2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Adaptive detection and accommodation of communication attacks on infinite dimensional systems with multiple interconnected actuator/sensor pairs

Michael A. Demetriou

Abstract— **The work provides a general model of communication attacks on a networked infinite dimensional system. The system employs a network of inexpensive control units consisting of actuators, sensors and control processors. In an effort to replace a reduced number of expensive high-end actuating and sensing devices implementing an observer-based feedback, the alternate is to use multiple inexpensive actuators/sensors with static output feedback. In order to emulate the performance of the high-end devices, the controllers for the multiple actuator/sensors implement controllers which render the system networked. In doing so, they become prone to communication attacks either as accidental or deliberate actions on the connectivity of the control nodes. A single attack function is proposed which models all types of communication attacks and an adaptive detection scheme is proposed in order to (i) detect the presence of an attack, (ii) diagnose the attack and (iii) accommodate the attack via an appropriate control reconfiguration. The reconfiguration employs the adaptive estimates of the controller gains and restructure the controller adaptively in order to minimize the detrimental effects of the attack on closed-loop performance. Numerical studies on a 1D diffusion PDE employing networked actuator/sensor pairs are included in order to further convey the special architecture of detection and accommodation of networked systems under communication attacks.**

I. Introduction

This paper examines the vulnerability of networked systems under communication attacks and the manner in which the attacks can be handled. The communication attack affects the inter-agent (controller) connectivity and takes the form of sign reversal of a network gain in a given link or nullification of a gain in a given link. Separate from the attack modelling of cyber-physical systems, see [1], [2], [3] and references therein, the attack model here cannot be viewed as either an additive or multiplicative actuator or sensor fault, [4].

The premise is that an infinite dimensional system, often representing partial differential equations, is using a set of inexpensive actuating and sensing devices in place of a small number of high-end/high-capacity actuators and sensors. In opting for the inexpensive devices that render the closed-loop system a networked one by the use of static output feedback controllers that use the weighted sum of the measurements of a subset of the sensors, the networked system becomes vulnerable to communication attacks. In the case of the nullification of a communication link, which corresponds to the zeroing of a gain multiplying a specific sensor output, the attack is viewed as benign. When the sign of a feedback gain of an output corresponding to a given controller is reversed, the attack is viewed as malicious and is classified as an adversarial action whose goal is to destabilize the system.

The modelling and detection of a communication attack in the earlier work [5] is generalized and expanded to include more than one gain (nullified or sign-reversed) in a given control signal. Additionally, a diagnostic observer is proposed whose role is to detect the presence of an attack in the networked system, and to diagnose the type of the attack. Such a diagnosis is made possible via the use of adaptive estimates of the static feedback gains. By the appropriate use of the residual signals and their corresponding time-varying thresholds, an attack is detected the instant any of the residuals exceed their corresponding time-varying thresholds. The use of time-varying thresholds ensures that the detection time, i.e. time between the occurrence of an attack and the declaration of its presence, is minimized. The adaptive estimates of the static gains serve also to accommodate the effects of the attack via the adaptive control reconfiguration.

The proposed infinite dimensional system is summarized in Section II and the static feedback controller is presented in Section III. The attack modelling along with the gain parametrization are given in Section IV. The attack detection, diagnosis and accommodation are presented in Section V and numerical studies are summarized in Section VI.

II. Mathematical framework and motivation

The infinite dimensional system under consideration evolves in a Hilbert space X and given by

$$\dot{x}(t) = Ax(t) + \mathbf{B}u(t), \quad y(t) = \mathbf{C}x(t) \tag{1}$$

Associated with the state space X is the additional space V and its conjugate dual V^*, which is the space of continuous conjugate linear functionals on V. The space V is a reflexive Banach space with norm $\|\cdot\|$, and assume that it is embedded densely and continuously in X. Let the usual uniform operator norm on V^* be denoted by $\|\cdot\|_*$. We then have $V \hookrightarrow X \hookrightarrow V^*$ with the embeddings dense and continuous, [6]. The state operator $A \in \mathcal{L}(V, V^*)$ and the (single) input and (single) output operators are $\mathbf{B} \in \mathcal{L}(\mathbb{R}^1, V^*)$ and $\mathbf{C} \in \mathcal{L}(V, \mathbb{R}^1)$. The state operator is self adjoint and generates an exponentially stable C_0 semigroup [7], and $x(0) = x_0 \in \mathrm{dom}(A)$.

The assumption is that the control devices-(actuator and sensor)-are high capacity and performance devices. It is also

M. A. Demetriou is with Worcester Polytechnic Institute, Aerospace Engineering, Worcester, MA 01609, USA, mdemetri@wpi.edu. The author gratefully acknowledge financial support from NSF-CMMI grant # 1825546.

implicitly assumed that both are high end devices with high costs (fixed and operating) associated with them. To design a controller, one may use LQR/$\mathbb{H}^2$ methods [8] to arrive at

$$u(t) = -\mathbf{K}x(t), \quad \mathbf{K} \in \mathcal{L}(\mathbb{R}^1, V^*). \tag{2}$$

The enabling condition imposed is that the operator pair $(A, \mathbf{B})$ be approximately controllable, [8]. However, even when the controllability condition is satisfied and the operator gain $\mathbf{K}$ is computed, the control law cannot be realized as it requires the infinite dimensional state $x(t)$. This is resolved by implementing a state observer and then replacing $x(t)$ in (2) by its estimate $\widehat{x}(t)$. The state estimate is generated by

$$\dot{\widehat{x}}(t) = A\widehat{x}(t) + \mathbf{B}u(t) + \mathbf{L}\left(y(t) - \mathbf{C}\widehat{x}(t)\right), \ \widehat{x}(0) = \widehat{x}_0 \neq x_0, \tag{3}$$

where the filter gain operator $\mathbf{L} \in \mathcal{L}(\mathbb{R}^1, V^*)$. In parallel to the above condition for controller design, one requires that the operator pair $(A, \mathbf{C})$ be approximately observable, [8]. The filter gain can be obtained by the Kalman filter or the Luenberger observer design. This implementation requires the simultaneous propagation of the observer (3) and possibly its covariance operator in order to implement the controller

$$u(t) = -\mathbf{K}\widehat{x}(t), \quad \mathbf{K} \in \mathcal{L}(\mathbb{R}^1, V^*). \tag{4}$$

Summarizing, (4) requires the following for implementation

- expensive actuating device with input operator $\mathbf{B}$,
- expensive sensing device with output operator $\mathbf{C}$,
- computation of controller gain $\mathbf{K}$,
- computation of filter gain $\mathbf{L}$,
- real time generation of the finite dimensional approximation of (3) and possibly its covariance equation.

In response to the high cost (computational and operating/fixed) requirements associated with the controller (4), one may consider the following inexpensive alternative

- use inexpensive actuating devices with associated input operators B_i, $i = 1, \ldots, N$,
- use inexpensive sensing devices with associated output operators C_i, $i = 1, \ldots, N$,
- use simplified controller architecture based on static feedback.

The economic advantages with the use of inexpensive sensing and actuating devices is straightforward. However, one must take into account their reliability. The computational and algorithmic advantages are seen in the controller architecture where a simple static output feedback controller can be implemented in place of (4).

Using the above considerations, (1) is revisited and is now

$$\begin{aligned} \dot{x}(t) &= Ax(t) + \sum_{i=1}^{N} B_i u_i(t), \\ y(t) &= \begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} = \begin{bmatrix} C_1 x(t) \\ \vdots \\ C_N x(t) \end{bmatrix}. \end{aligned} \tag{5}$$

The N input operators $B_i \in \mathcal{L}(\mathbb{R}^1, V^*)$, describe the manner in which the control signals $u_i(t)$, $i = 1, \ldots, N$ are injected in the system via the inexpensive actuators. Similarly, the N output operators $C_j \in \mathcal{L}(V, \mathbb{R}^1)$, $j = 1, \ldots, N$ describe the manner the state is obtained by the inexpensive sensors. For the regulation case, the static output feedback is given by

$$u_i(t) = -\sum_{j=1}^{N} f_{ij} y_j(t) \tag{6}$$

where f_{ij} are the gains, resulting in the closed-loop system

$$\dot{x}(t) = Ax(t) - \sum_{i=1}^{N} B_i \sum_{j=1}^{N} f_{ij} y_j(t). \tag{7}$$

Closer examination of (7) reveals that the closed loop system requires each control unit to communicate with all other control units via f_{ij}; the i^{th} control unit is comprised of the i^{th} actuator, the i^{th} sensor and the control gains f_{ij}, $i, j = 1, \ldots, N$. While the alternate solution (6) results in reduced fixed costs, reduced computational costs and reduced complexity, it also results in increased communication costs, as it requires an *all-to-all* communication amongst the N control units. To reduce this cost, one must consider a networked controller, namely a controller in which a subset of the feedback gains f_{ij} are nonzero. This selection of the nonzero entries is described by an appropriate communication topology which defines the information exchange between the sensing devices and the control units. An undirected connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is assumed to define the connectivity of the sensors and control units. The graph nodes $\mathcal{V} = \{1, 2, \ldots, N\}$ represent the N control units in (5). The edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ depict the communication links between the control units. For a given control unit, we define the set of neighbors it communicates with by $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}\}$. Finally we use L to denote the nominal graph Laplacian associated with the graph $\mathcal{G}$; it is given by $L = \mathcal{D} - \mathcal{A}$, where $\mathcal{D}$ denotes the degree matrix and $\mathcal{A}$ the adjacency matrix, [9].

Equipped with the background on data and information exchange, we can define the networked version of (6)

$$u_i(t) = -\sum_{j \in \mathcal{N}_i} g_{ij}(y_i(t) - y_j(t)), \quad i = 1, \ldots, N. \tag{8}$$

An equivalent expression involving the graph Laplacian L is

$$u_i(t) = -\sum_{j=1}^{N} L_{ij} k_{ij} y_j(t), \quad i = 1, \ldots, N, \tag{9}$$

where k_{ij}, $i, j = 1 \ldots, N$ are the feedback gains. To reduce the parametrization of the feedback gains, one may consider weighted graph Laplacians, as they provide both the memory of the connectivity between the sensors and the control units, and also absorb the value of the proportional gain. However, when considering the stability of the closed loop system and possible adaptation of the gains during an attack, it is better to consider a separate matrix that will keep track of the connectivity. Different from the earlier work in [5], the matrix to keep track of the "memory" of connectivity amongst the agents/control units is given by

$$M \triangleq \mathbf{I}_N + \mathcal{A}. \tag{10}$$

Using this memory matrix, the controller in (9) is re-written

$$u_i(t) = -\sum_{j=1}^{N} (M \circ K)_{ij} y_j(t), \quad i = 1, \ldots, N, \tag{11}$$

where $\circ$ denotes the Hadamard (entrywise) product [10] between matrices and thus $(M\circ K)_{ij}$ represents the $(ij)^{\text{th}}$ element of the result of the Hadamard product $M\circ K$.

On a first inspection, (9) appears to address all concerns: (i) reduced operating and fixed costs for the controller equipment (actuators and sensors), (ii) simplified controller architecture (proportional controllers via static feedback vs concurrent propagation of state estimator (3), (4)). However, the problem comes at the vulnerability of the proposed controller (5); the connectivity defined by the memory matrix M in (10) is prone to communication failures, either due to adversarial actions or due to accidental disruptions.

Due to the above, one is faced with the following dilemma

1) Keep the attack-immune controller (3), (4) and incur the high costs associated with it (fixed and operating) and computational demands for running in real-time the (finite dimensional approximation of) estimator (3)
2) Keep the low cost and minimal complexity networked controller (9) and consider a scheme to monitor the networked system for possible attacks; subsequently detect, diagnose and accommodate the attacks via an appropriate control reconfiguration.

III. Control design

In the face of prohibitively expensive actuating and sensing devices, the second option appears financially appealing. In order to provide a framework for the detection and accommodation of the network attacks, we must first provide the control objective. Then the model for the network attacks must be provided before a monitoring scheme is presented.

A. Control problem formulation

In order to generalize the control objective and include tracking instead of simple regulation, we consider an idealized behavior of the infinite dimensional system (5) also governed by another infinite dimensional system. Since the alternative to a single controller with reliable sensor and actuator device is a networked controller of the form (9), it is then befitting to view the reference model as a *virtual leader*. Thus one is interested in designing a networked controller similar to (9), so that the state of (5) follows the state of

$$\dot{x}_m(t) = A_m x_m(t) + \sum_{i=1}^{N} B_i r_i(t), \quad x_m(0) = x_{m0}, \tag{12}$$

where $r_i \in L_2(0,\infty)$, $i = 1\ldots,N$ denote the scalar reference signals and the state operator A_m generates an exponentially stable C_0 semigroup $T_m(t)$ in X with $\|T_m(t)\|_X \leq \mu e^{-\alpha t}$, [7].

In the ideal case (no attacks), one must make some admissibility conditions that enable one to find a controller for (5) so that the resulting closed loop system can track the state of (12). In essence, one must make the feasibility condition for the proposed networked controllers

$$u_i(t) = -\sum_{j=1}^{N} (M\circ K)_{ij} y_j(t) + r_i(t). \tag{13}$$

This is stated in the result below.

Lemma 1: Assume that there exists a communication topology with graph Laplacian matrix L and associated static gains $k_{ij} = \{K\}_{ij}$, with $i,j = 1,\ldots,N$, such that

$$A - \sum_{i=1}^{N} B_i \sum_{j=1}^{N} (M\circ K)_{ij} C_j = A_m. \tag{14}$$

Then the networked controllers (13) ensure that

$$\lim_{t\to\infty} |x(t) - x_m(t)| = 0. \tag{15}$$

Proof: Using the fact that A_m generates an exponentially stable semigroup and $r_i \in L_2(0,\infty)$, then we have that the model (12) is well-posed. Now define the tracking error $z(t) = x(t) - x_m(t)$. Using (5) with (13) and (12) we arrive at

$$\begin{aligned}\dot{z}(t) &= Ax(t) + \sum_{i=1}^{N} B_i\Big(-\sum_{i=1}^{N} (M\circ K)_{ij} y_j(t) + r_i(t)\Big)\\ &\quad -A_m x_m(t) - \sum_{i=1}^{N} B_i r_i(t) = A_m z(t)\end{aligned}$$

with $z(0) = x_0 - x_{m0}$ and $z(t) = T_m(t)z(0)$. Since $\|T_m(t)\|_X \leq \mu e^{-\alpha t}$, then (15) is satisfied with $|z(t)| \leq \mu e^{-\alpha t}|z(0)|$. ■

Remark 1: For the controller description above, one usually ensures first that static stabilizability is feasible and subsequently finds simultaneously the best communication topology and the optimal gains associated with the optimal topology. This then defines the closed-loop operator A_m

The condition in (14) falls under the system-theoretic property of *static stabilizability*, which an infinite dimensional system must satisfy in order to argue for the existence of an appropriate topology. Define the matrix operators

$$\mathbb{B} = \begin{bmatrix} B_1 & \ldots & B_N \end{bmatrix}, \quad \mathbb{C} = \begin{bmatrix} C_1 \\ \vdots \\ C_N \end{bmatrix}$$

The results in [11] on static stabilizability of (5) define the static stabilizability of a bounded linear system in terms on the existence of $\mathbb{G} \in \mathcal{L}(\mathcal{Y},\mathcal{U})$ such that the triple $(A - \mathbb{B}\mathbb{G}\mathbb{C}, \mathbb{B}, \mathbb{C})$ is a strongly stable bounded linear system, where $\mathcal{Y}$ and $\mathcal{U}$ denote the output and input spaces, respectively. In this case, $A - \mathbb{B}\mathbb{G}\mathbb{C}$ is the generator of a strongly stable semigroup $T_{\mathbb{G}}(t)$, i.e. $\lim_{t\to\infty} \|T_{\mathbb{G}}(t)x\| = 0$, for all $x \in X$.

The above result immediately define the class of infinite dimensional systems for which the alternate networked controller design in (13) is applicable. If one is only interested in the tracking problem in Lemma 1, then the condition of the state operator generating an exponentially stable C_0 semigroup can be relaxed to the case of generating a C_0 semigroup on X. However, this comes at the condition that $\mathbb{B} \in \mathcal{L}(\mathcal{U},X)$ and $\mathbb{C} \in \mathcal{L}(X,\mathcal{Y})$. If one is simply interested in designing a static output feedback controller of the form (13) without any considerations for attack detection and accommodation, then Lemma 1 can be used with the relaxed conditions on the state operator A. However, when an attack detection and accommodation is considered, then the conditions on the state operator must be strengthened to that of generating an exponentially stable C_0 semigroup on X.

Remark 2: The different controller parameterizations have different structures. Using the parametrization in [5], use the

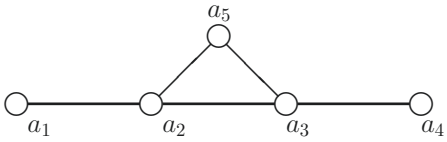

Fig. 1. An example of a connected graph with 5 vertices.

$N \times N$ gain matrix G with $\{G\}_{ij} = g_{ij}$ in the controller (8). Then the graph-weighted gain matrix $\mathbb{G}$ is defined via

$$\{\mathbb{G}\}_{ij} = -\big(\mathcal{A} \circ G\big)_{ij}, \;\; i,j = 1,\ldots,N, j \neq i,$$
$$\{\mathbb{G}\}_{ii} = -\sum_{j \in \mathcal{N}_i, j \neq i} \{\mathbb{G}\}_{ij}, \;\; i = 1,\ldots,N,$$

and the closed-loop system is written as

$$\dot{x}(t) = Ax(t) - \mathbb{B}\mathbb{G}\mathbb{C}x(t) + \mathbb{B}\mathbf{r}(t), \quad x(0) \in \mathrm{dom}(A). \quad (16)$$

Using the proposed controller in (13), with $\mathbb{K} \triangleq M \circ K$, then the closed-loop system is

$$\dot{x}(t) = Ax(t) - \mathbb{B}\mathbb{K}\mathbb{C}x(t) + \mathbb{B}\mathbf{r}(t), \quad x(0) \in \mathrm{dom}(A). \quad (17)$$

In the graph of Figure 1, $\mathbb{G}$ with 10 gains g_{ij} has the form

$$\begin{bmatrix} g_{12} & -g_{12} & 0 & 0 & 0 \\ -g_{21} & g_{21}+g_{23}+g_{25} & -g_{23} & 0 & -g_{25} \\ 0 & -g_{32} & g_{32}+g_{34}+g_{35} & -g_{34} & -g_{35} \\ 0 & 0 & -g_{43} & g_{43} & 0 \\ 0 & -g_{52} & -g_{53} & 0 & g_{52}+g_{53} \end{bmatrix}$$

whereas $\mathbb{K}$ with 15 gains $\mathbb{K}_{ij}$ has the form

$$\begin{bmatrix} \mathbb{K}_{11} & \mathbb{K}_{12} & 0 & 0 & 0 \\ \mathbb{K}_{21} & \mathbb{K}_{22} & \mathbb{K}_{23} & 0 & \mathbb{K}_{25} \\ 0 & \mathbb{K}_{32} & \mathbb{K}_{33} & \mathbb{K}_{34} & \mathbb{K}_{35} \\ 0 & 0 & \mathbb{K}_{43} & \mathbb{K}_{44} & 0 \\ 0 & \mathbb{K}_{52} & \mathbb{K}_{53} & 0 & \mathbb{K}_{55} \end{bmatrix}.$$

IV. Communication attack modeling

As was highlighted in [5] an attack on the networked control units takes the form of a compromised communication link. Such a model is specific to networked systems and cannot be modeled as an actuator or a sensor fault. If the attack (or fault), on a link is intended to cause disruption of the normal operation, then it is classified as a *malicious attack*; this can be associated with an adversarial action on the network. Its goal is to maximize damage to the networked system by possibly destabilizing it. This is realized through the sign reversal of a subset of the signals to a given node by negating the signals to a given control unit. If the fault is attributed to normal wear and equipment aging, then it is classified as a *benign attack*. This is modeled by a severed communication link by removing a graph edge. In the former case (malicious attack), this is equivalent to changing

$$\mathbb{K}_{ij} \longleftarrow -\mathbb{K}_{ij}, \quad \text{for some } j,$$

and in the latter case (benign attack), it is equivalent to setting

$$\mathbb{K}_{ij} \longleftarrow 0, \quad \text{for some } j.$$

Another feature of the attack is its *time profile*, as was considered in [5]. The time profile is defined as *abrupt* thereby designating an abrupt action which may be accidental or deliberate. The time profile is defined as *incipient* designating a slowly evolving action. It turns out that the abrupt case is a limiting case of the incipient case. Thus, a single expression can capture both time profiles. For the severed communication link, the time profile is given by

$$\beta(\tau) = \begin{cases} 1 & \text{if } \tau < 0 \\ e^{-\lambda\tau} & \text{if } \tau \geq 0 \end{cases} = 1 - \big(1 - e^{-\lambda\tau}\big)\mathcal{H}(\tau), \quad (18)$$

where $\mathcal{H}(t)$ denotes the Heaviside step function, i.e. it nullifies an entry of the gain matrix $\mathbb{K}$ since $\lim_{t\to\infty}\beta(t) = 0$. For the sign reversal of signals case, the time profile is

$$\beta(\tau) = \begin{cases} 1 & \text{if } \tau < 0 \\ -1 + e^{-\lambda\tau} & \text{if } \tau \geq 0 \end{cases} = 1 - \big(2 - e^{-\lambda\tau}\big)\mathcal{H}(\tau) \quad (19)$$

and $\lim_{t\to\infty}\beta(t) = -1$. In both cases, when the time profile rate $\lambda \to \infty$, then the incipient profile becomes abrupt since $\beta(t) = 1 - \mathcal{H}(t)$ or $\beta(t) = 1 - 2\mathcal{H}(t)$. This was used in [5] to model a single fault, i.e. a *single element* of $\mathbb{K}_{ij}$, $j = 1,\ldots,N$ was either nullified or had its sign reversed. To account for *more than one* entries of the i^{th} row of $\mathbb{K}$, define the sets

$$\mathcal{S}_i^a = \{j : \beta_{ij} \neq 0\}, \quad \mathcal{S}_i^h = \mathcal{S}_i \setminus \mathcal{S}_i^a, \quad \mathcal{S}_i = \{j : M_{ij} \neq 0\}$$

where $\mathcal{S}_i$ denotes the set of all non-zero entries of the matrix M, i.e. the set of nonzero entries of the i^{th} row of $\mathbb{K}$. Thus $\mathcal{S}_i^a$ denotes the set of elements of the i^{th} row of $\mathbb{K}$ that are attacked and its complement $\mathcal{S}_i^h$ denotes the set of elements of the i^{th} row of $\mathbb{K}$ that are healthy. Denote the time that the attack occurs, i.e. the *attack instance*, by t_a. Then the closed-loop system resulting from (5), (13) with attacks is

$$\begin{aligned} \dot{x}(t) = Ax(t) + \sum_{i=1}^{N} B_i r_i \\ - \sum_{i=1}^{N} B_i \Big(\sum_{j \in \mathbb{S}_i^h} \mathbb{K}_{ij} y_j(t) + \sum_{j \in \mathbb{S}_i^a} \beta_{ij}(t - t_a)\mathbb{K}_{ij} y_j(t) \Big). \end{aligned} \quad (20)$$

V. Attack detection, diagnosis and accommodation

If only the presence of an attack in the networked system (20) is desired to be detected, then a detection observer suffices and its architecture is simple. In fact, the virtual leader can serve as a detection observer. This was presented in [5] and it simply summarized here. Prior to the occurrence of an attack, the expression for the closed-loop system (20) is modified to include disturbances and is

$$\begin{aligned} \dot{x}(t) = Ax(t) + Ew(t) + \sum_{i=1}^{N} B_i r_i - \sum_{i=1}^{N} B_i \sum_{j \in \mathbb{S}_i^h} \mathbb{K}_{ij} y_j(t) \\ - \sum_{i=1}^{N} B_i \sum_{j \in \mathbb{S}_i^a} \beta_{ij}(t - t_a)\mathbb{K}_{ij} y_j(t), \end{aligned} \quad (21)$$

with $E \in \mathcal{L}(\mathbb{R}^1, V^*)$ the disturbance operator and $w \in L_2(0,\infty;\mathbb{R}^1)$ the disturbance signal. In this case the error between (21) and (12), denoted by $\upsilon(t) = x(t) - x_m(t)$, is

$$\dot{\upsilon}(t) = A_m\upsilon(t) + Ew(t), \;\; \upsilon(0) \neq 0 \quad t \in [0, t_a). \quad (22)$$

Its solution is given by

$$\upsilon(t) = T_m(t)\upsilon(0) + \int_0^t T_m(t - \tau)Ew(\tau)\,\mathrm{d}\tau, \quad t \in [0, t_a).$$

Since $T_m(t)$ is an exponentially stable semigroup and w is square integrable, then $\lim_{t\to\infty}|\upsilon(t)|=0$. The scalar *residuals*

$$\pi_i(t)=C_iT_m(t)\upsilon(0)+\int_0^t C_iT_m(t-\tau)Ew(\tau)\,\mathrm{d}\tau,$$

provide the *time varying thresholds*

$$\rho_i(t)=|C_i|\mu\left(|\upsilon(0)|e^{-\alpha t}+|E||w|_\infty\frac{e^{-\alpha t}-1}{\alpha}\right).$$

After the occurrence of the attack, the error $\upsilon(t)$ attains a different structure and the residuals deviate/exceed the thresholds, thus declaring the presence of an attack. In order to diagnose and accommodate the attack, an adaptive detection and diagnostic observer is required. Prior to the presence of an attack, it resembles the aforementioned detection observer. However, after the attack on the networked system (21) is declared, the adaptive estimates of the static gains $\mathbb{K}_{ij}$ are activated and subsequently used to accommodate the attack.

A. Adaptive detection and diagnostic observer

The adaptive detection/diagnostic observer is given by

$$\begin{aligned}\dot{\widehat{x}}(t)&=A\widehat{x}(t)-\mathbb{H}\left(y(t)-\mathbb{C}\widehat{x}(t)\right)+\mathbb{B}\mathbf{r}(t)\\&\quad-\sum_{i=1}^{N}B_i\sum_{j=1}^{N}\widehat{\mathbb{K}}_{ij}(t)y_j(t)\end{aligned}\tag{23}$$

where $\widehat{\mathbb{K}}_{ij}(t)$ denote the adaptive estimates of $\mathbb{K}_{ij}$, and $\mathbb{H}\in\mathcal{L}(\mathbb{R}^N,V^*)$ denotes the observer gain selected so that $A-\mathbb{H}\mathbb{C}$ is the generator of an exponentially stable C_0 semigroup $T_o(t)$ on X with $\|T_o(t)\|_X\le\kappa e^{-\zeta t}$. To help with the derivation of the adaptive laws, consider the control terms in (21)

$$\begin{aligned}&\sum_{i=1}^{N}B_i\Big(\sum_{j\in\mathbb{S}_i^h}\mathbb{K}_{ij}y_j(t)+\sum_{j\in\mathbb{S}_i^a}\beta_{ij}(t-t_a)\mathbb{K}_{ij}y_j(t)\Big)=\\&\quad\sum_{i=1}^{N}B_i\Big(\sum_{j\in\mathbb{S}_i}\mathbb{K}_{ij}y_j(t)+\sum_{j\in\mathbb{S}_i^a}(\beta_{ij}(t-t_a)-1)\mathbb{K}_{ij}y_j(t)\Big).\end{aligned}\tag{24}$$

Using the above, it is easily seen that the *detection error* $e(t)=x(t)-\widehat{x}(t)$ for $t<t_a$ is governed by

$$\begin{aligned}&\dot{e}(t)=\big(A-\mathbb{H}\mathbb{C}\big)e(t)+Ew(t)-\sum_{i=1}^{N}B_i\sum_{j=1}^{N}\widetilde{\mathbb{K}}_{ij}(t)y_j(t)\\&e(0)=e_0,\ \ \widetilde{\mathbb{K}}_{ij}(0)=0,\ \ \widetilde{\mathbb{K}}_{ij}(t)=\mathbb{K}_{ij}-\widetilde{\mathbb{K}}_{ij}(t).\end{aligned}\tag{25}$$

The above error is complemented with the update laws of the adaptive estimates $\widehat{\mathbb{K}}_{ij}(t)$. However, it must be ensured that they do not adapt till after the presence of an attack. In the presence of a nonzero disturbance $w(t)$, the choice $\widehat{\mathbb{K}}_{ij}(0)=\mathbb{K}_{ij}$ does not guarantee that $\widehat{\mathbb{K}}_{ij}(t)=\mathbb{K}_{ij}$ for $t<t_a$. Nonzero values of the parameter errors $\widehat{\mathbb{K}}_{ij}(t)$ will immediately force the residuals to exceed the thresholds and thus falsely declare the presence of an attack. This is achieved with a projection scheme [12] applied on the standard adaptive law

$$\dot{\widehat{\mathbb{K}}}_{ij}(t)=\gamma_{ij}\left(\varepsilon_i(t)y_j(t)-\widetilde{\mathbb{K}}_{ij}(t)\right),\quad\widehat{\mathbb{K}}_{ij}(0)=\mathbb{K}_{ij},$$

and modifying it to

$$\dot{\widehat{\mathbb{K}}}_{ij}(t)=\mathrm{Pr}\left\{\gamma_{ij}\left(\varepsilon_i(t)y_j(t)-\widetilde{\mathbb{K}}_{ij}(t)\right)\right\},\ \ \widehat{\mathbb{K}}_{ij}(0)=\mathbb{K}_{ij}.\tag{26}$$

The projection modification[13] ensures that no adaptation of the adaptive gains $\widehat{\mathbb{K}}_{ij}(t)$ takes place prior to $t<t_a$ for any values of the *residuals*

$$\varepsilon_i(t)=C_ie(t)=C_iT_o(t)e(0)+\int_0^t C_iT_o(t-\tau)Ew(\tau)\,\mathrm{d}\tau,\tag{27}$$

for $i=1,\dots,N$, in a given set. Prior to the presence of an attack, the residuals fall below their corresponding *time varying thresholds*

$$\sigma_i(t)=|C_i|\kappa\Big(|e(0)|_Xe^{-\zeta t}+|E|_X\|w\|_\infty\frac{e^{-\zeta t}-1}{\zeta}\Big).\tag{28}$$

In other words, prior to the instance where any of the residuals $\varepsilon_i(t)$ exceed their corresponding residuals $\sigma_i(t)$, the adaptive gains satisfy $\widehat{\mathbb{K}}_{ij}(t)=\mathbb{K}_{ij},\ \forall t\le\tau_a$ where τ_a denotes the *attack declaration time*, defined below. The attack declaration time $\tau_a\ge t_a$ is the time when the presence of an attack in the networked system is declared.

B. Attack accommodation

The accommodation is achieved via control reconfiguration using the estimates of $\mathbb{K}_{ij}$. Two similar control reconfiguration policies are considered. In the first one, the controller for the k^{th} unit is allowed to establish new connections with all controllers, even those not previously connected, and thus

$$u_k(t)\leftarrow-\sum_{j=1}^{N}\widehat{\mathbb{K}}_{kj}(t)y_j(t).\tag{29}$$

If new connections are not allowed, the original connections to the k^{th} controller are preserved and the reconfiguration is

$$u_k(t)\leftarrow-\sum_{j=1}^{N}\Big(M_{kj}\widehat{\mathbb{K}}_{kj}(t)\Big)y_j(t).\tag{30}$$

This is presented in the lemma below.

Lemma 2: Assume that the state operator A in (21) generates an exponentially stable C_0 semigroup. For a prescribed communication topology defined by the graph $\mathcal{G}$ with memory matrix given by (10), assume that the triple $(A,\mathbb{B},\mathbb{C})$ is statically stabilizable guaranteeing the existence of a static gain K such that the operator defined in (14) generates an exponentially stable C_0 semigroup on X. Further assume that the attack, representing a compromised communication link either due to an accidental (18) or a malicious (19) action, is described by (21). Then the observer (23) and the adaptive gains (26) ensure that prior to the presence of an attack, the N residuals defined by (27) are always below their corresponding thresholds (28). At the presence of an attack, the residuals march toward the boundary of the thresholds and eventually exceed them thereby detecting and declaring the presence of an attack at the *attack declaration time* τ_a

$$\tau_a=\min\left\{\arg_{t\ge t_a}\{t:|\varepsilon_i(t)|\ge\sigma_i(t),i=1,\dots,N\}\right\}.\tag{31}$$

Furthermore, the compromised link is diagnosed via

$$a_k=\arg_{i=1,\dots,N}\left\{|\varepsilon_i(t)|\ge\sigma_i(t)\right\}.\tag{32}$$

Once the compromised controller is diagnosed, then either (29) or (30) can be used for accommodation and thus ensure that $\lim_{\tau_a<t\to\infty}|x(t)-\widehat{x}(t)|=0$ and $\lim_{\tau_a<t\to\infty}|x(t)-x_m(t)|=0$.

VI. Numerical Results

Consider the PDE with $N=5$ collocated actuators/sensors

$$\frac{\partial x(t,\xi)}{\partial t} = 0.1\frac{\partial x^2(t,\xi)}{\partial \xi^2} + \sum_{i=1}^{5} b_i(\xi)u_i(t)$$
$$x(t,0) = x(t,1) = 0, \quad x(0,\xi) = x_0(\xi),$$

$$y(t) = \begin{bmatrix} y_1(t) & \dots & y_5(t) \end{bmatrix}, \quad y_i(t) = \int_0^1 b_i(\xi)x(t,\xi)\,\mathrm{d}\xi$$

The input distribution functions $b_i(\xi)$ were taken to be the same as in [5] and with initial condition $x_0(\xi) = 39.4\sin(1.3\pi\xi)e^{-7\xi^2}$. Using the graph Laplacian

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & 0 & 2 \end{bmatrix}$$

the matrix $\mathbb{K}$ was set as $\mathbb{K} = L$. To demonstrate the effectiveness of the monitoring scheme, the attack was assumed an adversarial action taking place at $t_a = 0.7$s negating the gains $\mathbb{K}_{32}$ and $\mathbb{K}_{35}$ resulting in the post-attack gain matrix

$$\mathbb{K} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 \\ 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & 0 & 2 \end{bmatrix}.$$

Before the attack one has $\lambda_{max}(A - \mathbb{BKC}) = -1.3443$ and after the attack $\lambda_{max}(A - \mathbb{BKC}) = 1.6447$, where $\lambda_{max}(\cdot)$ denotes the largest eigenvalue.

The time varying thresholds (28) are used to detect the presence of an attack in the networked system. The presence of an attack is detected at $\tau_a = 1.048$s resulting in an attack delay time of 0.348s. At the onset of the attack detection, the controller of the compromised control unit is reconfigured to

$$u_3(t) = -\widehat{K}_{32}(t)y_2(t) - \widehat{K}_{33}(t)y_3(t) - \widehat{K}_{34}(t)y_4(t) - \widehat{K}_{35}(t)y_5(t)$$

Figure 2 depicts the norm of the output error and it is observed that it falls below the time varying threshold before the attack. When it exceeds the threshold, the presence of the attack is declared. When the accommodation is not activated, then the system becomes unstable and $\varepsilon \to \infty$. The same is also observed in Figure 3, which depicts the evolution of the state L_2 norm. That also exhibits the same behavior with regards to the presence of an attack and also the negative effects of the attack when it is not accommodated.

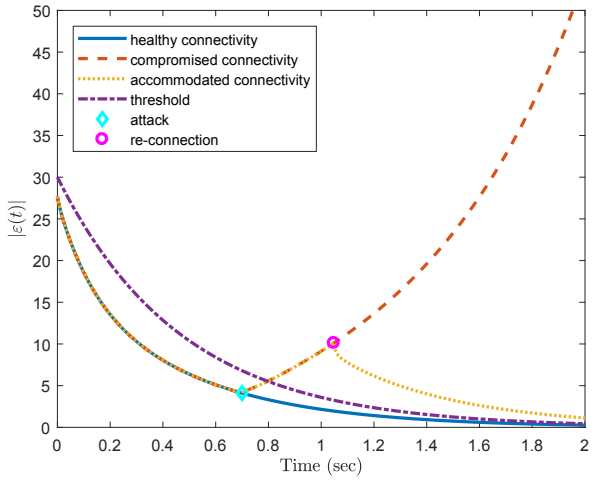

Fig. 2. Evolution of (Euclidean) norm of output detection error $\varepsilon(t)$.

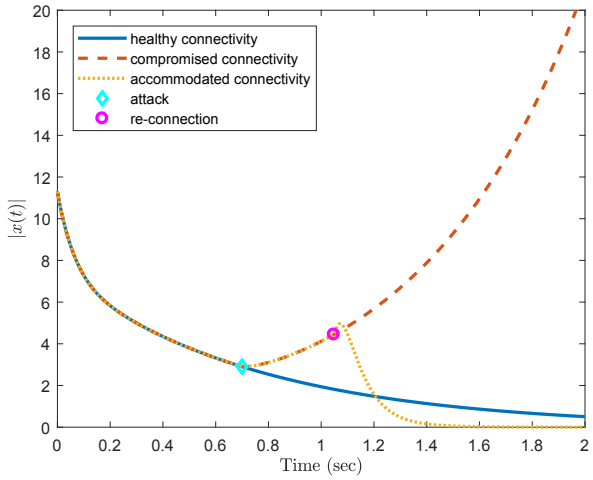

Fig. 3. Evolution of state L_2 norm.

References

[1] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. on Automatic Control*, vol. 58(11), pp. 2715–2729, 2013.

[2] K. Paridari, N. O'Mahony, A. El-Din Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, "A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration," *Proc. of the IEEE*, vol. 106, no. 1, pp. 113–128, Jan 2018.

[3] S. Weerakkody, X. Liu, S. H. Son, and B. Sinopoli, "A graph theoretic characterization of perfect attackability and detection in distributed control systems," in *Proc. of the American Control Conference*, July 2016, pp. 1171–1178.

[4] R. Isermann, *Fault-diagnosis applications*. Springer, Heidelberg, 2011, model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems.

[5] M. A. Demetriou, "Detection of communication attacks on spatially distributed systems with multiple interconnected actuator/sensor pairs," in *Proc. of the IEEE Conf. on Decision and Control*, Miami Beach, FL, USA, Dec. 17-19 2018, pp. 2896–2901.

[6] R. E. Showalter, *Hilbert Space Methods for Partial Differential Equations*. London: Pitman, 1977.

[7] A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, ser. Applied Mathematical Sciences. New York: Springer-Verlag, 1983, vol. 44.

[8] R. F. Curtain and H. J. Zwart, *An Introduction to Infinite Dimensional Linear Systems Theory*. Berlin: Springer-Verlag, 1995.

[9] C. Godsil and G. Royle, *Algebraic graph theory*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 2001, vol. 207.

[10] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd ed. Cambridge: Cambridge University Press, 2013.

[11] J. C. Oostveen and R. F. Curtain, "Robustly stabilizing controllers for dissipative infinite-dimensional systems with collocated actuators and sensors," *Automatica*, vol. 36(3), pp. 337–348, 2000.

[12] M. A. Demetriou and I. G. Rosen, "On-line robust parameter identification for parabolic systems," *International Journal of Adaptive Control and Signal Processing*, vol. 15(6), pp. 615–631, 2001.

[13] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Englewood Cliffs, NJ: Prentice Hall, 1995.