Resilient Monitoring and Control
Techniques, Analysis, Design, and Performance Evaluation

Semyon M. Meerkov

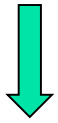
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109
USA

Scientific Council Meeting
Institute of Control Sciences
Moscow, Russia
November 23, 2017

Translations of “Resilient” and “Monitoring”

- Resilient
  - Dictionary translation:
    - Упругий
    - Жизнеспособный
    - Жизнерадостный
    - Неунывающий
  - Our translation:
    - **Выживающий**
- Monitoring
  - Dictionary translation:
    - Надзор
    - Отслеживание
    - Наблюдение
  - Our choice:
    - **Наблюдение**

Resilient Monitoring and Control Systems



Выживающие Системы Наблюдения и Управления

Outline

1. Introduction: Concepts and Results

PART I: RESILIENT MONITORING

2. Scenario: Model and problem
3. Data quality, process variable, and plant condition assessments
4. Sensor network adaptation
5. Decentralization and knowledge fusion
6. Decentralized RMS for a power plant

PART II: RESILIENT CONTROL

7. PMF-based resilient control
8. Synchronous detection-based resilient control

9. CONCLUSIONS AND FUTURE WORK

1 Introduction: Concepts and Results

- Cyber-physical systems considered in this work consist of the following:
  - *Plant*, which is characterized by several process variables and which may operate in various regimes; in each regime the plant is characterized by different dynamics (e.g., transfer functions); the plant may be under a physical attack.
  - *Monitoring system*, which is a sensor network, intended to monitor process variables and inform the plant operator about plant regime – normal of anomalous; the sensors may be under a cyber-attack;
  - *Feedback control system*, which uses the outputs on the monitoring system to maintain the plant in the desired (e.g., Normal) regime.

1 Introduction: Concepts and Results (cont)

- Monitoring system is called *resilient* (RMS) if it provides the least uncertain (in terms of minimum entropy) process variables and plant condition assessment.
- Control system is called *resilient* (RCS) if, given the output of RMS, it maintains the plant in the Normal regime with the largest probability.
- Jointly, RMS and RCS are referred to as a *resilient monitoring and control system* (RM&C).

1 Introduction: Concepts and Results (cont)

- The problem addressed in this work: Develop methods for design and analysis of RM&C systems.
- To attain monitoring resiliency, the methods developed here are:
  - *Sensor data quality assessment*, $DQ \in [0, 1]$ (based on a probing signals approach)
  - *Process variable assessment* (based on sensor measurements, its DQ and Dempster- Shafer rule)
  - *Sensor network adaptation* to the state corresponding to minimum entropy of process variable assessment (based on the rational controllers approach)
  - *Knowledge fusion* (based on inference calculations)
  - *Plant condition assessment* (based on Jeffrey's rule)
  - *RMS efficacy assessment* (based on Kullback-Leibler divergence).
  - This leads to an "*eclectic theory*" of RMS (based on non-classical statistics, signal processing, information theory, and control).
- These methods are outlined in Part I of the talk.

1 Introduction: Concepts and Results (cont)

- To attain control resiliency, the methods developed in are:
  - Calculation of a *safe control signal*, U_{safe}, which maintains the plant is a safe domain, even if the monitoring system output (i.e., plant condition assessment) has the maximum entropy
  - Calculation of a *desired control signal*, U_{des}, which forces the plant operation in a desired regime, if the plant status were known precisely
  - Calculation of a *resilient control signal*, U_{res}, as a linear combination of U_{safe} and U_{des} with optimal weights, which maintains the plant in the desired regime with the largest probability (using MPC approach)
  - This leads to a novel control method, wherein the plant is described by regime-dependent transfer functions, while RMS provides the probability of each regime (pmf-based control)
  - In addition, for "stand-alone" feedback systems, we provide a synchronous detection-based method for identification of attacks on actuators and sensors and, if at all possible, mitigation of their effects.
- These methods are outlined in Part II of the talk.

1 Introduction: Concepts and Results (cont)

- Related literature:

  Foundational issues (M. Amin, C. Rieger)
  Control-theoretic approach (S. Sastry, F. Bullo)
  Fault-tolerant control (M. Balnke, J. Schroder, H. Noura)
  Network anomaly detection (C. Cassandras, I. Paschalidis)
  False data injection (O. Kosut, V. Poor, T. Basar).

- The current work is different in that:

  Sensors may provide misleading information about process variable statuses (due to cyber-attacks)
  Plant may be in different statuses (due to physical attacks or malfunction)

  - Our publications: *IEEE Tr. Cybernetics* (2014), *JPC* (2015), *IEEE Tr. Cybernetics* (2017).

2 SCENARIO: MODEL AND PROBLEM

- Plant, **G**, may operate in several regimes, $G \in \Sigma_G = \{\mathrm{N}_G, \mathrm{A}_{G_1}, \dots, \mathrm{A}_{G_K}\}$; plant's status is defined by a pmf $p[G]$.
- Process variable, $\mathbf{V}_i, i = 1, \dots, M$, status is defined by $p[V_i]$, where

$$V_i \in \Sigma_{V_i} = \{\mathrm{L}_{V_i}, \mathrm{N}_{V_i}, \mathrm{H}_{V_i}\}\ , i = 1, \dots, M.$$

- Plant models: $(A_\sigma, B_\sigma, C_\sigma), \sigma \in \Sigma$

$$P[V_1, \dots, V_M \,|G]$$
$$P[V_i|V_j], i, j = 1, \dots M, i \neq j.$$

- Process variable model:

$$\tilde{V}_i \in [\,\tilde{V}_{i,min}, \tilde{V}_{i,max}\,]$$
$$\tilde{V}_i \in [\,\tilde{V}_{i,min}, R_1] \Longrightarrow V_i = \mathrm{L}_{V_i};\ \tilde{V}_i \in [\,R_1, R_2] \Longrightarrow V_i = \mathrm{N}_{V_i};$$
$$\tilde{V}_i \in [R_2, \tilde{V}_{i,max}] \Longrightarrow V_i = \mathrm{H}_{V_i}$$
$$T_{V_i,\sigma}(s), i = 1, \dots, M,\ \sigma \in \{\mathrm{L}_{V_i}, \mathrm{N}_{V_i}, \mathrm{H}_{V_i}\}.$$

2 Scenario: Model and Problem (cont)

- Sensor, $\mathbf{S}_i$, reports data, $\tilde{S}_i \in [\,\tilde{V}_{i,min}, \tilde{V}_{i,max}\,]$, which induces S_i with $p[S_i]$, where

$$S_i \in \Sigma_{S_i} = \{\mathrm{L}_{S_i}, \mathrm{N}_{S_i}, \mathrm{H}_{S_i}\}\ , i = 1, \ldots, N_S.$$

$$\tilde{S}_i \in [\,\tilde{V}_{i,min}, R_1] \Longrightarrow S_i = \mathrm{L}_{S_i};\ \tilde{S}_i \in [\,R_1, R_2] \Longrightarrow S_i = N_{S_i};\ \tilde{S}_i \in [R_2, \tilde{V}_{i,max}] \Longrightarrow S_i = \mathrm{H}_{S_i}.$$

- Attacker model:
  - If the sensor is not attacked, $E[\,\tilde{S}_i\,] = E[\,\tilde{V}_i\,]$.
  - If the sensor is attacked, $E[\,\tilde{S}_i\,] \neq E[\,\tilde{V}_i\,]$.
- Sensor network, **SN**, consists of N_S sensors. The state, X, of SN is N_S-tuple of 1's and 0's, so $|X| = 2^{N_S}$.
- *Problem*: Design RMS, which (a) evaluates the data quality (*DQ*) projected by each sensor; (b) calculates $\hat{p}[G]$ as a functional of the sensor data and function of *DQ*; and (c) adapts to a state of **SN**, resulting in $\hat{p}[G]$ with smallest entropy,

$$I\{\hat{p}[G]\} = -\sum_{\sigma \in \Sigma_G} \hat{p}[G = \sigma] \log_{|\Sigma_G|} \hat{p}[G = \sigma]$$

- This problem is solved based on series of sub-problems described next.

3a Data Quality Assessment

- *Problem*: Quantify sensor's "trustworthiness".
- *Approach*: Probing signals.
- *Solution*:
  - Consider **V** monitored by **S**. Introduced the probing signal:

$$u_{\mathbf{V}}(t) = A_{\mathbf{V}}\text{rect}_T(t - t_0)$$

  - If the sensor is not attacked,

$$\mu'_{\tilde{S}} - \mu_{\tilde{S}} = A_{\mathbf{V}}\alpha_{\mathbf{V}}(\mu_{\tilde{S}}),$$

where $\mu_{\tilde{S}}$ and $\mu'_{\tilde{S}}$ are $E\left[\tilde{S}\right]$ before and after the test, respectively

  - If the sensor is attacked, the difference between both sides is the *probing inconsistency*:

$$PIC_{\mathbf{S}} := \left|(\mu'_{\tilde{S}} - \mu_{\tilde{S}}) - A_{\mathbf{V}}\alpha_{\mathbf{V}}\left(\mu_{\tilde{S}}\right)\right|.$$

3a Data Quality Assessment (cont)

- Introduce the measure of *data quality*:

$$DQ_{\mathbf{S}} = e^{-F(PIC_{\mathbf{S}})},$$

where

$$F(PIC_{\mathbf{S}}) := -\frac{\ln \epsilon}{PIC^2_{\max,\mathbf{S}}} PIC^2_{\mathbf{S}},$$

$$\epsilon \ll 1,$$

$PIC_{\max,\mathbf{S}}$ is the largest value attainable by $PIC_{\mathbf{S}}$.

- As it follows from the above,

$$\min DQ_{\mathbf{S}} = \epsilon$$

and ϵ is a design parameter.

3b Process Variable Assessment

- *Problem*: Evaluate $\hat{p}[V]$ as a functional of sensor data and function of DQ.
- *Approach*: Modeling of S and V coupling and a recursive statistical procedure based this coupling, sensor data, and DQ.
- *Solution*: Single sensor case:
  - Introduce the sensor *believability*:

$$\beta_{\mathbf{S}} = \frac{|\Sigma_V| - 1}{|\Sigma_V|} DQ_{\mathbf{S}} + \frac{1}{|\Sigma_V|}$$

  - Postulate V and S *coupling* :

$$P[V = \sigma | S = \sigma] = \beta_{\mathbf{S}},$$

$$P[V = \bar{\sigma} | S = \sigma] = \frac{1 - \beta_{\mathbf{S}}}{|\Sigma_V| - 1}$$

  - Let $\hat{p}_n[V = \sigma] = P[V = \sigma | s_1, s_2, ..., s_n; DQ_{\mathbf{S}}], \ \forall \sigma \in \Sigma_V$ and introduce the notation:

$$h_\sigma(n) := \hat{p}_n[V = \sigma], \ \forall \sigma \in \Sigma_V.$$

3b Process Variable Assessment (cont)

- Introduce a procedure (*h-procedure*) for $\hat{p}_n[V]$ evaluation:

$$h_\sigma(n+1) = h_\sigma(n) + \epsilon_h \left[h^*_\sigma(s_{n+1}) - h_\sigma(n)\right], \ \forall \sigma \in \Sigma_V,$$

$$h_\sigma(0) = \frac{1}{|\Sigma_V|}, \ \forall \sigma \in \Sigma_V,$$

$$h^*_\sigma(s_{n+1}) = \begin{cases} \beta_{\mathbf{S}}, & \text{if } s_{n+1} = \sigma \\ \frac{1-\beta_{\mathbf{S}}}{|\Sigma_V|-1}, & \text{if } s_{n+1} \neq \sigma, \end{cases}$$

$$0 < \epsilon_h(n) \leq 1, \ \sum_{n=0}^{\infty} \epsilon_h(n) = \infty, \ \sum_{n=0}^{\infty} \epsilon_h^2(n) < \infty,$$

or

$$0 < \epsilon_h << 1.$$

3b Process Variable Assessment (cont)

- *Theorem*:

$$\lim_{n\to\infty} h_\sigma(n) = p[S=\sigma]DQ_{\mathbf{S}} + \frac{1-DQ_{\mathbf{S}}}{|\Sigma_V|},\ \forall\sigma\in\Sigma_V$$

  - The convergence is in probability if $\epsilon_h(n) = const$ and almost sure if $\epsilon_h(n)$ is a decreasing function defined above.

- Thus, h-procedure results in:

$$\hat{p}[V=\sigma] = p[S=\sigma]DQ_{\mathbf{S}} + \frac{1-DQ_{\mathbf{S}}}{|\Sigma_V|},\ \forall\sigma\in\Sigma_V.$$

3b Process Variable Assessment (cont)

- *Solution*: Multiple sensors case
  - Need to calculate:

$$\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V=\sigma]=\lim_{n\to\infty}P[V=\sigma|s_1^1,...,s_n^1;DQ_{\mathbf{S}_1};s_1^2,...,s_n^2;DQ_{\mathbf{S}_2}],\ \forall\sigma\in\Sigma_V.$$

  - This is accomplished by calculating $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$ and then using the Dempster-Shafer rule:

$$\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V=\sigma]=\frac{\hat{p}^{\mathbf{S}_1}[V=\sigma]\hat{p}^{\mathbf{S}_2}[V=\sigma]}{\sum_{\sigma\in\Sigma_V}\hat{p}^{\mathbf{S}_1}[V=\sigma]\hat{p}^{\mathbf{S}_2}[V=\sigma]},\ \forall\sigma\in\Sigma_V.$$

  - The final pmf is selected as

$$\hat{p}^*[V]=\arg\min\left[I\{\hat{p}^{\mathbf{S}_1}[V]\},I\{\hat{p}^{\mathbf{S}_2}[V]\},I\{\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]\}\right],$$

    where

$$I\{p[V]\}=-\sum_{\sigma\in\Sigma_V}p[V=\sigma]\log_{|\Sigma_V|}p[V=\sigma].$$

3c Plant Condition Assessment

- *Problem*: Evaluate $\hat{p}[G]$
- *Approach*: Jeffrey's rule
- *Solution*:
  - Based on a single process variable,

$$p_0[G] = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right], \quad p_0[V_i, G] = P[V_i|G]p_0[G], \quad p_0[V_i] = \sum_{G \in \Sigma_G} p_0[V_i, G],$$

$$\hat{p}[V_i, G] = p_0[V_i, G]\frac{\hat{p}[V_i]}{p_0[V_i]}, \quad \hat{p}^{V_i}[G] = \sum_{V_i \in \Sigma_{V_i}} \hat{p}[V_i, G].$$

  - For multiple process variables, combine the pmf's obtained using Dempster-Shafer rule:

$$\hat{p}[G = \sigma_G] = \frac{\prod_{i=1}^{M} \hat{p}^{V_i}[G = \sigma_G]}{\sum_{\sigma_G \in \Sigma_G} \prod_{i=1}^{M} \hat{p}^{V_i}[G = \sigma_G]}, \ \sigma_G \in \Sigma_G$$

4 Sensor Network Adaptation

- *Problem*: Adapt to the network state resulting in the smallest entropy of $\hat{p}[G]$.
- *Approach*: Based on rational controllers, i.e., dynamical systems operating in the decision space and having two properties: ergodicity and rationality.
  - *Ergodicity* implies that that all decisions are visited with non-zero probability
  - *Rationality* implies that residence time in decisions with a smaller penalty function are larger than in those with a larger one
  - The degree to which this variance takes place is called the *level of rationality*, $N \geq 1$.
  - Example: $\dot{x} = f^N(\{x\})$, where $f(\cdot) > 0$. Then

$$\lim_{N\to\infty} \frac{1}{T_N} \int_0^{T_N} x(t)dt = \min_{x\in[0,1)} f(x).$$

4 Sensor Network Adaptation (cont)

- *Solution:*
  - Decision space – state space of the network
  - Penalty function – entropy of $\hat{p}[G]$ in network state x, i.e., $\hat{p}_x[G]$.
  - Ergodicity is ensured by visiting all states in round-robin manner
  - Rationality is ensured by selecting residence time in $x \in X$ as

$$T_x = \begin{cases} T_{\max}, & \text{if } \hat{I}_x(G) \le \beta \\ \left(\frac{\beta}{\hat{I}_x(G)}\right)^N T_{\max}, & \text{if } \hat{I}_x(G) > \beta \end{cases}$$

  - Let

$$\tau_x = \frac{T_x}{\sum_{x \in X} T_x}$$

  - Then the plant assessment reported to the plant operator is

$$\bar{p}[G] = \sum_{x \in X} \tau_x \hat{p}_x[G].$$

4 Sensor Network Adaptation (cont)

- *Measure of resiliency:*
  - Based on Kullback-Leibler divergence:

$$D\left(p_1[G]||p_2[G]\right) = \sum_{\sigma_G \in \Sigma_G} p_1\left[G = \sigma_G\right] \log_{|\Sigma_G|} \frac{p_1[G = \sigma_G]}{p_2[G = \sigma_G]}.$$

  - Select $p_1[G]$ as $p[G]$ and $p_2[G]$ as either $\bar{p}[G]$ or $p_{\rm nr}[G]$. Then *MR* is defined as

$$MR = \frac{D\left(p[G]||p_{\rm nr}[G]\right) - D\left(p[G]||\bar{p}[G]\right)}{D\left(p[G]||p_{\rm nr}[G]\right)}.$$

  - Clearly, $MR \leq 1$, and the equality is reached when

$$\bar{p}[G] = p[G]$$

4 Sensor Network Adaptation (cont)

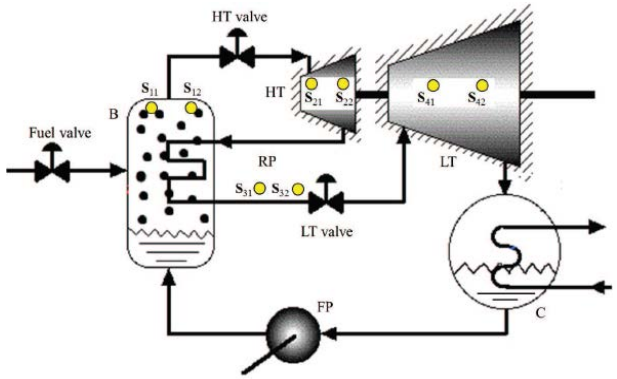
- *Temporal properties*
  - The adaptation process consists of epochs; $|X|$ epochs comprise a cycle; at the end of each cycle $\bar{p}[G]$ is reported to plant operator.
  - For each $x \in X$, the epoch consists of the three periods:
    - DQ acquisition (T_{DQ})
    - process variable and plant pmf acquisition (T_{eval})
    - residence time in state x (T_x).
  - Assuming sensor data is provided every 0.01sec, $T_{DQ} \cong$ 5sec, $T_{eval} \cong$ 6sec, and T_x can be selected as 1sec. Thus, $T_{epoch} \leq$ 12sec.
  - Based on the above, $T_{cycle} = 12|X|$, i.e., in general,

$$T_{cycle} = \gamma 2^{Ns}$$

    resulting in the *curse of dimensionality*.
  - A method for combatting this effect is described next.

5 DECENTRALIZATION WITH KNOWLEDGE FUSION

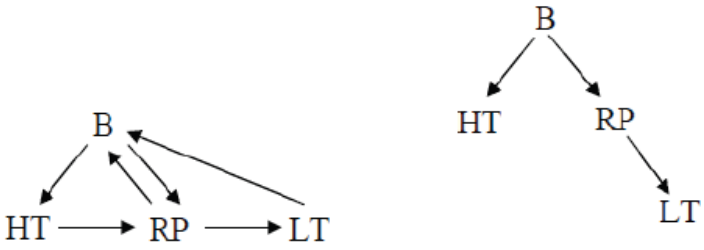
- *Problem*: Reduce the effect of exponential dependence of T_{cycle} on N_S.
- *Approach*: Plant decomposition with subsequent knowledge fusion.
  - To illustrate this approach, consider a simplified model of power plant:



  - Having 8 sensors, $T_{cycle} \approx 51$min, must be reduced dramatically
  - Decompose the plant into sub-plants B, HT, RP, LT
  - This induces sensor network decomposition into subnetworks, each monitoring one process variable.

5 Decentralization and Knowledge Fusion (cont)

- Influence diagram of process variables:



(a) Cyclic graph (b) Tree graph

- Decentralized representation:
  - Plant: $\mathbf{G}_{\mathrm{B}}$, $\mathbf{G}_{\mathrm{HT}}$, $\mathbf{G}_{\mathrm{RP}}$, and $\mathbf{G}_{\mathrm{LT}}$ with $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, and $\mathbf{V}_4$
  - Sensor network: $\mathbf{SN}_{\mathrm{B}}$, $\mathbf{SN}_{\mathrm{HT}}$, $\mathbf{SN}_{\mathrm{RP}}$, and $\mathbf{SN}_{\mathrm{LT}}$.
  - Need to assess: $\bar{p}_{\mathbf{G}_{\mathrm{B}}}[V_1]$, $\bar{p}_{\mathbf{G}_{\mathrm{HT}}}[V_2]$, $\bar{p}_{\mathbf{G}_{\mathrm{RP}}}[V_3]$, and $\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_4]$
  - Cardinality of each subnetwork state space:

$$|X_k| = 2^{N_{S_k}} \text{ and } T_{cycle} \approx 48\text{sec}$$

- To account for loss of information, use knowledge fusion (inference calculations).

5 Decentralization and Knowledge Fusion (cont)

- Inference calculations:

$$\bar{p}_{\mathbf{G}_{\mathrm{HT}}}[V_1] = \sum_{\sigma_2 \in \Sigma_{V_2}} P[V_1|V_2 = \sigma_2]\bar{p}_{\mathbf{G}_{\mathrm{HT}}}[V_2 = \sigma_2]$$

$$\bar{p}_{\mathbf{G}_{\mathrm{RP}}}[V_1] = \sum_{\sigma_3 \in \Sigma_{V_3}} P[V_1|V_3 = \sigma_3]\bar{p}_{\mathbf{G}_{\mathrm{RP}}}[V_3 = \sigma_3]$$

$$\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_1] = \sum_{\sigma_3 \in \Sigma_{V_3}} P[V_1|V_3 = \sigma_3]\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_3 = \sigma_3],$$

where

$$\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_3] = \sum_{\sigma_4 \in \Sigma_{V_4}} P[V_3|V_4 = \sigma_4]\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_4 = \sigma_4].$$

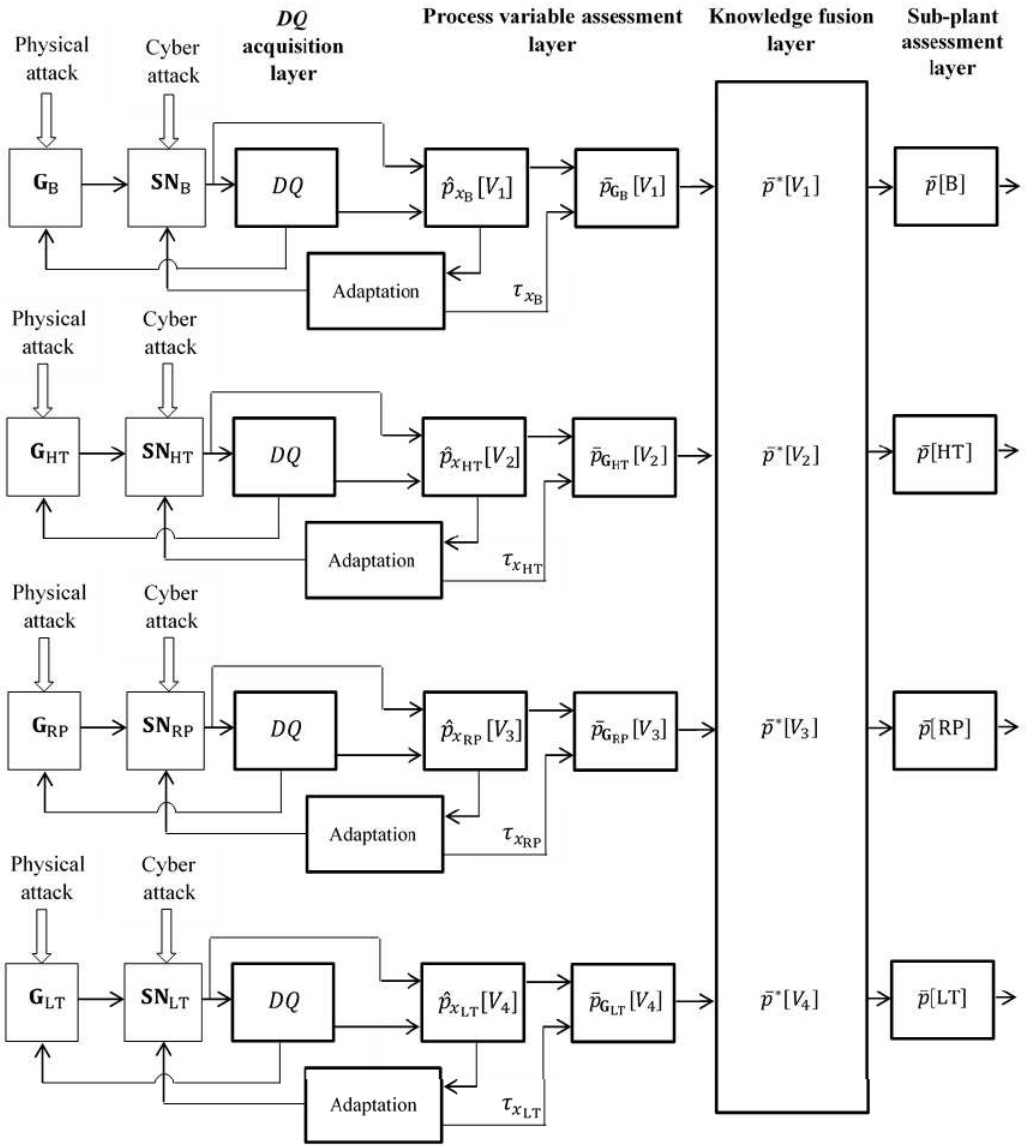
- Then, pmf of V_1 based on all subplants is:

$$\bar{p}_{\mathbf{G}_{\mathrm{B,HT,RP,LT}}}[V_1 = \sigma_1] = \frac{\prod_{k=\mathrm{B,HT,RP,LT}} \bar{p}_{\mathbf{G}_k}[V_1 = \sigma_1]}{\sum_{\sigma_1 \in \Sigma_{V_1}} \prod_{k=\mathrm{B,HT,RP,LT}} \bar{p}_{\mathbf{G}_k}[V_1 = \sigma_1]}$$

- Finally, the pmf of V_1 is selected as:

$$\bar{p}^*[V_1] = \arg\min \left\{ I\left\{\bar{p}_{\mathbf{G}_{\mathrm{B}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{HT}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{RP}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{B,HT,RP,LT}}}[V_1]\right\}\right\}.$$

6 Decentralized RMS for a Power Plant



6 Decentralized RMS for a Power Plant (cont)

- Attack scenarios and resulting performance:
  - Scenario 1: All sub-plants in N; Cyber-attack on B projecting A:
    - RMS reports:
$$\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.9, 0.1],$$
$$\bar{p}[G_{\mathrm{RP}}] = [0.91, 0.09],\ \bar{p}[G_{\mathrm{LT}}] = [0.92, 0.08]$$
    - Non-resilient system reports:
$$p_{\mathrm{nr}}[G_{\mathrm{B}}] = [0.05, 0.95]$$
    - Measure of resiliency:
$$\overrightarrow{MR} = [0.98, -, -, -]$$
  - Scenario 2: All sub-plants in N; Cyber-attack on LT projecting A:
    - RMS reports:
$$\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.9, 0.1],$$
$$\bar{p}[G_{\mathrm{RP}}] = [0.91, 0.09],\ \bar{p}[G_{\mathrm{LT}}] = [0.49, 0.51]$$
    - Non-resilient system reports:
$$p_{\mathrm{nr}}[G_{\mathrm{LT}}] = [0.09, 0.91]$$
    - Measure of resiliency :
$$\overrightarrow{MR} = [-, -, -, 0.7].$$

6 Decentralized RMS for a Power Plant (cont)

- Attack scenarios and resulting performance (cont):
  - Scenario 3: Coordinated cyber-physical attack on RP:
    - RMS reports:

$$\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.9, 0.1],$$
$$\bar{p}[G_{\mathrm{RP}}] = [0.12, 0.88],\ \bar{p}[G_{\mathrm{LT}}] = [0.92, 0.08]$$

    - Non-resilient system reports:

$$p_{\mathrm{nr}}[G_{\mathrm{RP}}] = [0.91, 0.09]$$

    - Measure of resiliency:

$$\overrightarrow{MR} = [-, -, 0.95, -].$$

    - If the attack was not coordinated, e.g., physical attack on RP and cyber attack on LT, the status of LT would be undetermined: ,

$$\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.9, 0.1],$$
$$\bar{p}[G_{\mathrm{RP}}] = [0.12, 0.88],\ \bar{p}[G_{\mathrm{LT}}] = [0.49, 0.51].$$

6 Decentralized RMS for a Power Plant (cont)

- Attack scenarios and resulting performance:
  - Scenario 4: Coordinated cyber-physical attack on HT:
    - RMS reports:

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.51, 0.49],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],$$

$$\bar{p}[G_{\text{LT}}] = [0.92, 0.08],$$

    - Non-resilient system reports:

$$p_{\text{nr}}[G_{\text{HT}}] = [0.9, 0.1]$$

    - Measure of resiliency:

$$\overrightarrow{MR} = [-, 0.69, -, -]$$

    - If only one sensor of HT was attacked, the status of all sub-plant would be ascertained correctly:

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.11, 0.89],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],$$

$$\bar{p}[G_{\text{LT}}] = [0.92, 0.08].$$

    - If the attack was non-coordinated, e.g., physical attack on HT and cyber attack on B:

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.1, 0.9],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],$$

$$\bar{p}[G_{\text{LT}}] = [0.92, 0.08],$$

6 Decentralized RMS for a Power Plant (cont)

- **Take-away points:**
  - RMS provides no erroneous information.
  - Attacks on HT and LT are more dangerous than on B and RP.

    Reason: structure of $P[V_i|V_j]$, which permit inferences from HT and LT to B and RP but not vice-versa (terminal points of the influence graph).
  - Coordinated cyber-physical attacks are not more dangerous than non-coordinated ones. Important is if the terminal point are involved or not.
  - The minimum number of non-attacked sensors, which is necessary and sufficient for correct assessment, is two – one for HT and one for LT. If possible, they should be made "known secure".

PART II: RESILIENT CONTROL

7 PMF-BASED RESILIENT CONTROL

- Resilient control problems addressed:
  - Problem 1: Feedback control based on the output of the Resilient Monitoring System
  - Problem 2: Feedback control of a “stand-alone” closed-loop system with its actuator and /or sensor under a malicious attack.
- Approaches:
  - Problem 1 is addressed based on developing a theory for pmf-based feedback control.
  - Problem 2 is addressed based on the method for synchronous detection as a tool for identifying actuators and sensors “health” (with subsequent mitigation of the attack effect, if at all possible).
- Initial results on Problem 1 are presented in this section. Section 8 presents the results on Problem 2.

7 PMF-BASED RESILIENT CONTROL (CONT)

- Classical control configurations:
  - Output-based feedback control (needs output measurements)
  - State space-based feedback control (needs states measurements)
  - Observer-based feedback control (needs inputs and outputs measurements).
- None is applicable if just the pmf of process variable is available.
- This leads to a new control configuration: *pmf-based control*.
- Initial results in this direction obtained to-date are described next.

7 PMF-BASED RESILIENT CONTROL (CONT)

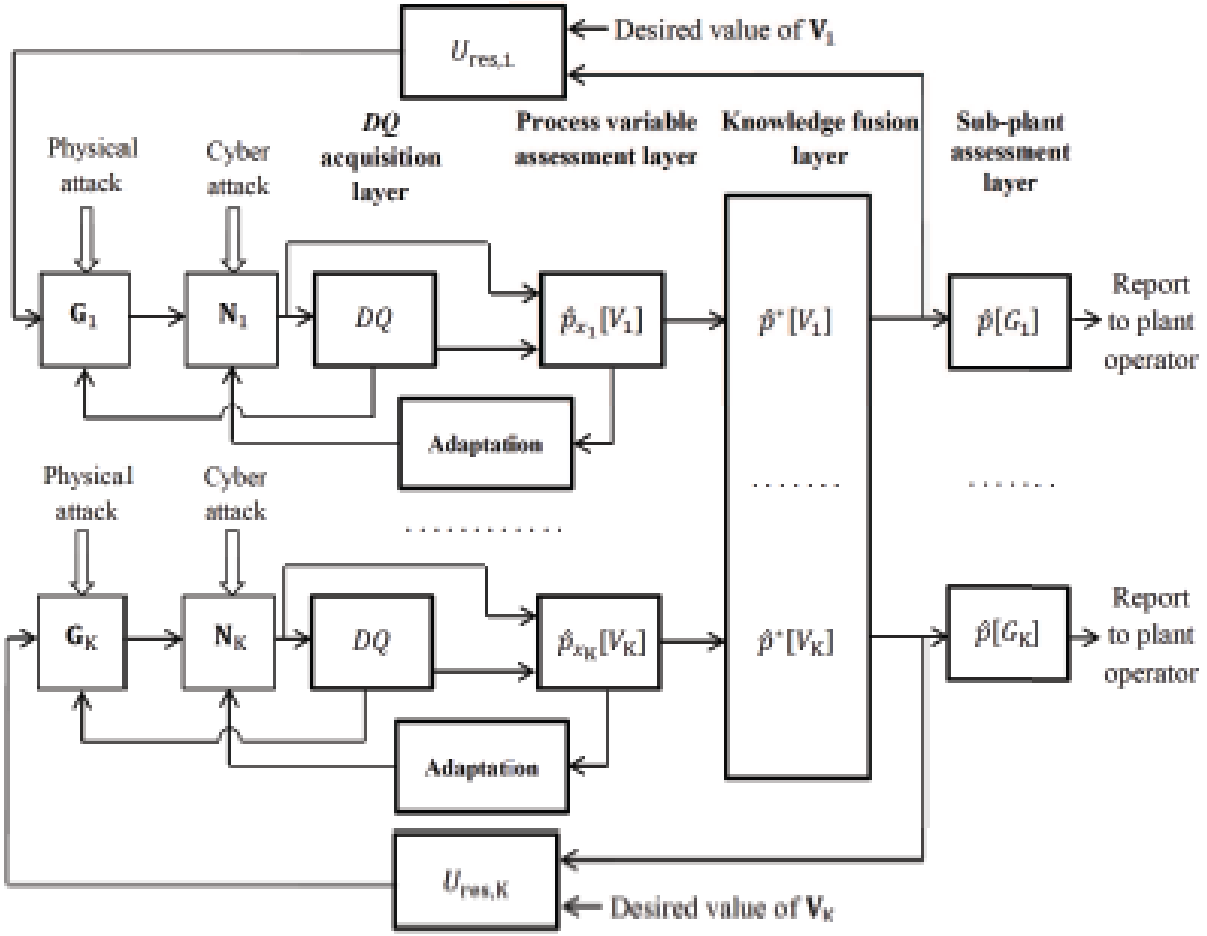
- Approach:
  - Denote:
    - U_{safe} – the control input, which maintains process variable in a safe domain, irrespective on plant's status
    - U_{des} – the control input, which ensures desired value of the output, if the process variable pmf had a zero entropy.
  - Introduce the resilient control input as follows:

$$U_{res} = \Delta U_{des} + (1 - \Delta) U_{safe},$$

where $0 \leq \Delta \leq 1$ is a weighting factor selected based on an optimization procedure (model-predictive control-like), so that when $I\{\hat{p}[V_i]\}$ is close to zero, Δ is close to 1; if $I\{\hat{p}[V_i]\}$ is large, Δ is close to 0.

7 PMF-BASED RESILIENT CONTROL (CONT)

- The architecture of pmf-based control:



- A few details on its development are given next.

7 PMF-BASED RESILIENT CONTROL (CONT)

- Sub-plant equations:

$$\begin{aligned} \mathbf{x}_\sigma(n+1) &= A_\sigma \mathbf{x}_\sigma(n) + B_\sigma U_{\text{res}}(n),\ \mathbf{x}_\sigma \in \mathbb{R}^q,\ U_{\text{res}} \in \mathbb{R},\ n = 0, 1, ..., \\ \tilde{V}(n) &= C_\sigma \mathbf{x}_\sigma(n),\ \tilde{V} \in \mathbb{R},\ n = 0, 1, ...,\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \end{aligned}$$

- Calculation of U_{safe}:
  - d.c. gain:

$$\alpha_\sigma = C_\sigma [I - A_\sigma]^{-1} B_\sigma,\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$$

$$\alpha_{\min} = \min\{\alpha_{\mathrm{L}_V}, \alpha_{\mathrm{N}_V}, \alpha_{\mathrm{H}_V}\},\ \alpha_{\max} = \max\{\alpha_{\mathrm{L}_V}, \alpha_{\mathrm{N}_V}, \alpha_{\mathrm{H}_V}\}.$$

  - Assumption:

$$\frac{\alpha_{\max}}{\alpha_{\min}} < \frac{V_{\max}}{V_{\min}}$$

  - Proposition: U_{safe} can be selected as any constant from the interval

$$U_{\text{safe}} \in \left[\frac{V_{\min}}{\alpha_{\min}}, \frac{V_{\max}}{\alpha_{\max}}\right].$$

7 PMF-BASED RESILIENT CONTROL (CONT)

- Calculation of U_{des}:
  - Introduce a two-degree of freedom feedback law, which would steer the process variable to V_{des} if the plant status, σ, were known:

$$U_\sigma(n) = -K_{1,\sigma}\hat{\mathbf{x}}_{\sigma,U_\sigma}(n) + K_{2,\sigma}V_{\text{des}},\ n = 0, 1, 2, ...,\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}$$

$$\hat{\mathbf{x}}_{\sigma,U_\sigma}(n+1) = A_\sigma\hat{\mathbf{x}}_{\sigma,U_\sigma}(n) + B_\sigma U_\sigma(n),\ n = 0, 1, ...,\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\},$$

$$\hat{\mathbf{x}}_{\sigma,U_\sigma}(0) = \mathbf{x}_\sigma(0),\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}.$$

  - When σ is not known, this control may lead to a disaster. To alleviate this problem, synthesize U_{des} as follows (under the assumption that attacker dynamics are much slower that those of the plant:

$$\begin{aligned} U_{\text{des}}(n) \quad = \quad & \hat{p}[V(0) = \text{L}_V]U_{\text{L}_V}(n) + \hat{p}[V(0) = \text{N}_V]U_{\text{N}_V}(n) \\ & + \hat{p}[V(0) = \text{H}_V]U_{\text{H}_V}(n),\ n = 0, 1, ... \end{aligned}$$

7 PMF-BASED RESILIENT CONTROL (CONT)

- Calculation of Δ:
  - Optimization problem:

$$\underset{\Delta(n),\Delta(n+1),\ldots,\Delta(n+N_{\mathrm{p}}-1)}{\text{minimize}} \quad \sum_{i=1}^{N_{\mathrm{p}}} \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} \frac{1}{2} W_\sigma \left[\hat{V}_{\sigma,U_{\mathrm{res}}}(n+i;\Delta(n+i-1)) - V_{\mathrm{des}}\right]^2$$

$$\text{subject to} \quad V_{\min} \le \hat{V}_{\sigma,U_{\mathrm{res}}}(n+i;\Delta(n+i-1)) \le V_{\max},$$

$$n = 0, 1, \ldots, \; i = 1, \ldots, N_{\mathrm{p}}, \; \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\},$$

(CONT)

  - Given the solution $\Delta^*(n), \ldots, \Delta^*(n+N_p-1)$, only $\Delta^*(n)$ is utilized for $U_{des}(n)$ calculation, while all other are discarded.
- Resilient control signal

$$U_{res}(n) = \Delta^*(n) U_{des}(n, \Delta^*(n)) + (1 - \Delta^*(n)) U_{safe}.$$

7 PMF-BASED RESILIENT CONTROL (CONT)

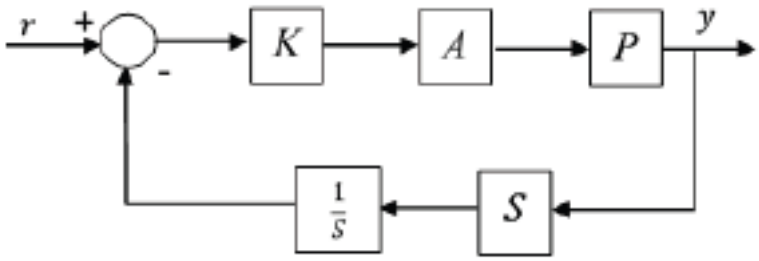
- Definition: The process variable pmf $\hat{p}[V_{ss}]$ is said to be the *proper permutation* of $\hat{p}[V_0]$ if:
  - both pmf's are comprised of the same probabilities;
  - the pmf $\hat{p}[V_{ss}]$ takes the largest probability for $\sigma = \mathrm{N}$, irrespective of $\hat{p}[V_0]$.
- Proposition: Under some technical conditions (related to the steady state gains of the sub-plant transfer functions for $\sigma \in [\mathrm{L}, \mathrm{N}, \mathrm{H}]$ and process variable domains L, N, and H), the pmf of the steady-state process variable under the resilient control signal is the proper permutation of the initial process variable.
- In other words, irrespective of $\hat{p}[V_0]$, the pmf $\hat{p}[V_{ss}]$ has the $\sigma =$ Normal status with the largest probability.

7 PMF-BASED RESILIENT CONTROL (CONT)

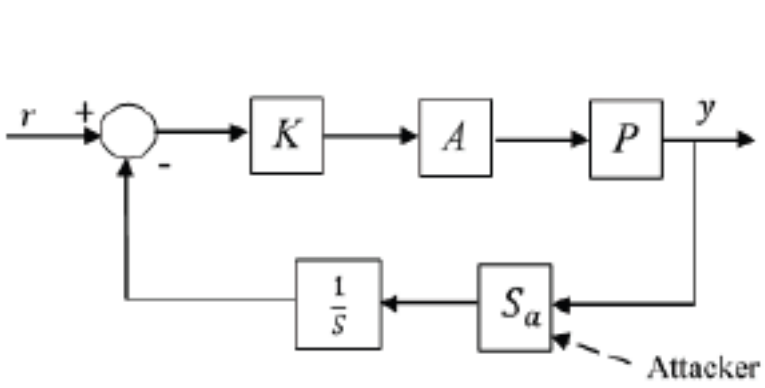
- **Take-away point**: Resilient (pmf-based) control does not “create” information. It just “reshuffles” the process variable’s pmf (provided by RMS) so that in the steady state this process variable is in the Normal status with the largest probability

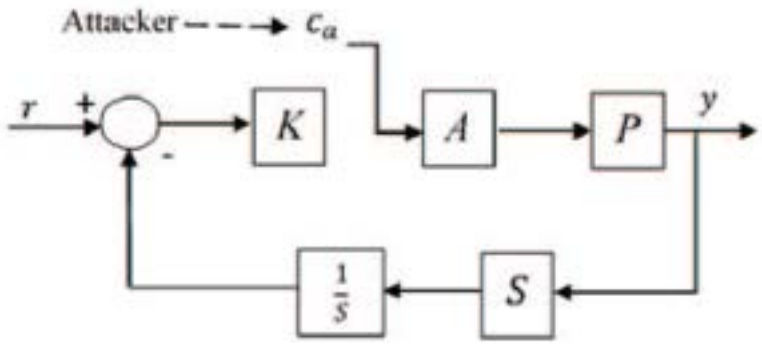
8 Synchronous Detection-based Resilient Control

- Systems considered:
  - Original:



  - Under Type 1 and Type 2 deception attacks:

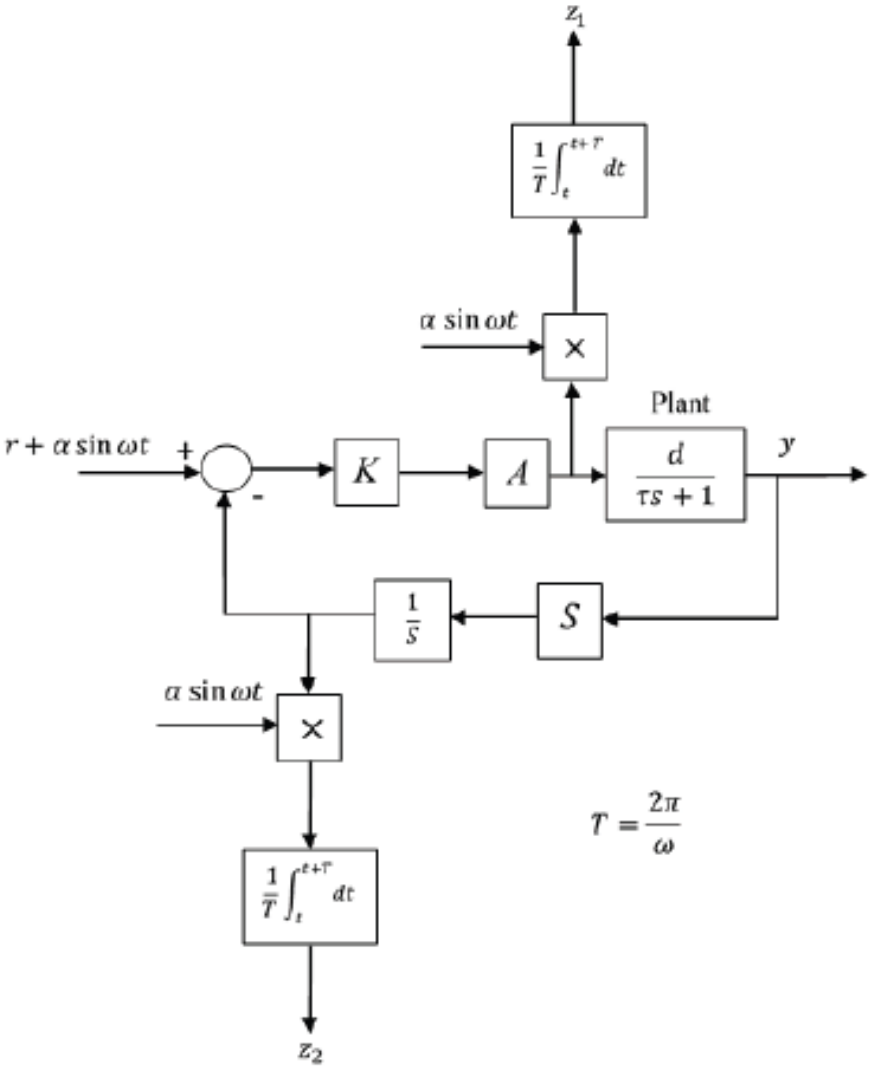




  - Type 3 attack – combination of Type 1 and Type 2.

8 Synchronous Detection-based Resilient Control (cont)

- Approach: Singular detection technique



8 Synchronous Detection-based Resilient Control (cont)

- Analysis:
  - Assumption: Closed-loop system is asymptotically stable
  - Results:

Scenario\Signal	$z_{1,ss}$	$z_{2,ss}$
Nominal system	$\left[\frac{KA(1+KAd+\omega^2\tau^2)}{\omega^2\tau^2+(1+KAd)^2}\right]\frac{\alpha^2}{2}$	$\left[\frac{KAd(1+KAd)}{\omega^2\tau^2+(1+KAd)^2}\right]\frac{\alpha^2}{2}$
Type 1 attack on S and A	$\left[\frac{KA_a\left(1+KA_ad\frac{S_a}{S}+\omega^2\tau^2\right)}{\omega^2\tau^2+\left(1+KA_ad\frac{S_a}{S}\right)^2}\right]\frac{\alpha^2}{2}$	$\left[\frac{KA_ad\frac{S_a}{S}\left(1+KA_ad\frac{S_a}{S}\right)}{\omega^2\tau^2+\left(1+KA_ad\frac{S_a}{S}\right)^2}\right]\frac{\alpha^2}{2}$
Type 2 attack on S	$KA\frac{\alpha^2}{2}$	0
Type 2 attack on A	0	0
Type 2 attack on S and A	0	0
Type 3 attack: Type 2 attack on S and Type 1 attack on A	$KA_a\frac{\alpha^2}{2}$	0
Type 3 attack: Type 1 attack on S and Type 2 attack on A	0	0

8 Synchronous Detection-based Resilient Control (cont)

- Attack mitigation:
  - Type 1 attack:
    - Calculate:

$$S_a = S\left[\frac{4z_{2,ss,a}-\alpha^2+\sqrt{(4z_{2,ss,a}-\alpha^2)^2-8z_{2,ss,a}(1+\omega^2\tau^2)(2z_{2,ss,a}-\alpha^2)}}{4z_{1,ss,a}d}\right]$$

$$A_a = \frac{2z_{1,ss,a}}{K[\alpha^2-2z_{2,ss,a}]}.$$

    - Use $K\frac{A}{A_a}$ instead of K and $\frac{1}{S_a}$ instead of $\frac{1}{S}$.
  - Type 2 and 3 attacks:
    - Operation of control system must be stopped.

8 Synchronous Detection-based Resilient Control (cont)

- Example: Uranium enrichment centrifuge control:
  - Plant:

$$P(s) = \frac{157}{4s + 1}$$

  - Control systems parameters:

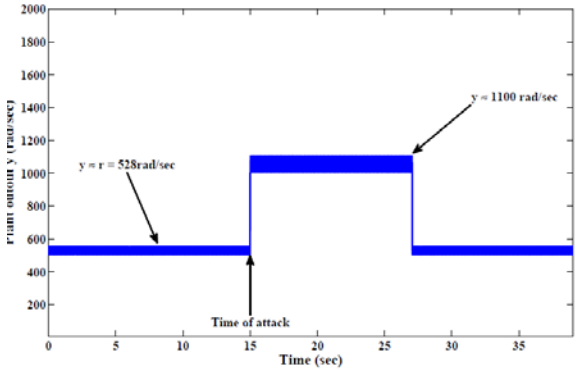
Gains of controller, actuator, and sensor	$K = 20, A = 2, S = 1$
Value of reference signal	$r = 528$
Amplitude and frequency of sinusoidal signal	$\alpha = 25, \omega = 100$

  - Type 1 attack parameters:

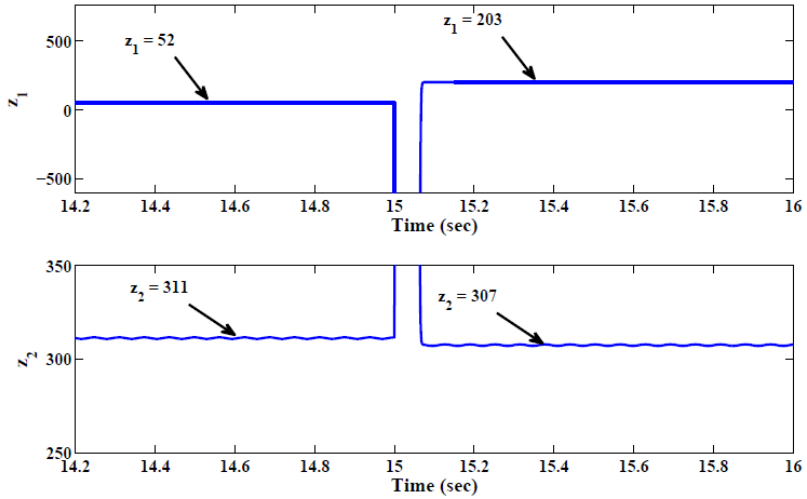
Attacked DC gain of sensor	$S_a = 0.5$
Time of attack	15sec

8 Synchronous Detection-based Resilient Control (cont)

- Example: Uranium enrichment centrifuge control (cont):
  - Results:
    - Trajectory of the output ($T_{id+mitig} < 15$ sec) :



    - Zoomed trajectories of z_1 and z_2 around $t = 15\ sec$:



9. Conclusion and Future Work

- This work provides techniques for resilient monitoring and control of cyber-physical systems and demonstrates their efficacy.
- Numerous research problems remain opened:
  - Resilient monitoring:
    - *DQ* acquisition: efficacy of probing signals for different attackers; improvement of temporal properties; *DQ* acquisition by inferences.
    - $\hat{p}[V]$ assessment: improving temporal properties; believability robustness; other models of coupling between $p[V]$ and $p[S]$.
    - Network adaptation: faster rational controllers; novel penalty functions.
    - Efficacy of knowledge fusion: conditions for lossless decomposition.
  - Resilient control:
    - Further development of the synchronous detection approach.
    - Further development of pmf-based control theory.
    - Possibly, other novel resilient control techniques.
  - Most importantly – APPLICATIONS!

ACKNOWLEDGEMENT

- Results reported have been developed in collaboration with H. Garcia (INL), W.-C. Lin (INL/GM), and M. Ravichandran (UM/Ford).

- This research was supported in part by the US Department of Energy.