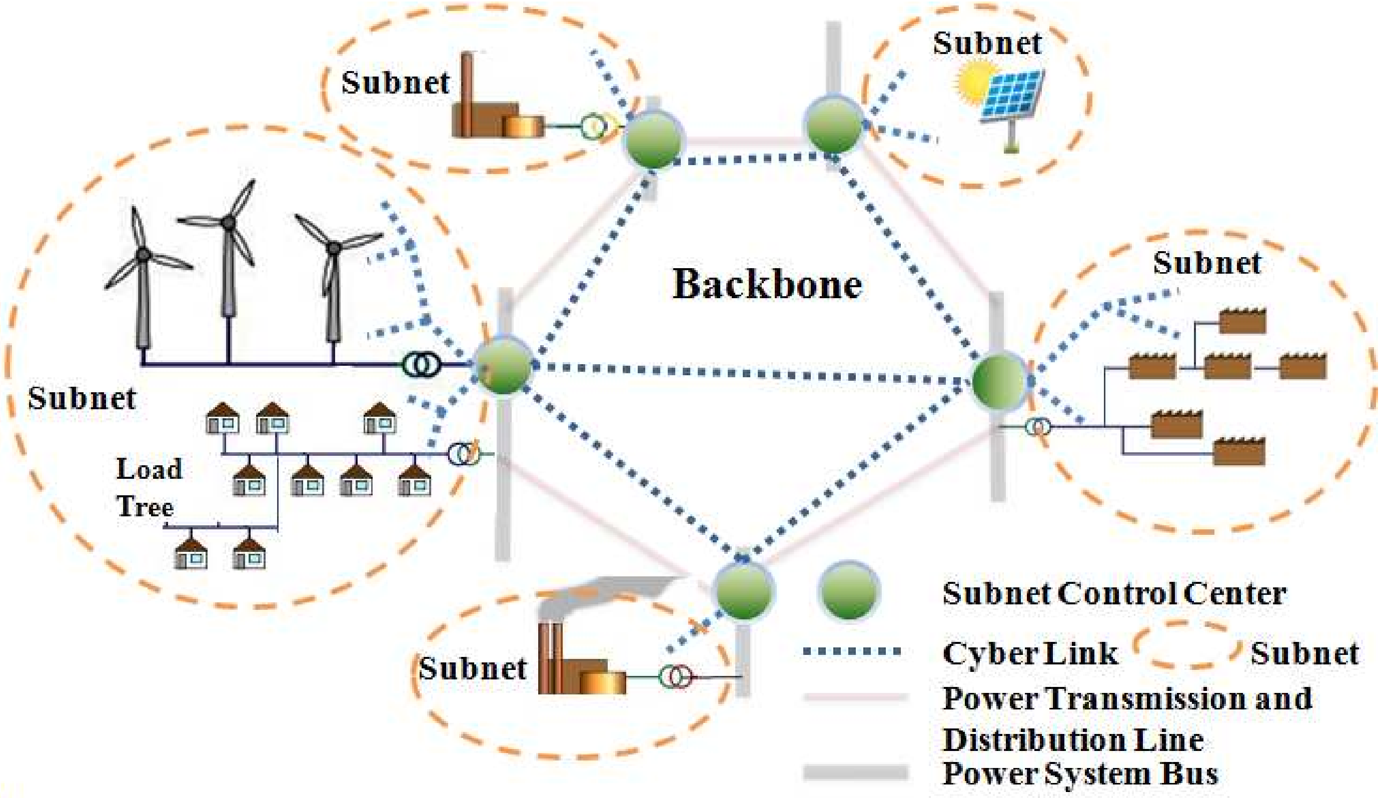
Distributed Estimation and Detection for Smart Grid

Shuguang (Robert) Cui

Texas A&M University

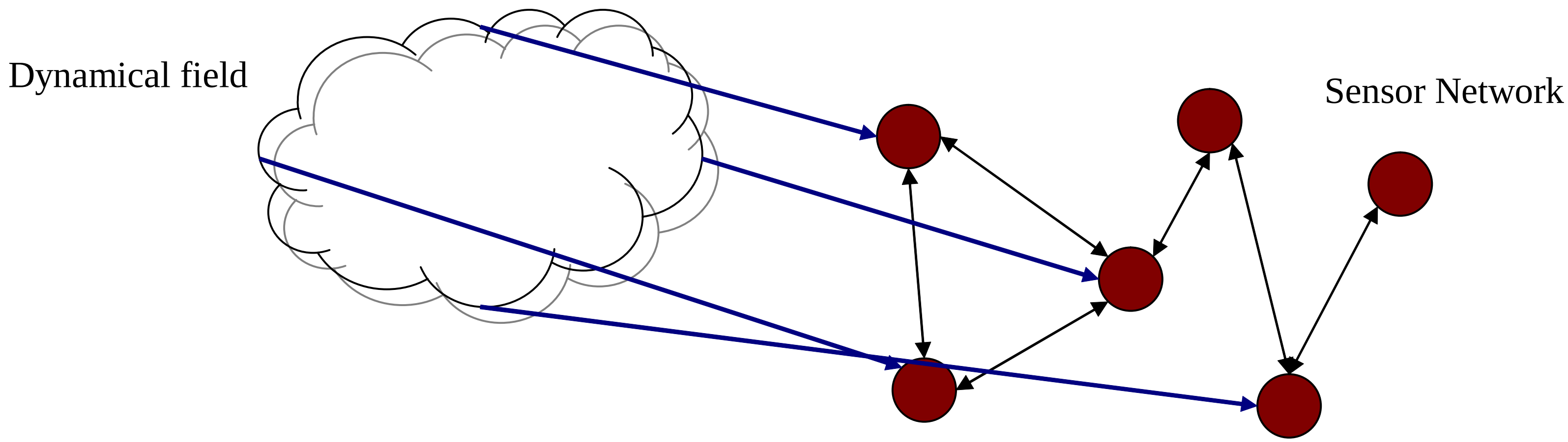
- Joint Work with: S. Kar (CMU), R. Tandon (Princeton), H. V. Poor (Princeton), and J. M. F. Moura (CMU)

Distributed Estimation/Detection in Smart Grid



Estimation: Large-scale power system state
Detection: Attack or anomaly in the system

Distributed Estimation over Linear Dynamical Systems



Field/Signal Process: $\{\mathbf{x}_k\} \in \mathbb{R}^M$

$$\mathbf{x}_{(k+1)\Delta} = \mathcal{F}\mathbf{x}_{k\Delta} + \mathbf{w}_k$$

Local observation at sensor n: $\{\mathbf{y}^n_{k\Delta}\} \in \mathbb{R}^{m_n}$

$$\mathbf{y}^n_{k\Delta} = \mathcal{C}_n \mathbf{x}_{k\Delta} + \mathbf{v}^n_k$$

- Applications: Smart Grid, Social Network, etc...

Assumptions on the Signal Model

We impose the following global assumptions on the signal/observation model:

- **Stabilizability-Assumption (S.1)**: The pair $(\mathcal{F}, \mathcal{Q}^{1/2})$ is stabilizable. The non-degeneracy (positive definiteness) of $\mathcal{Q}$ ensures this.

- **Global Detectability-Assumption (D.1)**: The pair $(\mathcal{C}, \mathcal{F})$ is detectable, where $\mathcal{C} = [\mathcal{C}_1^T \cdots \mathcal{C}_N^T]^T$.

Remark 1. *We do not assume local detectability at each agent. These assumptions are required even by the centralized estimator to achieve a stable estimation error. We later show that under rather weak conditions on the inter-sensor communication, the global detectability assumption is also sufficient for our distributed scheme to achieve stable estimation errors at each sensor.*

Distributed Filtering Architecture

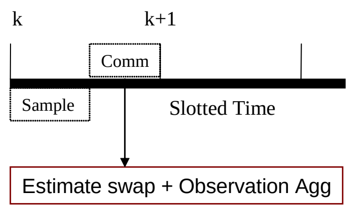
- The agents (sensors) need to collaborate for *successful* estimation of $\{\mathbf{x}_k\}$.

- Basic Sensing and Communication Architecture:
  - Inter-agent communication rate is measured as the link activation rate
  - Inter-agent communication is assumed perfect

- Communication is constrained by the network topology; and only $\overline{\gamma} > 0$ rounds of inter-agent message passing may be allowed per epoch $(k\Delta, (k+1)\Delta]$ on an average.

M-GIKF Scheme

The overall communication rate $\overline{\gamma}$ is split over two phases:



- **Estimate Swapping**:
  - Agents perform pairwise swapping of previous epoch's estimates (if the corresponding link is active)
  - *Estimate Swapping guarantees asymptotic stability of local estimation errors under weak global detectability assumptions*

- **Observation Aggregation**:
  - Agents use the remaining communication to disseminate their instantaneous observations to the neighbors
  - *Observation Aggregation improves estimation performance*

Communication Rate Constraint

Define:

- $M^e(k)$: Number of sensor communications at the k-th epoch for estimate swapping
- $M^o(k)$: Number of sensor communications at the k-th epoch for observation aggregation

Communication Constraint: Since $\overline{\gamma} > 0$ is given, we require:

$$\limsup_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} (M^e(k) + M^o(k)) \leq \overline{\gamma} \text{ a.s.}$$

GS Protocol: Estimate Swapping

Connectivity - Assumption (C.1): The maximal graph $(V, \mathcal{E})$ is connected, where V denotes the set of N agents and $\mathcal{E}$ the set of allowable links.

Proposition 1. *Under (C.1) we have the following:*

- *The sequence $\{A^e(k)\}$ of estimate swapping adjacency matrices is independent and identically drawn from a special distribution $\mathcal{D}$*
- *The mean adjacency matrix $\overline{A}^e$ is irreducible and aperiodic*
- *The average number of inter-sensor communications due to estimate swapping satisfies almost surely (a.s.)*

$$\overline{M}^e = \lim_{k \to \infty} (1/k) \sum_{i=0}^{k-1} M^e(k) < \overline{\gamma}/2$$

GS Protocol: Observation Aggregation

Adjacency matrix sequence drawn following a Poisson process of rate $\overline{\gamma}/2$

Proposition 2. *Under* ***(C.1)*** *we have the following:*

- *The average number of inter-agent communications due to observation aggregation per epoch is* $\overline{\gamma}/2$.
- *The sequence* $\{\mathcal{I}_k^n\}$ *denotes the set of observations (w.r.t. node indices) available at sensor* n *at the end of the epochs; then for every epoch* k *and every sensor* n,

$$\mathbb{P}(\mathcal{I}_k^n = [1, \cdots, N]) > 0$$

Reasonable Communication Scheme

Definition 1. *A communication scheme (for estimate swapping and observation aggregation) is said to be* reasonable *if*

(E.1) *The estimate swapping adjacency matrices $\{A^e(k)\}$ are i.i.d. The mean matrix $\overline{A}$ is doubly stochastic, irreducible, and aperiodic.*

(E.2) *The sequences $\{\mathcal{I}_k^n\}$ are i.i.d. for each n, independent of the estimate swapping and satisfy $\mathbb{P}(\mathcal{I}_k^n = [1, \cdots, N]) > 0$.*

(E.3) *The average number of inter-sensor communications (including both the estimate swapping and observation aggregation steps) per epoch is less than or equal to $\overline{\gamma}$, where $\overline{\gamma} > 0$ is a predefined upper bound on the communication rate.*

Remark 2. *The previous GS protocol is a reasonable scheme under* **(C.1)**.

M-GIKF: Estimate Update

Define:

- $(\widehat{\mathbf{x}}^n_{k|k-1}, \widehat{P}^n_k)$: Estimate (state) at sensor n of $\mathbf{x}_k$ based on *information* till time $k-1$
- $n_k^{\rightarrow}$: Neighbor of sensor n at time k w.r.t. $A^e(k)$.

  With *estimate swapping*: swap $\left(\widehat{\mathbf{x}}^n_{k|k-1}, \widehat{P}^n_k\right)$ and $\left(\widehat{\mathbf{x}}^{n_k^{\rightarrow}}_{k|k-1}, \widehat{P}^{n_k^{\rightarrow}}_k\right)$;

  With *observation aggregation*: sensor n collects: $\mathbf{y}_k^{\mathcal{I}_n^k}$

<u>**Update Rule**</u>:

$$\widehat{\mathbf{x}}^n_{k+1|k} = \mathbb{E}\left[\mathbf{x}_{k+1} \;\middle|\; \widehat{\mathbf{x}}^{n_k^{\rightarrow}}_{k|k-1}, \widehat{P}^{n_k^{\rightarrow}}_k, \mathbf{y}_k^{\mathcal{I}_n^k}\right]$$

$$\widehat{P}^n_{k+1} = \mathbb{E}\left[\left(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}^n_{k+1|k}\right)\left(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}^n_{k+1|k}\right)^T \;\middle|\; \widehat{\mathbf{x}}^{n_k^{\rightarrow}}_{k|k-1}, \widehat{P}^{n_k^{\rightarrow}}_k, \mathbf{y}_k^{\mathcal{I}_n^k}\right]$$

M-GIKF: Estimate Update (*Contd.*)

- The filtering steps can be implemented through time-varying Kalman filter recursions
- The sequence $\{\widehat{P}_k^n\}$ of conditional predicted error covariance matrices at sensor n satisfies the random Riccati recursion:

$$\widehat{P}_{k+1}^n = \mathcal{F}\widehat{P}_k^{n_k^{\rightarrow}}\mathcal{F}^T + \mathcal{Q} - \mathcal{F}\widehat{P}_k^{n_k^{\rightarrow}}\mathcal{C}_n^T\left(\mathcal{C}_n\widehat{P}_k^{n_k^{\rightarrow}}\mathcal{C}_n^T + \mathcal{R}_n\right)^{-1}\mathcal{C}_n\widehat{P}_k^{n_k^{\rightarrow}}\mathcal{F}^T$$

- The sequence $\{\widehat{P}_k^n\}$ is random due to the random neighborhood selection functions $n_k^{\rightarrow}$ and $\mathcal{I}_n^k$
- The goal is to study asymptotic properties of $\{\widehat{P}_k^n\}$ at every sensor n
  - In what sense $\{\widehat{P}_k^n\}$ is stable
  - In what sense they reach *agreement*

Switched Riccati Iterates

Let $\mathfrak{P}$ denote a generic subset of $[1, \cdots, N]$.
Define:

- $\mathcal{C}_{\mathfrak{P}}$: The stack of $\mathcal{C}_j$'s for all $j \in \mathfrak{P}$
- $f_{\mathfrak{P}}(\cdot)$: The Riccati operator given by

$$f_{\mathfrak{P}}(X) = \mathcal{F}X\mathcal{F}^T + \mathcal{Q} - \mathcal{F}X\mathcal{C}_{\mathfrak{P}}^T \left(\mathcal{C}_{\mathfrak{P}}X\mathcal{C}_{\mathfrak{P}}^T + \mathcal{R}_n\right)^{-1} \mathcal{C}_{\mathfrak{P}}X\mathcal{F}^T$$

The conditional prediction error covariances are then updated as:

$$\widehat{P}_n(k+1) = f_{\mathcal{I}_k^n}\left(\widehat{P}_{n_k^{\rightarrow}}(k)\right), \quad \forall n$$

Remark 3. *The sequence* $\{\widehat{P}_n(k)\}$ *evolves as a random Riccati iterate with non-stationary switching.*

M-GIKF: Main Results on Convergence

Theorem 1. *Consider the M-GIKF under assumptions* **(S.1)**, **(D.1)**, **(C.1)**, *and* **(E.1)-(E.3)**. *Then,*

(1) *Let* q *be a uniformly distributed random variable on* $[1, \cdots, N]$ *independent of the sequences* $\{A^e(k)\}$ *and* $\{\mathcal{I}_n^k\}$. *Then, the sequence* $\left\{\widehat{P}_q(k)\right\}$ *converges weakly to an invariant distribution* $\mu^{\overline{\gamma}}$ *on* $\mathbb{S}_+^N$, *i.e.,*

$$\widehat{P}_q(k) \Longrightarrow \mu^{\overline{\gamma}}.$$

(2) *The performance approaches the centralized one exponentially over* $\bar{\gamma}$.

(3) *For each* n, *the sequence of conditional error covariances* $\left\{\widehat{P}_n(k)\right\}$ *is stochastically bounded,*

$$\lim_{J\to\infty} \sup_{k\in\mathbb{T}_+} \mathbb{P}(\|\widehat{P}_n(k)\| \geq J) = 0.$$

Proof Methodology

- An appropriate stationary modification of the switched Riccati iterate governing the covariance evolution leads to a hypothetical Random Dynamical System (RDS) in the sense of Arnold.
- The RDS thus constructed is shown to be order preserving and strongly sublinear.
- Together with the global detectability and network connectivity assumptions, the weak convergence of the stationary hypothetic RDS is established.
- The ergodicity of the actual covariance sequence $\{\widehat{P}_n(k)\}$ is then applied to obtain the same convergence results for the original non-stationary system.

Summary of Estimation Results

- Stability of distributed Kalman filtering errors is established under assumptions of network connectivity and global detectability.
- The analysis requires a new approach to the random Riccati equation (with non-stationary switching).
- Distributional convergence of the random Riccati equation is established;
- Approaching centralized performance exponentially fast over the inter-sensor communication rate.

Distributed Detection

- N sensors decide between two hypotheses: H_0 vs. H_1.

- Sensor n observes $Y_n(k)$ at time k; observation vector: $\mathbf{Y}(k) = [Y_1(k), \ldots, Y_N(k)] \in \mathbb{R}^N$

- Conditional PDFs f_0 and f_1:

$$\text{Under } H_l : \quad f_l(\mathbf{Y}(k))d\nu, \quad l = 0, 1.$$

- Global Kullback-Lieber (KL) divergence $D(P_0||P_1)$ vs. $D(P_1||P_0)$:

$$D(P_0||P_1) = \mathbb{E}_0\left[\ln\left(\frac{f_0(\mathbf{Y}(k))}{f_1(\mathbf{Y}(k))}\right)\right] = \int_{\mathbb{R}^N} \ln\left(\frac{f_0(\mathbf{Y})}{f_1(\mathbf{Y})}\right) f_0(\mathbf{Y})d\nu(\mathbf{Y}).$$

Mixed Time Scale Distributed Detection: Algorithm $\mathcal{MD}$

- Each sensor maintains a local decision variable: $X_n(k)$.
- Connected sensor pair (i, j) communicates as $U_{i,j}(k) = X_i(k) + W_{i,j}(k)$.
- Local decision: H_1 if $X_n(k) > 0$; otherwise H_0.
- Type-I error:

$$\mathbb{P}_e^1(n, k) = \mathbb{P}_0(X_n(k) > 0),$$

and the Type-II error:

$$\mathbb{P}_e^2(n, k) = \mathbb{P}_1(X_n(k) \leq 0).$$

Local Decision Variable $X_n(k)$

- At time k, local observation generates local intelligence (LLR): $S_n(k) = \log\left(\frac{f_1(Y_n(k))}{f_0(Y_n(k))}\right)$.

- At time $k+1$, local decision variable is updated (mixed time scale) as

$$\begin{aligned}
X_n(k+1) &= X_n(k) - \beta(k) \sum_{j\in\Omega_n} \big[X_n(k) - U_{n,j}(k)\big] + \alpha(k)\big[S_n(k) - X_n(k)\big] \\
&= X_n(k) - \beta(k) \sum_{j\in\Omega_n} [X_n(k) - X_j(k) - W_{n,j}(k)] \\
&\quad +\alpha(k)\big[S_n(k) - X_n(k)\big] \\
&= \Big[1 - \beta(k) \sum_{j\in\Omega_n} 1 - \alpha(k)\Big] X_n(k) + \beta(k) \sum_{j\in\Omega_n} X_j(k) \\
&\quad +\beta(k) \sum_{j\in\Omega_n} W_{n,j}(k) + \alpha(k) S_n(k).
\end{aligned}$$

Key Assumptions

- **(E.1)**: The sensor observations are conditionally independent, i.e.,

$$\text{Under } H_l: \quad f_l(\mathbf{Y}(k))d\nu = \prod_{n=1}^{N} f_{n,l}(Y_n(k))d\nu, \quad l = 0, 1,$$

- **(E.2)**: We assume global detectability, i.e., $D(P_0||P_1) > 0$ (and $D(P_1||P_0) > 0$).
- **(E.3)**: The inter-sensor communication network is connected.
- **(E.4)**: The time varying weight sequences $\{\beta(k)\}_{k=0}^{\infty}$ and $\{\alpha(k)\}_{k=0}^{\infty}$ associated with the update process satisfy

$$\beta(k) = \frac{b_0}{(k+1)^\tau}, \quad \frac{1}{2} < \tau < 1, \text{ and } \alpha(k) = \frac{a}{(k+1)}.$$

($\lim_{k\to\infty} \beta(k)/\alpha(k) = \infty$: Consensus potential vs. Innovation potential.)

Main Theorem and Summary

Theorem 2. *Let assumptions* ***(E.1)*****-*****(E.4)*** *hold. Then,*

$$\lim_{k\to\infty} \mathbb{P}_e^1(n,k) = 0 \quad \textit{and} \quad \lim_{k\to\infty} \mathbb{P}_e^2(n,k) = 0, \quad n = 1, \ldots, N.$$

- When local observations are i.i.d., a single time scale distributed algorithm is achievable;
- With Gaussian noise assumptions, error decay exponent is obtainable.

About My School and Research Group

- Texas A&M University: One of the largest public schools in USA;
- Dwight Look Engineering School (USNews Rank 13th)
- Dept. of Electrical and Computer Engineering (USNews Rank 18th)
- My research Group:
  - 2 PostDoc, 9 Ph.D students, and 2 MS students
  - 8 NSF and DoD research projects
  - Main topics: network information theory, cognitive radio networks, and stochastic signal processing for smart grid
  - Openings: Postdocs and students for the coming year