Cooperative Design of Systems of Systems Against Attack on One Subsystem

Shahriar Talebi, *Student Member, IEEE,* Marwan A. Simaan, *Life Fellow, IEEE,* Zhihua Qu, *Fellow, IEEE,*

***Abstract*—Future large-scale control systems will be cyber-physical System of Systems (SoS)s each consisting of many interacting subsystems operating harmoniously and making decisions that benefit the entire enterprise. In terms of accomplishing their own objectives, the subsystems in an SoS will be designed so that their decision-making process achieves a Nash equilibrium for the entire SoS. This equilibrium ensures that no one subsystem can improve its position by deviating from the agreed decision. One of the weaknesses of this equilibrium is that it is vulnerable to cyber-attacks directed towards the objectives of one subsystem. In this paper, we consider an attack on a linear quadratic SoS that consists of altering some parameters in the objective function of one subsystem by injecting negative definite matrices into the attacked subsystem's objective function. Such an attack would be designed to disrupt the operation of the entire SoS by pitting one subsystem against the others and causing a change in the behavior and trajectory of the entire SoS. We consider two possible outcomes of the attack. The first is when the attack is not detected and all subsystems continue to operate as designed but now reaching a different Nash equilibrium that may or may not yield the outcome desired by the attacker. In this case, existence results related to the new Nash equilibrium will be derived. The second is when the attack is detected causing the remaining subsystems to react by forming a cooperative team to counter the actions of the attacked subsystem. The team will then establish a new Nash equilibrium with the attacked subsystem. Both of these possible outcomes of the cyber-attack will be analyzed. A 3-subsystems SoS example illustrating the results will also be presented. This example illustrates the ability of an attacker to drive the SoS's state trajectory away from the origin if the attack on one subsystem is not detected. It also illustrates the advantage that the remaining subsystems have in restoring the trajectory to one that is close to the pre-attack equilibrium trajectory by cooperating as a team if the attack is detected.**

I. Introduction

A system of systems (SoS) is a large enterprise that consists of many independent but interacting subsystems who make collaborative decisions based on their shared objectives that enhance the performance of the entire enterprise. The SoS offers more functionality to its constituent subsystems than when they operate on their own as isolated systems. Future SoSs will be cyber-physical-human in nature consisting of machines, distributed human and software agents and virtual subsystems equipped with artificial intelligence and machine learning all interconnected through technologies such as the Internet of Things (IoT). The requirements for keeping these SoSs operational will be more stringent than what is currently required for conventional isolated systems. The SoS will operate optimally when all subsystems have a harmonious non-conflicting relationship that governs their collective behavior. A reasonable decision process for the entire SoS is one that yields what is known as a Nash equilibrium among its subsystems [1]. Such an equilibrium guarantees all subsystems that no one subsystem can achieve an advantage over the others by altering its decision making process. Another advantage of this equilibrium is that the resulting SoS will evolves in a way that is agreeable to all subsystems. One of its main weaknesses, however, is that the entire functionality of the SoS can be easily disrupted if the objectives of one subsystem are altered in such a way that they become inconsistent with the objectives of the remaining subsystems. A change in the objectives of one subsystem can be easily initiated by a malicious cyber-attack especially if the subsystem is vulnerable and does not have the necessary protections. When this happens, the operation of the entire SoS suffers resulting in an adversarial environment that may affect not only the behavior of the attacked subsystem but also possibly the behavior of the other subsystems as well. The disruption caused by an attack on one subsystem may cause the remaining subsystems either to adjust to the new adversarial environment by reaching a new Nash equilibrium with the attacked subsystem or to react by taking a unified team action to counter the effect of the attacked subsystem.

The consequences of one subsystem in an SoS changing its behavior in a manner that is inconsistent and adversarial with the other subsystems have not been extensively studied in the control systems literature. In the differential games literature [2]–[7], the proposed SoS framework has recently been considered in studying multi-pursuers one-evader problems where the evader, who is trying to escape, has an objective function that is directly opposed to the objective functions of several pursuers who collectively are trying to capture it [8]. It has also been applied to study all-against-one linear quadratic differential games where one player in a multi-player game decides to dissent by modifying certain matrices in its objective function from positive definite to negative definite [9], [10]. This change causes the remaining players to consider several options to deal collectively with the adversarial environment that results from the dissension of one player.

The proposed SoS framework also applies to cyber-physical-

The authors are with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL. Emails: shahriar@ece.ucf.edu, simaan@ucf.edu, qu@ucf.edu.

This work was supported in part by the US National Science Foundation under grants ECCS-1308928.

human systems such as smart grids in which consumers and energy producers as well as potential attackers interact at both physical and cyber layers [11]. Game theoretical approaches involving Nash or Stackelberg equilibria can be formulated depending on the cyber-physical structure as well as specific malicious actions [12]. Even if the cyber information structure is local, cooperative SoSs can collaboratively defend against attacks [13], and their distributed defense mechanism also has a game theoretical interpretation [14].

The paper is organized as follows: In section II we describe the SoS model and the attack on one of its subsystems designed to modify certain parameters in its objective function. In section III, we consider the case when the attack is not detected. We derive necessary conditions for definiteness of the solution of the Riccati equation with $S_{1f} < 0$ and $Q_1 < 0$ and for exponential boundedness of the resulting SoS state trajectory. In section IV, we consider the case when the attack is detected. We discuss the design of a team strategy among the remaining subsystems that forms a Nash equilibrium with the strategy of the attacked subsystem. In section V, we present an illustrative example of a system of three subsystems when one subsystem is subjected to an attack that transforms it into an adversary subsystem with the remaining two subsystems. Both cases, when the attack is and is not detected are analyzed. Concluding remarks are presented in section VI.

Notation: The real maximum (minimum) eigenvalue of a symmetric $n \times n$ matrix $Q(t)$ at each instant of time t is denoted by $\lambda_Q^{max}(t)$ ($\lambda_Q^{min}(t)$). The maximal (minimal) eigenvalue of the same matrix $Q(t)$ over the interval $[t_0, t_f]$ is a real constant defined as $\bar{\lambda}_Q := \max_{t\in[t_0,t_f]} \lambda_Q^{max}(t)$ ($\underline{\lambda}_Q := \min_{t\in[t_0,t_f]} \lambda_Q^{min}(t)$). The Euclidean norm of a vector v is denoted by $\|v\|$. Function $blkdiag\{Q_1, \ldots, Q_m\}$ constructs a larger matrix with diagonal blocks consist of matrices Q_i. Matrix I_n is the identity matrix of dimension n.

II. System of Systems Formulation

Without loss of generality, let us consider an LQ control SoS described by the linear differential equation:

$$\dot{x} = Ax + \textstyle\sum_{j=1}^{M} B_j u_j, \quad x(t_0) = x_0; \tag{1}$$

where x is the state and u_i is control vector of subsystem i whose objective is to minimize the quadratic function:

$$J_i = \frac{1}{2} x^\intercal(t_f) S_{if} x(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left[x^\intercal Q_i x + u_i^\intercal R_i u_i\right] dt \tag{2}$$

for $i = 1, 2, \cdots, M$. Matrices A, B_i, Q_i and R_i are all matrices of proper dimensions and satisfy $S_{if} \geq 0$ and $Q_i \geq 0$ (positive semidefinite) and $R_i > 0$ (positive definite). Under normal operating conditions, it is well known that a closed-loop Nash equilibrium of this SoS is accomplished if each subsystem implements its control as:

$$u_i^* = -R_i^{-1} B_i^\intercal S_i x \tag{3}$$

for $i = 1, 2, \cdots, M$, where S_i's satisfy the following M-coupled differential Riccati equations [4]:

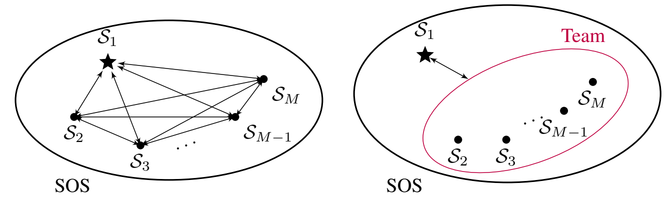


Fig. 1. SoS consisting of subsystems $\mathcal{S}_1, \cdots, \mathcal{S}_M$ where $\mathcal{S}_1$ is attacked; (a) Attack is not detected and subsystems continue to operate as before; (b) Attack is detected and remaining subsystems react by forming a team.

$$\begin{aligned}\dot{S}_i + S_i A + A^\intercal S_i + Q_i + S_i H_i S_i \\ - \textstyle\sum_{j=1}^{M} \left(S_i H_j S_j + S_j H_j S_i\right) = 0,\end{aligned} \tag{4}$$

with boundary conditions $S_i(t_f) = S_{if}$, and $H_i = B_i R_i^{-1} B_i^\intercal$. The resulting system trajectory will satisfy

$$\dot{x} = \overline{A} x, \quad \overline{A} = A - \textstyle\sum_{j=1}^{M} H_j S_j. \tag{5}$$

In a cyber-physical SoS, the coupled Riccati equations in (4) are solved by a separate Riccati Equations Solver (RES) unit that determines the matrices S_i's and transmits each S_i to the controller of its corresponding ith subsystem who then implements its closed-loop control as indicated in (3). The entire SoS will then evolve according to (5). By choosing positive semidefinite matrices $S_{if} \geq 0$ and $Q_i \geq 0$ all subsystems are expressing a common desire to regulate the SoS by bringing its state vector as close as possible to the origin, albeit using their own individual preferences through their own choices of S_{if} and Q_i.

Now let us assume that one subsystem's decision-making process (say subsystem 1) is attacked. The attack consists of altering the matrices S_{1f} and Q_1 in the subsystem's objective function by making them negative definite ($S_{1f} < 0$ and $Q_1 < 0$) in order to make this subsystem adversarial to the other subsystems. Such an attack would be designed to destabilize the entire SoS and to prevent it from being regulated by driving its trajectory away from the origin. An attack of this nature can be accomplished by an attacker accessing the computer code in the RES Unit and introducing the necessary change in matrices S_{1f} and Q_1.

This new structure of the SoS puts subsystems $2, \ldots, M$ directly in conflict with subsystem 1. That is, while subsystems $2, \ldots, M$ are trying to regulate the trajectory of the SoS by minimizing its deviation from the origin as in the intended Nash equilibrium, subsystem 1 now has an objective whose goal is to de-regulate (or de-stabilize) the SoS by maximizing the deviation of the trajectory away from the origin. In this paper, we consider two possible outcomes as a result of this attack as illustrated in Fig. 1. The first, Fig 1(a), is when the attack is not detected (i.e. change in matrices to $S_{1f} < 0$ and $Q_1 < 0$ is not detected in the RES unit) and the decision variables are implemented as in (3) except that the S_i matrices that the controllers receive from the RES unit are the solution of (4) with $S_{1f} < 0$ and $Q_1 < 0$. In this case, two theorems will be derived that address the definiteness of the solution of (4) and the boundedness of the resulting trajectory (5). The second, Fig 1(b), is when the attack is detected and the subsystems react by organizing as a team against the attacked subsystem. The derivation of the team decision process and the formation of the Nash strategy against the attacked subsystem will be considered. Both of these cases will be discussed in the next two sections.

III. Undetected Attack

For the standard SoS structure where $S_{if} \geq 0$ and $Q_i \geq 0$ for $i = 1, 2, \cdots, M$, it is well known [15] that all matrices satisfying (4) will be positive semi-definite. That is, $S_i(t) \geq 0$ for $t \in [t_0, t_f]$ and for $i = 1, 2, \cdots, M$. When the attack is not detected and the matrices in the objective function of the attacked subsystem change $S_{1f} < 0$ and $Q_1 < 0$ the above result will no longer be valid. The following theorem provides the equivalent result for the new S_i matrices.

Theorem 1. *For the attacked SoS with $S_{1f} < 0$ and $Q_1 < 0$, and $S_{if} \geq 0$ and $Q_i \geq 0$ for $i = 2, 3, \cdots, M$, the solutions of (4) will satisfy $S_1(t) < 0$ and $S_i(t) \geq 0$ for $i = 2, 3, \cdots, M$ for all $t \in [t_0, t_f]$.*

Proof. Let $\lambda_{S_1}^{max}(t)$ and $v(t)$ be the maximum eigenvalue and corresponding unit eigenvector of $S_1(t)$. The matrix $S_1(t)$ is piecewise continuously differentiable and symmetric for $t \in [t_0, t_f]$ but not necessarily analytic. Thus, its eigenvectors (or eigenspaces) are not necessarily continuous (therefore not differentiable) [1]. Continuity (differentiability) of the eigenvectors of $S_1(t)$ requires $S_1(t)$ to be holomorphic (analytic) (see [16], Theorem 1.10). However, in our case we only have continuity of $S_1(t)$ and thus we cannot differentiate $v(t)$. Instead, we can show [see [15], Theorem 3.6.1] that $\lambda_{S_1}^{max}(t)$ is not only continuous but also differentiable almost everywhere, and at any point that it is differentiable its derivative satisfies

$$\begin{aligned}\tfrac{d}{dt}\lambda_{S_1}^{max}(t) =& v^\intercal(t)\dot{S}_1 v(t)\\ =& -v^\intercal(t)\left(S_1 A + A^\intercal S_1 + S_1 H_1 S_1\right)v(t)\\ &+ \textstyle\sum_{j=1}^{M}\left(v^\intercal(t) S_1 H_j S_j v(t) + v^\intercal(t) S_j H_j S_1 v(t)\right)\\ &- v^\intercal(t) Q_1 v(t)\end{aligned}$$

which can be simplified as

$$\begin{aligned}\tfrac{d}{dt}\lambda_{S_1}^{max}(t) = \lambda_{S_1}^{max}(t) v^\intercal(t)\Big[&-A - A^\intercal - \lambda_{S_1}^{max}(t) H_1\\ &+ \textstyle\sum_{j=1}^{M}\left(H_j S_j + S_j H_j\right)\Big] v(t) - v^\intercal(t) Q_1 v(t).\end{aligned}$$

Now for sufficiently small $\lambda_{S_1}^{max}(t)$ we have

$$\tfrac{d}{dt}\lambda_{S_1}^{max}(t) \approx -v^\intercal(t) Q_1 v(t) > 0 \tag{6}$$

since $Q_1 < 0$. Given that $S_1(t_f) = S_{1f} < 0$, we know $\lambda_{S_1}^{max}(t_f) < 0$ and hence by (6) and the continuity of $\lambda_{S_1}^{max}(t)$, the maximal eigenvalue $\lambda_{S_1}^{max}(t) < 0$ for all $t \in [t_0, t_f]$. This proves that $\lambda_{S_1}^{max}(t)$ starts from a negative value at $t = t_f$ and stays negative as it evolves backward in time. Noting that $\lambda_{S_1}^{max}(t)$ is the maximum eigenvalue of $S_1(t)$, this proves that $S_1(t) < 0$ for all $t \in [t_0, t_f]$. Now, for $i = 2, 3, \cdots, M$, let $\overline{A}$ be as defined in (5) then (4) can be written as

$$\dot{S}_i + S_i\overline{A} + \overline{A}^\intercal S_i + Q_i + S_i H_i S_i = 0$$

with $S_i(t_f) = S_{if}$. Let $\Psi(t,\tau)$ be the state transition matrix of $-\overline{A}^\intercal(t)$. It is known that for $t, \tau \in [t_0, t_f]$,

$$\tfrac{\partial}{\partial t}\Psi(t,\tau) = -\overline{A}^\intercal(t)\Psi(t,\tau), \quad \Psi(\tau,\tau) = I_n; \tag{7}$$

Then $S_i(t)$ will satisfy

$$\begin{aligned}S_i(t) =& \Psi(t,t_f) S_i(t_f) \Psi^\intercal(t,t_f)\\ &+ \int_t^{t_f}\Psi(t,\tau)[Q_i + S_i H_i S_i]\Psi^\intercal(t,\tau)d\tau.\end{aligned} \tag{8}$$

Now, since $S_i(t_f) \geq 0$, $Q_i \geq 0$ and $R_i > 0$ for $i = 2, 3, \cdots, M$, it follows from (8) that for all $t \in [t_0, t_f]$ we have $S_i(t) \geq 0$ for $i = 2, 3, \cdots, M$. This completes the proof. □

We provide the following lemma which is used in the proof of the next corollary and Theorem 2. Let,

$$P(t) = S_1(t) + \textstyle\sum_{i=2}^{M} S_i(t) \tag{9}$$

$\{S_i(t),\ i = 1, 2, \cdots, M\}$ satisfying (4).

Lemma 1. *For the attacked SoS, if $Q_1 + \sum_{i=2}^{M} Q_i > 0$ and $S_{1f} + \sum_{i=2}^{M} S_{if} > 0$, then for every solution set $\{S_i(t),\ i = 1, 2, \cdots, M\}$ satisfying (4) it follows that $P(t) > 0$ for all $t \in [t_0, t_f]$.*

Proof. For every solution set $\{S_i(t),\ i = 1, 2, \cdots, M\}$ that satisfies (4), let then it would satisfy the following

$$\dot{P} + P\overline{A} + \overline{A}^\intercal P + \textstyle\sum_{i=1}^{M} Q_i + \sum_{i=1}^{M} S_i H_i S_i = 0 \tag{10}$$

with $P(t_f) = S_{1f} + \sum_{i=2}^{M} S_{if}$. Then $P(t)$ will satisfy

$$\begin{aligned}P(t) = \Psi(t,t_f) P(t_f) \Psi^\intercal(t,t_f) + \int_t^{t_f}\Psi(t,\tau)\Big[&Q_1\\ + \textstyle\sum_{i=2}^{M} Q_i + S_1 H_1 S_1 + \sum_{i=2}^{M} S_i H_i S_i&\Big]\Psi^\intercal(t,\tau)d\tau.\end{aligned} \tag{11}$$

where $\Psi(t,\tau)$ is as in (7). Since matrices $R_i > 0$ for $i = 1, 2, \ldots, M$, this yields $H_i \geq 0$, and from the hypothesis of the lemma and (11), it follows that $P(t) > 0$ for all $t \in [t_0, t_f]$. This completes the proof. □

The next theorem provides a sufficient condition for the boundedness of the SoSs state trajectory. Essentially, this theorem provides conditions that guarantee that the attacker cannot accomplish its goal of destabilizing the SoS.

Theorem 2. *For the attacked SoS with $S_{1f} < 0$ and $Q_1 < 0$, if $Q_1 + \sum_{i=2}^{M} Q_i > 0$ and $S_{1f} + \sum_{i=2}^{M} S_{if} > 0$, then the state trajectory $x(t)$ in (5) will be exponentially bounded and for any $r > 0$ it follows that $\|x(t_f)\| \leq r$ provided that*

$$t_f - t_0 \geq \frac{\overline{\lambda}_P}{\underline{\lambda}_Q}\left\{2\ln\left(\frac{\|x_0\|}{r}\right) + \ln\left(\frac{\overline{\lambda}_P}{\underline{\lambda}_P}\right)\right\}.$$

Proof. Let

$$V(t) = x^\intercal(t) P(t) x(t),$$

[1]An example of such a matrix is presented in [Kato 16, p. 111]

where the state vector $x(t)$ is defined in (1) and $P(t)$ is as in (9). Matrix $P(t)$ is positive definite for all t based on Lemma 1. Then, using (5) and (10) it follows that

$$\dot{V}(t) = -x^{\intercal}\left(\sum_{i=1}^{M}(Q_i + S_i H_i S_i)\right)x$$

Since, $Q_1 + \sum_{i=2}^{M} Q_i > 0$ and $H_i \geq 0$ for $i = 1, 2, \ldots, M$ it follows that $\dot{V}(t) < 0$ which yields a decaying dynamics for the system. Now, defining $Q = \sum_{i=1}^{M} Q_i$ implies that

$$\dot{V} \leq -x^{\intercal}Qx \leq -(\underline{\lambda}_Q/\overline{\lambda}_P)V$$

Then one can conclude that

$$V(t) \leq V(t_0)\exp\left\{-(\underline{\lambda}_Q/\overline{\lambda}_P)(t-t_0)\right\}$$

which is an exponential decaying bound for the system. Also, we can conclude that

$$\left(\|x(t)\|/\|x_0\|\right)^2 \leq (\overline{\lambda}_P/\underline{\lambda}_P)\exp\left\{-(\underline{\lambda}_Q/\overline{\lambda}_P)(t-t_0)\right\}.$$

Finally, by rearranging the terms, in order to achieve $\|x(t_f)\| \leq r$, $t_f - t_0$ should be greater than the expression provided in the theorem. This ends the proof. □

IV. Detected Attack

If the attack is detected, it is natural to expect that one possible reaction of the subsystems would be to form a cooperative team and adopt a collective team strategy to counter the actions of the attacked subsystem. The team strategy will be selected from the noninferior set of their objective functions and will be designed to achieve a Nash equilibrium with the objective function of the attacked subsystem [17]. The noninferior set will be determined by forming a team objective function as a convex combination of the individual objective functions of the $(M-1)$ subsystems in the team. That is,

$$J_T = \sum_{i=1}^{M-1}\alpha_i J_{i+1} \tag{12}$$

where $\alpha_i > 0$, $\sum_{i=1}^{M-1}\alpha_i = 1$ and J_i for $i = 2, \cdots, M$ as they are defined in (2). The weight α_i can be viewed as the relative contribution of subsystem $(i+1)$ to the team. So the cost function for the team of subsystems becomes

$$J_T = \frac{1}{2}x^{\intercal}(t_f)S_{T_f}x(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left[x^{\intercal}Q_T x + u_T^{\intercal}R_T u_T\right]dt$$

where $S_{T_f} = \sum_{i=1}^{M-1}\alpha_i S_{(i+1)f}$, $Q_T = \sum_{i=1}^{M-1}\alpha_i Q_{i+1}$ and $R_T = blkdiag\{\alpha_1 R_2,\ \alpha_2 R_3,\ \ldots,\ \alpha_{M-1}R_M\}$. The attacked subsystem's objective function stays the same as J_1. Also, the control inputs of the subsystems in the group are combined as one vector but it will be reallocated to each subsystem later at the implementation stage. Consequently, the system dynamics of (1) can be reorganized as follows

$$\dot{x} = Ax + B_1 u_1 + B_T u_T \tag{13}$$

where $u_T = [u_2^{\intercal}, u_3^{\intercal}, \cdots, u_M^{\intercal}]^{\intercal}$ is the team control vector and $B_T = [B_2, \cdots, B_M]$. The system dynamics (13) and cost functions J_T and J_1 form a 2-subsystem LQ game. The closed-loop Noninferior Nash strategies [17] are obtained as

$$u_1^* = -R_1^{-1}B_1^{\intercal}S_1 x, \quad u_T^* = -R_T^{-1}B_T^{\intercal}S_T x \tag{14}$$

where the matrices S_T and S_1 are the solutions to the following coupled differential Riccati equations

$$\dot{S}_1 + S_1 A + A^{\intercal}S_1 + Q_1 - S_1 H_1 S_1 - S_1 H_T S_T - S_T H_T S_1 = 0, \tag{15}$$

$$\dot{S}_T + S_T A + A^{\intercal}S_T + Q_T - S_T H_T S_T - S_T H_1 S_1 - S_1 H_1 S_T = 0, \tag{16}$$

where $H_T = B_T R_T^{-1} B_T^{\intercal}$ and with boundary conditions $S_1(t_f) = S_{1f}$ and $S_T(t_f) = S_{T_f}$. Note that assigning different values of α_i will yield different teams with different strengths and weaknesses, and correspondingly different Nash equilibria. Also, note that even though this problem is solved as one team of subsystems in conflict with the attacked subsystem, upon implementation of the decisions, the team decision is reallocated to each member subsystem and the SoS remains as an enterprise of M subsystems.

Upon detecting the attack, the next corollary provides conditions that guarantee the attacker cannot accomplish its goal of destabilizing the SoS if the remaining subsystems choose their relative contributions to the team appropriately.

Corollary 2.1. *Upon detecting the attack, the cooperative team can ensure that the new Nash equilibrium also converges to the origin by simply choosing the α_is such that $Q_T > Q_1$ and $S_{T_f} > S_1$.*

Proof. It follows directly by applying Theorem 2 to an attacked SoS consisting of two subsystems with costs J_1 as in (2) and J_T as in (12). □

V. Illustrative Example

To illustrate the concept of cooperative design against an attack on a subsystem of an SoS, we consider a simple three-subsystems SoS consisting of $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ that have reached a Nash equilibrium designed to converge to the origin. We consider an attack on subsystem $\mathcal{S}_1$ whose objective is to separate it from subsystems $\mathcal{S}_2$ and $\mathcal{S}_3$ and to cause it to diverge from the equilibrium by altering some parameters in its objective function. Let $x = \begin{bmatrix} x_{\mathcal{S}_1} & x_{\mathcal{S}_2} & x_{\mathcal{S}_3}\end{bmatrix}^{\intercal} \in \mathbb{R}^3$ be the SoS state vector. The parameters of the SoS based on our model are as shown below:

$$A = \begin{bmatrix} -0.04 & 0.01 & 0.01 \\ 0.01 & -0.04 & 0.01 \\ 0.01 & 0.01 & -0.04 \end{bmatrix}, B_1 = \begin{bmatrix} 0.1 \\ 0 \\ 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0.15 \\ 0.18 \\ 0.18 \end{bmatrix}, B_3 = \begin{bmatrix} 0.15 \\ 0.11 \\ 0.31 \end{bmatrix}.$$

We consider a final time $t_f = 70$ (which is relatively long for the given system dynamics). Before the attack takes place, each subsystem of $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ has plans to minimize its objective function stated respectively as follows:

$$J_1 = 10\,x_{\mathcal{S}_1}^2(t_f) + \int_0^{70}\left(10\,x_{\mathcal{S}_1}^2 + 30\,u_1^2\right)dt$$

$$J_2 = 8\,x_{\mathcal{S}_2}^2(t_f) + \int_0^{70}\left(8\,x_{\mathcal{S}_2}^2 + 100\,u_2^2\right)dt$$

$$J_3 = 5\,x_{\mathcal{S}_3}^2(t_f) + \int_0^{70}\left(5\,x_{\mathcal{S}_3}^2 + 120\,u_3^2\right)dt$$

This represents the situation that each subsystem is only considering its own state variable in its objective function. The initial conditions at $t = 0$ are $x_0 = \begin{bmatrix} 5 & 7 & 10 \end{bmatrix}^{\intercal}$. If the

SoS has not been attacked, the planned behavior of the state variables is shown in Fig. 2a. Clearly, all state variables have planned to converge toward the origin as t_f approaches 70.

Now let us assume that subsystem $\mathcal{S}_1$ is subjected to an attack at $t = 5$. The attacker would alter the objective function J_1 so that now it looks as follows:

$$J_1 = -10\, x^2_{\mathcal{S}_1}(t_f) + \int_5^{70} \big(-10\, x^2_{\mathcal{S}_1} + 30\, u_1^2\big) dt$$

Such an attack can be accomplished by the attacker making two sign changes at $t = 5$ in the software code of the RES unit. After the attack takes place, we consider two scenarios which correspond to whether the attack is or is not detected.

A. *Undetected Attack*

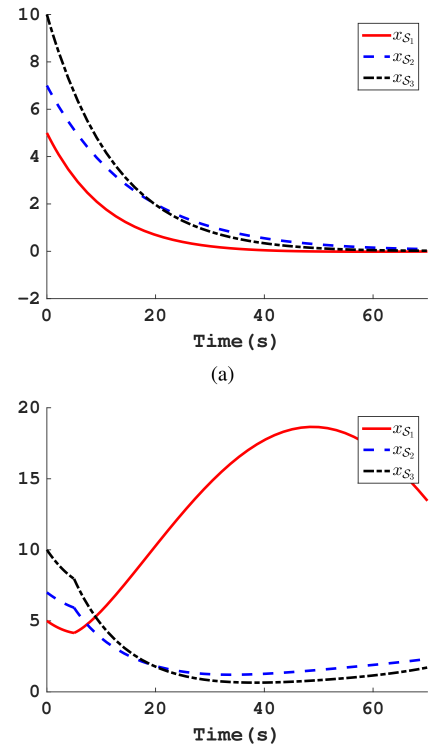
If the attack is not detected, the subsystems will accomplish a Nash equilibrium with their objective functions having the parameters as described above. The state variables of the subsystems $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ are plotted versus time in Fig. 2b. As can be seen, the attacker has successfully destroyed the convergence towards the origin that could have been established among the subsystems if the attack had not occurred. Specifically, it can be seen that even though subsystems $\mathcal{S}_2$ and $\mathcal{S}_3$ are trying to recover the convergence, due to the adversarial behavior of the attacked subsystem $\mathcal{S}_1$, the planned convergence is completely destroyed.

B. *Detected Attack*

If the attack is detected, we assume that subsystems $\mathcal{S}_2$ and $\mathcal{S}_3$ form a team according to (12). In this example we will consider different teaming options according to different choices of α_1 and α_2, which represent contributions of $\mathcal{S}_2$ and $\mathcal{S}_3$ to the team, respectively. These choices are done only for illustration purposes and may not necessarily be the best choices that optimize the contribution of each subsystem to the team. In this case, the subsystems will accomplish a Non-inferior Nash equilibrium with their team objective function against the objective function of the attacked subsystem $\mathcal{S}_1$.

The state variables of the subsystems $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ for different choices of α_1 and α_2 are plotted versus time in Fig. 3. While not all of these choices have been able to bring the trajectories back into full conformity, these options produced outcomes that are much more favorable to the entire SoS than the outcome of Fig. 2b which corresponds to the case that the attack has not been detected. In particular, Fig. 3h which shows the teaming with $(\alpha_1, \alpha_2) = (0.8, 0.2)$ which represents 80% contribution by subsystem $\mathcal{S}_2$ and 20% by $\mathcal{S}_3$, illustrates a design which gives the subsystems the ability to successfully defend against the attack and bring the entire SoS back into conformity so that its trajectories are close to the original un-attacked trajectories.

The advantage of teaming choice $(\alpha_1, \alpha_2) = (0.8, 0.2)$ is due to the fact that subsystem $\mathcal{S}_2$ has relatively more power than $\mathcal{S}_3$ in terms of the parameters in its objective function, so it may as well be required to contribute more to the team.



(a)

(b)

Fig. 2. State variables of subsystems $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ (a) when the attack had not taken place; The convergence to equilibrium has easily been established. (b) when subsystem $\mathcal{S}_1$ is attacked at $t = 5$ but the attack is not detected; The planned convergence has been destroyed.

VI. Conclusion

The current literature on Systems of Systems has emphasized a structure in which all subsystems operate harmoniously for the benefit of the entire enterprise. This structure does not consider situations when one or more subsystems are subjected to an attack designed to alter their behavior so they become adversarial towards the remaining subsystems. In this paper, we considered linear quadratic Systems of Systems where one of the subsystems is cyber-attacked by modifying parameters in its objective function so that its resulting behavior becomes hostile to the remaining subsystems. We considered two possible cases. The first is when the attack is not detected and the Riccati Equation Solver (RES) unit continues to generate solutions based on the original objective functions including the attacked subsystems modified objective function. The second is when the attack is detected and the remaining subsystems decide to take defensive action by forming a team to counter the hostile behavior of the attacked subsystem. Optimizing the formation of the team by determining the most effective contribution of each member is still a problem under investigation. As an illustrative example, we simulated a system consisting of three subsystems operating with similar objectives when one subsystem is attacked and its objective is modified so that it became adversarial with the other two. We derived the corresponding system's behavior by examining its subsystems state trajectories in both cases when the attack is not detected and when it is detected. We showed that if

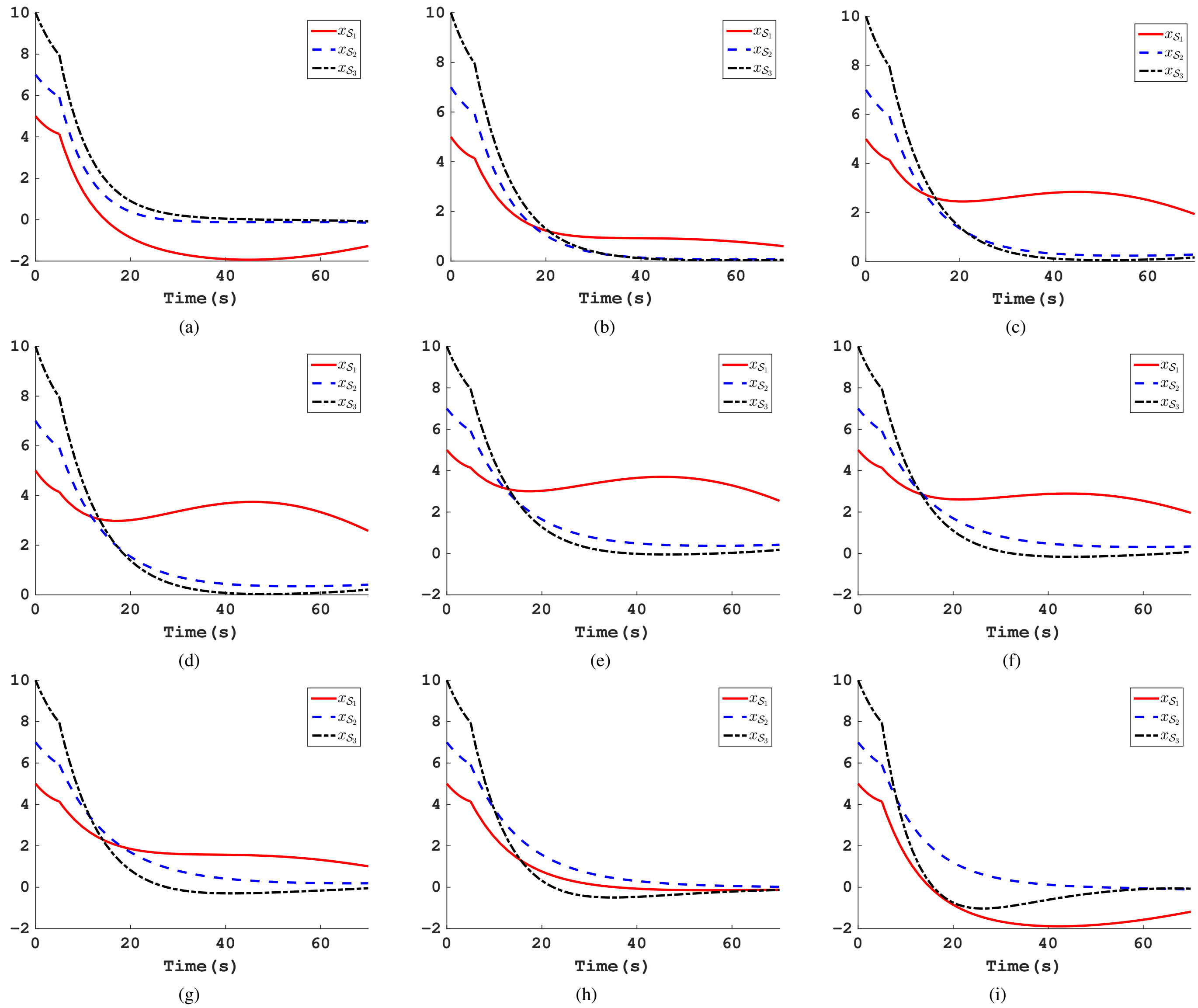


Fig. 3. State variables of subsystems $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ after subsystem $\mathcal{S}_1$ is attacked and it is detected; Subsystems $\mathcal{S}_2$ and $\mathcal{S}_3$ form a team with $(\alpha_1,\ \alpha_2)$ equals (a) $(0.1,\ 0.9)$, (b) $(0.2,\ 0.8)$, (c) $(0.3,\ 0.7)$, (d) $(0.4,\ 0.6)$, (e) $(0.5,\ 0.5)$, (f) $(0.6,\ 0.4)$, (g) $(0.7,\ 0.3)$, and (h) $(0.8,\ 0.2)$, and (i) $(0.9,\ 0.1)$.

the attack is detected, there is an advantage for the remaining subsystems to form a collaborative team.

References

[1] J. Nash, "Non-cooperative games," *Annals of math.*, pp. 286–295, 1951.
[2] R. Isaacs, *Differential Games*. John Wiley and Sons, 1965.
[3] Y. Ho, A. Bryson, and S. Baron, "Differential games and optimal pursuit-evasion strategies," *IEEE Trans. on AC*, vol. 10, pp. 385–389, 1965.
[4] A. W. Starr and Y.-C. Ho, "Nonzero-sum differential games," *Journal of Optimization Theory and Applications*, vol. 3, pp. 184–206, 1969.
[5] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Society for Industrial and Applied Mathematics, 1999.
[6] J. Engwerda, *LQ Dynamic Optimization & Differential Games*. John Wiley & Sons, 2005.
[7] L. Lambertini, *Differential Games in Industrial Economics*. Cambridge University Press, 2018.
[8] S. Talebi, M. A. Simaan, and Z. Qu, "Cooperative, non-cooperative and greedy pursuers strategies in multi-player pursuit-evasion games," in *Proc. of the IEEE Conf. on CCTA*, Aug 27-30 2017, pp. 2049–2056.
[9] ——, "All-against-one linear-quadratic differential games," *arXiv preprint arXiv:1902.10292*, 2019.
[10] ——, "Decision-making in complex dynamical systems of systems with one opposing subsystem," in *Proc. of European Control Conference*, June 25-28 2019, pp. 2789 – 2795.
[11] S. Backhaus and et al., "Cyber-physical security: A game theory model of humans interacting over control systems," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2320–2327, 2013.
[12] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyber-physical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, p. 4665, 2015.
[13] A. Gusrialdi, Z. Qu, and M. A. Simaan, "Competitive interaction design of cooperative systems against attacks," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3159–3166, 2018.
[14] ——, "Game theoretical designs of resilient cooperative systems," in *Proc. of European Control Conference*, July 2015, pp. 1699–1705.
[15] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank, *Matrix Riccati equations in control and systems theory*. Birkhäuser, 2003.
[16] T. Kato, *Perturbation theory for linear operators*. Springer Science & Business Media, 2013, vol. 132.
[17] Y. Liu and M. Simaan, "Noninferior Nash strategies for multi-team systems," *Journal of Optimization Theory and Applications*, vol. 120, no. 1, pp. 29–51, 2004.