form only for independent data

Routing for Inference: For Special Correlation Models

- Independent Measurements (Yang & Blum, Yu & Ephremides)
- 1-D Gauss-Markov process (Sung et al. 06, Chamberland & Veeravalli)

Routing for Belief Propagation (Kreidl & Willsky 06, Williams et al. 05)

Local MAP estimate of raw data at each node: not global decision at FC