JDMS

Special Issue: Modeling & Simulation for Cyber Security of Autonomous Vehicle Systems

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
1–16

DOI: 10.1177/1548512915594703
dms.sagepub.com
SAGE

Optimal defense and control of dynamic systems modeled as cyber-physical systems

Haifeng Niu and S Jagannathan

Abstract

With the increasing connectivity among computational cyber-connected elements and physical entities, a unified representation that captures the interrelationship between the cyber and the physical systems becomes increasingly important. In this paper, we propose a novel representation for developing cybersecurity schemes for physical systems wherein the cyber system states affect the physical system and vice versa. Subsequently by using this representation, an optimal strategy via Q-learning is derived for the cyber defense in the presence of an attack. Since the cyber system under attack will affect the physical system stability and performance, an optimal controller by using Q-learning is considered for the physical system with uncertain dynamics. As an example, cyber-attacks that increase the network delay and packet losses are considered and the goal of the proposed cyber defense and optimal controller is to thwart the attack and mitigate the performance degradation of the physical system due to increased delays and packet losses. An illustrative example is given where the proposed theory is evaluated on the yaw-channel control of an unmanned aerial vehicle. Simulation results show that on the cyber side, both the attacker and the defender gains their greatest payoff whereas on the physical system side, the optimal controller is able to maintain the linear system in a stable manner when the cyber state vector meets a certain desired criterion.

Keywords

Cyber-physical systems, cybersecurity, optimal control, zero-sum game

1. Introduction

Cyber-physical systems (CPS) refer to engineered systems constructed as networked interactions of physical and computational cyber components.[1] Examples of CPS can be found in areas as diverse as automobiles, air transportation, civil infrastructure, power grid, embedded medical devices, and consumer appliances. Recently, with the development of information technology (IT) such as IT management and networking growth, the security in CPS has received attention. Moreover, as cyber and physical capabilities are becoming increasingly intertwined, a comprehensive framework that models the cyber system, the physical plant dynamics, and their interrelationship is also increasingly needed.

In general, there are two types of the representations for the security analysis of CPS in the existing literature: one that models the effect on the cyber systems under a certain specific attack;[2–6] and the other includes the effect of cyber-attacks on physical systems.[7–12] The former effort explores the behavior of the attacker as well as the defender, formulates the cyber changes under attacks, and presents appropriate strategies that bring the cyber system back to normal. For example, the study by Baumman and Sandmann introduces denial of service (DoS) flooding attacks by a continuous-time Markov chain and utilizes

Electrical and Computer Engineering Department, Missouri University of Science and Technology, Rolla, MO, USA

Corresponding author:
Haifeng Niu, Electrical and Computer Engineering Department, Missouri University of Science and Technology, Rolla, MO, 65401, USA.
Email: hnhy6@mst.edu

the state space method to compute security measures accurately.[2]

Different from Baumman and Sandmann,[2] Zhu and Basar studied the cyber defense by modeling the actions of the attacker and the defender as a stochastic zero-sum game.[3] In the study by Ten et al.,[4] the measure of vulnerabilities in cyber-physical systems with application to power systems is defined and a security framework including anomaly detection and mitigation strategies is provided. Sallhammar et al. evaluated cybersecurity by computing the expected probabilities of the attacker and using the probabilities to build a transition model through a game-theoretic approach.[5] In the investigation by Aenes et al.,[6] the cyber vulnerability was dynamically evaluated by using a hidden Markov model which provides a mechanism for handling sensor data with different trustworthiness. However, this type of representation mainly focuses on the cyber system and neglects the fact that the states of the physical system also affect the cyber defense strategy.

In contrast, others concentrate on characterizing the dynamics of the physical system under attacks by extending the classic state-space description in order to include the attacks.[7–12] For instance, in the report by Kwon et al.,[7] the system dynamics include an extra term to model the deception attack. In the study by Liu et al.,[8] the system state under attack is represented with an additive term, where the additive term is used to simulate the false data injection attack. Unlike Liu et al.,[8] Teixeira et al. characterize the deception attackers by a set of objectives and propose policies to synthesize stealthy deceptions attacks in both linear and nonlinear estimators.[9] In the investigation by Fawzi et al.,[10] the estimation and control of linear systems when sensors or actuators are corrupted by an attacker is provided, together with a secure local control loop that can improve the resilience of the system. On the other hand, Amin et al. define the control input under attacks as the product of the given input and a coefficient to characterize the effect induced by the DoS attacks.[11] A class of human adversaries, who are called correlated jammers, was considered by Zhu and Martínez.[12] By modeling the coupled decision making process as a two-level receding-horizon dynamic Stackelberg game, the authors propose a control law and analyze the performance and the closed-loop stability under attacks.

However, there are many weaknesses in the above reported works.[13] First, the representation can only describe a single type of attack due to the fact that attacks affect the system dynamics in a variety of ways. In particular, Pasqualetti et al. proposed a unified framework that is able to detect attacks;[13] however, it still has the two drawbacks mentioned next. Second, it is difficult to implement the representation developed in the literature so far since the system dynamics under attacks are considered known. For instance, due to random delays and packet losses caused by certain cyber-attacks, the physical system dynamics can be uncertain. Last but not the least, these representations fail to take the interactions between the cyber defense policy and the system controller under consideration.

In summary, to the best knowledge of the authors, little effort has been carried out in the literature to develop a representation that precisely characterizes the interplay between the cyber and the physical systems. Such a representation is necessary because inadequate decisions can be made for the cyber defense if the physical states are ignored. Likewise, the physical plant may not be stable if the controller is designed without considering the impact due to the changes in the cyber system.

Therefore, in this paper, we propose a framework for cyber-physical systems to (i) study optimal defense to mitigate attacks and (ii) to derive an associated optimal control policy for physical systems. First, we introduce a mathematical representation for the cyber-physical system, in which it was shown that the activities of the cyber system affect the states of the physical system and vice versa. Then based on this representation, we derive the optimal strategies for the defender and the attacker by considering them as two players in a zero-sum game. Since the cyber state influences the behavior of the physical system, next, an optimal controller for the physical system in the presence of uncertainties induced by the cyber system is revisited, based on that reported by Xu et al.[14] In addition, a condition on the cyber state vector is derived under which the physical system is stable. Finally, an illustrative example is given in which we show that on the cyber side, both the attacker and the defender gain their greatest payoff while on the physical side, the optimal controller is able to maintain the plant stable when the state vector of the cyber system meets a certain condition.

Thus, the main contributions of this work include: (i) a novel and comprehensive representation of the cyber-physical system that captures the interrelationship between the cyber and the physical elements; (ii) the development of the optimal strategies for the defender and the attacker; (iii) the application of the optimal controller for the physical system in the presence of uncertain dynamics induced by the cyber system;[14] and (iv) the demonstration of how the proposed theory can be applied to the control of the yaw-channel of an unmanned aerial vehicle (UAV) in the presence of an attack.

The rest of this paper is organized as follows. The proposed representation for the cyber-physical systems is introduced in Section 2. In Section 3, the optimal defense and attack policies are derived and presented, followed by the optimal controller design for the physical system introduced in Section 4. The illustrative example including policy derivation as well as the simulation results are

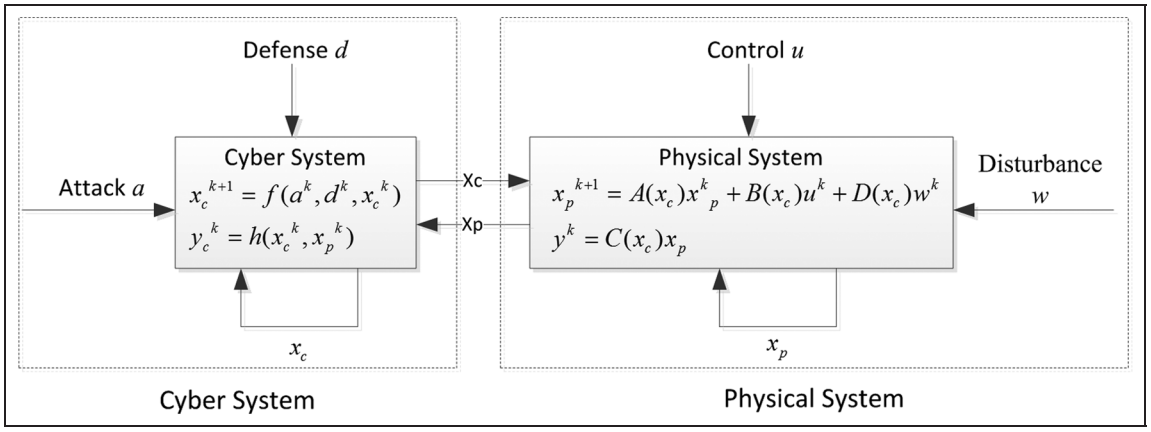

Figure 1. Proposed representation of a cyber-physical system.

presented in Section 5 and this paper is concluded in Section 6.

2. Proposed representation for cyber-physical systems

In this section, the proposed framework for the cyber-physical systems is introduced. Figure 1 depicts the proposed representation for the optimal defense/control scheme.

2.1. Cyber system

Consider the cyber system described by a nonlinear discrete-time system given by

$$x_c(k+1) = f(a(k), d(k), x_c(k)), \tag{1}$$

where $x_c \in \mathbb{R}^{N_c}$ is the state of the cyber system, N_c being the dimension of the state vector of the cyber system, $a \in \mathbb{R}$ is malicious action taken by the attacker, and $d \in \mathbb{R}$ is the defense strategy taken by the cyber system.

The cyber state x_c represents a set of network performance metrics such as latency, throughput, packet loss rate, and so on. Since it was shown in the literature that most attacks on the cyber system will cause an increase in network delay and packet losses,[15] in this paper we mainly consider these two as the cyber state vector in the controller design (Section 4) and in the illustrative example (Section 5). In some cases, x_c also needs to include a few network security metrics such as the number of successive failed authentications or the changes of IP addresses. It is obvious that the cyber state can be affected by the action of both the attacker and the defense strategy and a relationship is described by the function f.

In particular, we propose a more concrete representation of the cyber state as

$$x_c(k+1) = A_c(k)F_c(x_c(k))D_c(k) = \sum_{i=0}^{N_a}\sum_{j=0}^{N_d} a_i d_j f_{ij}(x_c(k)), \tag{2}$$

where $A_c = [a_0, a_1, ..., a_{N_a}]$ is a vector consisting of all N_a number of possible attacks, each $a_i \in \{0, 1\}$ stands for a type of attack (except for a_0) wherein $a_i = 1$ implies the ith attack has been launched and $a_i = 0$ otherwise. In particular, we let $a_0 = 1$ if and only if there is no active attack at that moment. Similarly, $D_c = [d_0, d_1, ..., d_{N_d}]^T$ is a vector describing the status of the defense strategies and $d_0 = 1$ if and only if there is no active defense. Finally, $F = [f_{00}, f_{01}, ..., f_{0N_d}; ...; f_{N_a0}, f_{N_a1}, ..., f_{N_aN_d}]$ is a matrix of functions and each element $f_{ij} : \mathbb{R}^{N_c \times 1} \to \mathbb{R}^{N_c \times 1}$ describes the effect to the cyber state x_c brought by the ongoing attack/defense pair (a_i, d_j). In other words, at each sampling time instant k, the active attack/defense pair (a_i, d_j) corresponds to a function f_{ij} which characterizes the system dynamics for the following sampling interval. An assumption is made in that when there are two or more attacks (and defense) simultaneously being launched, the effect of each attack (and defense) to the cyber system state is independent.

As depicted in Figure 1, the cyber system output in the proposed representation is described as

$$y_c(k) = h\big(x_c(k), x_p(k)\big), \tag{3}$$

where $y_c \in \mathbb{R}$ is the output of the cyber system and $x_p \in \mathbb{R}^{N_p}$ is the state of the physical system with N_p being the dimension of the state vector. The output y_c, which is a function of x_c and x_p, is a quantized value indicating the condition of the cyber system. A simple example of y_c is presented in Remark 1 whereas more complicated forms can be found in Remark 2.

One can assess the health condition or even the specific attack on the system by exploiting the cyber state x_c as well as the physical system state x_p. For example, if the network is reported with a significant drop in throughput and a considerable mean delay in a short time, then it is possible that the system is experiencing a DoS attack. The importance of introducing the cyber output y_c stems from the fact that the states needs to be organized and interpreted in order to be useful for the administrator to make suitable defense strategies.

It is important to note that the physical system state x_p is also necessary at the cyber system in order to obtain a comprehensive and accurate estimation of the system condition. For example, if an attacker manages to get the administrative privilege without being detected by cracking the password or exploiting the security bugs, then he/she is able to give malicious instructions that may lead to the failure of the physical system. In this case, only the abnormality in the physical system state (not the cyber state) could be detected. Therefore, by including the physical system state when assessing the condition of the cyber system, the administrator can still trigger the alert mechanism and launch the defense even if no abnormities in the cyber systems have been observed. Therefore, by using both x_c and x_p in y_c, the cyber defense decision becomes more insightful and reliable. The relationship between y_c, x_c, and x_p is characterized by the function h.

Remark 1: A simple example of the cyber output y_c is presented here, in which y_c is defined as

$$y_c = \begin{cases} 1, & \text{if } x_c \in X_{cd} \text{ and } x_d \in X_{dd} \\ 0, & \text{otherwise} \end{cases},$$

where X_{cd} and X_{dd} are the set of desired values of the cyber state x_c and physical state x_p, respectively. Therefore, in this example, $y_c = 1$ represents a healthy system while $y_c = 0$ represents a compromised one.

Remark 2: The function h may take various forms on the basis of the system security requirement. The selection of h is critical to the system security level, considering that the output of h is used to assess the system health condition and determine the defense strategies that will be launched. The objective of selecting function h is that it should make use of the observed states and precisely predict the ongoing or even potential attacks. A few examples of function h can be given as follows:

1. Threshold form: $y_c = \left(\text{sgn}\left(\begin{bmatrix} x_c - x_{c\_\min} \\ x_p - x_{p\_\min} \end{bmatrix}\right) + \text{sgn}\left(\begin{bmatrix} x_c - x_{c\_\max} \\ x_p - x_{p\_\max} \end{bmatrix}\right)\right)/2$, where $y_c \in \mathbb{R}^{(N_c+N_p)\times 1}$, $x_{c\_\min}, x_{c\_\max}(x_{p\_\min}, x_{p\_\max})$ are the predefined lower, upper threshold vectors for each cyber (physical) state respectively and $\text{sgn}(\cdot)$ is the sign function. As a result, the corresponding row of y_c becomes "-1" if a state is smaller than the lower limit, "0" if within the interval, and "1" if higher than the upper limit. This form of function h provides a straightforward assessment of whether the states are in the desired zone or not.
2. Linear form: $y_c(k) = \eta_c x_c(k) + \eta_p x_p(k)$ where $y_c \in \mathbb{R}$; $\eta_c \in \mathbb{R}^{1\times N_c}$ and $\eta_p \in \mathbb{R}^{1\times N_p}$ denote the coefficient vectors for each state. By making use of these weighting factors, this form maps the state vector onto a scalar that provides an approximate description of the system healthy condition.
3. Quadratic form: $y_c(k) = x_c^T(k)\Lambda_c x_c(k) + x_p^T(k)\Lambda_p x_p(k)$, where $y_c \in \mathbb{R}, \Lambda_c \in \mathbb{R}^{N_c\times N_c}$ and $\Lambda_p \in \mathbb{R}^{N_p\times N_p}$ represent the weighting matrices for each state. Similar to the linear form, this quadratic form also maps the state vector onto a scalar except that it takes the correlation between each state into consideration.

In this paper, the attacks considered will increase the network delay and packet losses which in turn will make the linear time-invariant system as an uncertain stochastic time-varying system. The goal of the cyber defense and optimal controller is to mitigate the increase in random delays and packet losses and performance degradation of the physical system.

2.2. Physical system

As shown in the right block in Figure 1, the physical system is described as a linear discrete system in the presence of a disturbance given by

$$\begin{aligned} x(k+1) &= A(x_c)x(k) + B(x_c)u(k) + D(x_c)w(k) \\ y^k(k) &= Cx(k) \end{aligned}, \qquad (4)$$

where $x_p \in \mathbb{R}^{n_p}$ is the state of the physical system, $u \in \mathbb{R}^{m_u}$ is the control input, $w \in \mathbb{R}^{m_w}$ is the disturbance input, $y \in \mathbb{R}^r$ is the output, and $A \in \mathbb{R}^{n_p\times n_p}$, $B \in \mathbb{R}^{n_p\times m_u}$, $C \in \mathbb{R}^{r\times n_p}$, and $D \in \mathbb{R}^{n_p\times m_w}$ denote the system matrices.

It is important to note that unlike the classical linear discrete system, the system matrices described by equation (4) are a function of the cyber state x_c. In other words, the state of the cyber system will influence the dynamics of the physical system. For instance, a large network-induced delay or packet loss can degrade the system performance or even results in instability. Therefore, this framework is able to capture the cyber system activities because when a cyber-attack occurs, the physical system matrices $\{A(x_c), B(x_c)\ldots\}$change.

In conclusion, the cyber state vector, whose update is subject to the attack/defense decisions, changes the

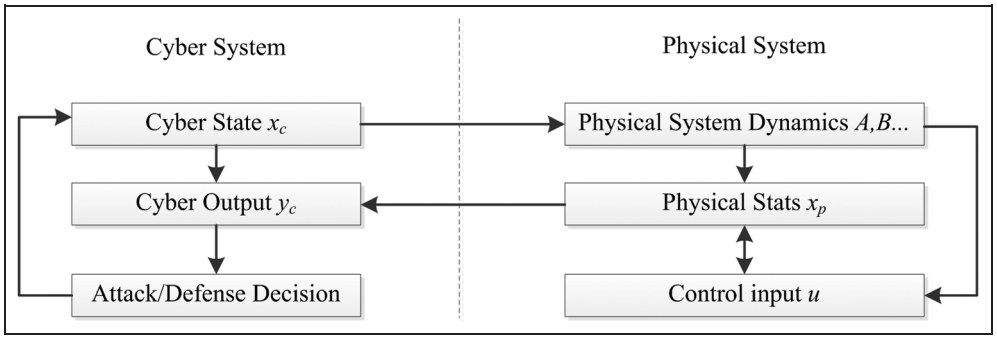

Figure 2. Interrelationship between the cyber and the physical system.

physical system dynamics. As a result, the control input needs to be adjusted to drive the physical states back to the desired value. The changes in the cyber and physical states, in turn determine the cyber output and hence the attack/defense decisions. A summary of the interrelationship between the cyber and the physical systems is shown in Figure 2.

Hence, the objective is to design an optimal policy by using a cost function for the physical system with unknown system dynamics induced by the cyber system. Therefore, by (i) including the physical system state in the assessment of cyber health condition and (ii) considering the influence on the physical system dynamics induced by the cyber states when designing the optimal controller, the proposed optimal defense/control scheme offers a coupled design which is able to capture the influence of the cyber and the physical systems.

3. Optimal attack/defense policy for cyber systems

In this section, the optimal attack and defense policies for the cyber system are derived, while in the next section we derive the optimal controller for the physical system with the presence of the delay and packet loss. We also derive the condition for the delay and packet loss under which the physical controller can be stabilized. The optimal controller gain will be computed and applied to the physical system once the delay and packet loss satisfy the condition. Otherwise appropriate defense strategy needs to be launched in order to drive the cyber states (delay and packet loss) to meet the criterion.

In this section, we first model the interactions between the attacker and the defender as a two-player zero-sum Markov game.[16] Then after defining the instant payoff as well as the expected discounted payoff function, we introduce two lemmas to show the existence of the solution of the game and the optimal policy. Next, the Q-function is proposed and it is shown in Theorem 1 that, using the Minimax-Q algorithm,[17] the Q-function converges to the game value. As a result, the optimal strategies for the defender and the attacker in order to gain their greatest discounted payoff are also derived.

Consider the cyber system with dynamics described by equation (2) and an output function in quadratic form of the state vectors, i.e. as

$$x_c(k+1) = A_c(k)F_c(x_c(k))D_c(k) = \sum_{i=0}^{N_a}\sum_{j=0}^{N_d} a_i d_j f_{ij}(x_c(k))$$
$$y_c(k) = x_c^T(k)\Lambda_c x_c(k) \tag{5}$$

where the cyber state vector x_c consists of delay and packet loss for illustrative purpose. Then the system can be modeled as a Markov decision process in which the state at the next sampling interval, $x_c(k+1)$, is determined by the state at the current instant, $x_c(k)$, together with the action pair $(A_c(k), D_c(k))$ launched by the defender and the attacker. The defender and the attacker update their defense strategies based on the condition indicated by y_c, which is a quantified value computed based on the delay and packet loss of the cyber system. In other words, the defender and the attacker launch appropriate actions so as to drive the delay and packet loss into preferred values.

Let Y be the set of all possible values of y_c. Since it is based on the value of y_c that the defender and the attacker decide which action should be taken, the problem becomes deriving the optimal action for each single value of y_c, which is impractical and unnecessary due to the tremendous computation. Therefore, we divide Y into several subsets and study the optimal strategies for each subset rather than for each element. Suppose that Y is divided into N_{yd} disjoint subsets (i.e. $Y = Y_1 \cup Y_2 \cup ... \cup Y_{N_{yd}}$ and $Y_i \cap Y_j = \varnothing$ for $i \neq j$) and each subset corresponds to a level of health status.

As illustrated in Figure 3, Y is divided into eight subsets where subset Y_0 is the secure state (with the smallest delay and packet loss) and subset Y_8 is the failed state of the system (with the largest delay and packet loss). The defender

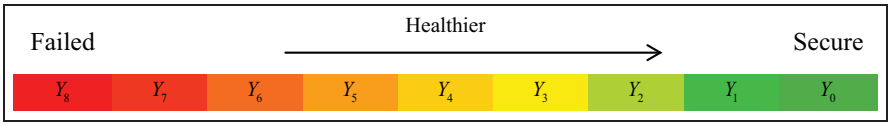

Figure 3. Each subset of Y corresponds a level of health condition.

decides which action should be taken based on the subset that current y_c is in. For example, if $y_c \in Y_4$, the defender may choose to load the defense more frequently to drive y_c into a more secure subset. As a result, the delay and packet loss are reduced and the physical system becomes more robust and resilient. Obviously, the more subsets Y is divided into, the more accurate the model is. However, more computation is needed as the optimal strategies need to be derived for each subset. Next, the definition of instant reward and discounted payoff are introduced in order to obtain the optimal strategy for each subset Y_i.

Let $r(A_c(k), D_c(k), Y_i(k))$ be the instant payoff (reward or cost) at time instant k in region $Y_i(k)$ for the action pair $(A_c(k), D_c(k))$. Let the instant payoff of the attack and the defender be r_a and r_d, respectively, and assume the game is zero-sum; we then have the relationship

$$\begin{aligned} &r(A_c(k), D_c(k), Y_i(k) := r_a(A_c(k), D_c(k), \\ &Y_i(k)) = -r_d(A_c(k), D_c(k), Y_i(k)), \end{aligned} \tag{6}$$

Specifically, we let the instant reward be defined as

$$\begin{aligned} &r(A_c(k), D_c(k), Y_i(k)) \\ &= x_c^T(k)\Lambda_c x_c(k) + x_p^T(k)\Lambda_p x_p(k) + \xi_d D_c(k) - \xi_a A_c^T(k), \end{aligned} \tag{7}$$

which consists of the cost of the cyber state, physical state, defense, and attack. The defense cost is defined as $\xi_d D_c(i)$, where $\xi_d = [\xi_{d,1}, \xi_{d,2}, ..., \xi_{d,N_d}]$ and each element $\xi_{d,i} \in \mathbb{R}^+$ is the corresponding cost of launching defense d_i. Likewise, $\xi_a = [\xi_{a,1}, \xi_{a,2}, ..., \xi_{a,N_a}]$ is the vector describing the cost of launching attacks. Next, we will derive the optimal strategy for the attacker and the optimal defense can be obtained in the same manner.

After introducing the definition of the instant payoff, we now consider the expected discounted payoff function over multiple stages. Let $\Xi_A = \{A_c(1), A_c(2), ...A_c(k)...\}$ and $\Xi_D = \{D_c(1), D_c(2), ...D_c(k)...\}$ be the *policies* for the attack and defense, respectively, where $A_c(k)$ and $D_c(k)$ stand for the actions at the time instant k. A policy, which is a sequence of decisions over time, is the mathematical description of a plan of the player for the game.[18] Now define the expected discounted cost function V for each subset Y_i as

$$V(\Xi_A, \Xi_D, Y_i) = \sum_{k=0}^{\infty} \left[\beta^k E(r(k)|\Xi_A, \Xi_D, y_c \in Y)_i\right], \tag{8}$$

where $\beta \in [0, 1)$ is the discount factor. As a result, the objective of the attacker becomes finding the appropriate policy Ξ_A in each subset Y_i such that the expected discounted payoff function V is maximized. Correspondingly, the defender aims to find the appropriate defense policy Ξ_D for each Y_i to minimize V. That is to say, we need to solve $\Xi_A = \arg\max_{\Xi'_A} V_a(\Xi'_A)$ and $\Xi_D = \arg\max_{\Xi' D} V_d(\Xi'_D)$. Next, the following two lemmas are introduced before we derive the optimal policies.

Lemma 1. The discounted zero-sum game always possesses a unique solution yielding the optimal game value.[19]

Lemma 2. The policy (Ξ_A^*, Ξ_D^*) is guaranteed to be optimal if $V(\Xi_A^*, \Xi_D^*, Y_i)$ satisfies the following fixed-point Bellman equation given by[20]

$$\begin{aligned} &V(\Xi_A^*, \Xi_D^*, Y_i) = \min_{\Xi_D} \max_{\Xi_A} \\ &\left\{ r(A_c, D_c, Y_i) + \beta \sum_{Y_i'} p(Y_i'|Y_i, A_c, D_c) V(\Xi_A^*, \Xi_D^*, Y_i') \right\}, \end{aligned} \tag{9}$$

where p is the probability of transitioning from current state Y_i to the next state Y_i' after taking action pair (A_c, D_c).

Based on these two lemmas, we use an iterative Q-learning method to search for the game value $V(\Xi_A^*, \Xi_D^*, Y_i)$ in equation (9). Now define the Q-function for each region Y_i as

$$\begin{aligned} &Q(A_c, D_c, Y_i) = r(A_c, D_c, Y_i) \\ &+ \beta \sum_{Y_i^i \in Y} p(Y_i'|Y_i, A_c, D_c) V(\Xi_A, \Xi_D, Y_i'). \end{aligned} \tag{10}$$

Accordingly, the optimal action dependent value function Q^* of the game is defined as

$$\begin{aligned} &Q^*(A_c, D_c, Y_i) = r(A_c, D_c, Y_i) \\ &+ \beta \sum_{Y_i^i \in Y} p(Y_i'|Y_i, A_c, D_c) V(\Xi_A H^*, \Xi_D^*, Y_i'). \end{aligned} \tag{11}$$

From equations (9) to (11), one can conclude that if the action pair sequence (Ξ_A, Ξ_D) is optimal, the optimal Q-function $Q^*(A_c, D_c, Y_i)$ is equal to the game value function $V(\Xi_A^*, \Xi_D^*, Y_i)$. In other words, we have

$$V(\Xi_A^*, \Xi_D^*, Y_i) = \min_{\Xi_D} \max_{\Xi_A} Q^*(A_c, D_c, Y_i) = Q^*(A_c^*, D_c^*, Y_i). \quad (12)$$

The Minimax-Q algorithm proposed by Littman is adopted to obtain $Q^*(A_c, D_c, Y_i)$ since it provides strong convergence guarantees according to the following theorem.[17]

Theorem 1. Let the Q-function $Q(A_c, D_c, Y_i)$ and the optimal action dependent value function, $Q^*(A_c, D_c, Y_i)$, be defined as in equations (10) and (11), respectively. Then $Q(A_c, D_c, Y_i)$ converges to the optimal value $Q^*(A_c, D_c, Y_i)$ after an infinite number of iterations with the following tuning law given by

$$Q_{i+1}(A_c, D_c, Y_i) = (1-\alpha(i))Q_i(A_c, D_c, Y_i) + \alpha(i)\left(r(A_c, D_c, Y_i) + \beta\Theta_a(Y_i')\right), \quad (13)$$

where $\alpha(i) \in \mathbb{R}^+$ is the learning rate that satisfies $\sum_{i=1}^{\infty} \alpha(i) < \infty$ and $\sum_{i=1}^{\infty} \alpha^2(i) < \infty$, and $\Theta_a(Y_i)$ is called the *state value function* calculated by[17]

$$\Theta_a(Y_i) = \min_{D_c} \sum_{A_c} Q(A_c, D_c, Y_i)\pi_a(A_c, Y_i), \quad (14)$$

where $\pi_a(A_c, Y_i)$ denotes the probability for the attacker to take action A_c given $y_c \in Y_i$.

The proof of Theorem 1 is similar to the theorem given by Littman.[21]

In addition, since $\pi_a(A_c, Y_i)$ is unknown, linear programming is employed to approximate it at each iteration. An appropriate update law for $\pi_a(A_c, Y_i)$ is given by[21]

$$\pi_a(A_c, Y_i) := \arg\max_{\pi_a(Y_i, \cdot)} \left\{ \min_{D_c} \left\{ \sum_{A_c} Q(A_c, D_c, Y_i)\pi_a(A_c, Y_i) \right\} \right\}. \quad (15)$$

A flowchart of the proposed method to obtain the optimal defense/attack strategy is shown in Figure 4.

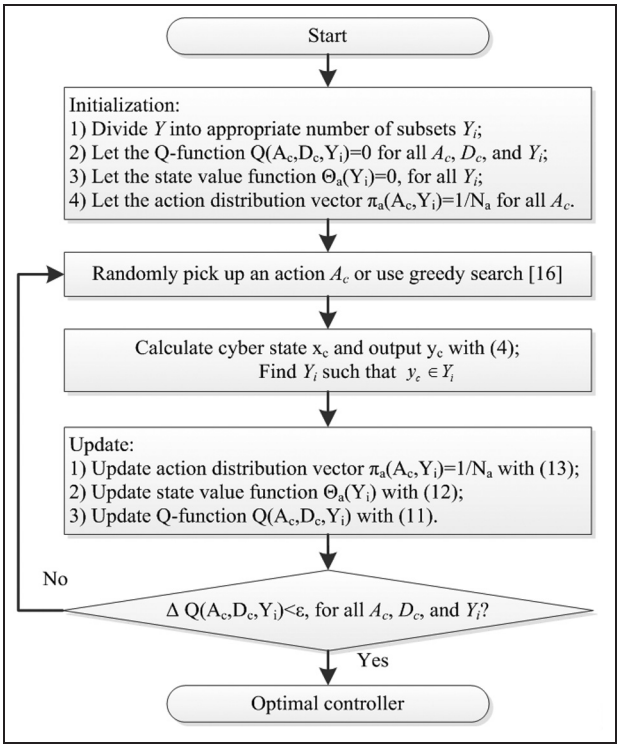

Figure 4. Flowchart of the optimal policy for the defender/attacker.

4. Optimal controller design

In this section, we introduce the optimal control scheme for the physical system based on the previous work.[14] First, we model the linear discrete-time system with dynamics that is unknown and altered by the cyber state vector, which includes packet losses and time delays since these are two important metrics for the network that may cause deterioration or potential instability of the system.[22] We then introduce the optimal control gain and show that the system is stable only when the cyber state vector satisfies a certain criterion. The cyber system needs to launch the appropriate defense if its state vector fails to satisfy the criterion. The development of the system dynamics as well as the Q-function update law is taken from the paper by Xu et al.[14] In summary, we show that the cyber state vector affects the optimal controller design and meanwhile the states of physical system also have an impact on designing the defense for the cyber system.

In cyber-physical systems, there are two types of network-induced delays: the sensor-to-controller delay and the controller-to-sensor delay. With the assumption that the former is negligible, the linear continuous system can be described as[14]

$$\dot{x}(t) = Ax(t) + \gamma(t)Bu(t-\tau(t)); \quad y(t) = Cx(t), \quad (16)$$

where

$$\gamma(t) = \begin{cases} \mathbf{I}^{n\times n} & \text{if the control input is received at time } t \\ \mathbf{0}^{n\times n} & \text{if the control input lost at time } t \end{cases}$$

and τ is the delay which is discrete-value. It is important to note that the data information needed to be discretized before transmitting into the communication network.

Moreover, to avoid the infinite-dimensional issue, authors assume that the delays are bounded. Let T_s be the sampling time, the system can be discretized as

$$x_{k+1} = A_s x_k + \sum_{i=0}^{b} \gamma_{k-i} B_i^k u_{k-i}; \quad y_k = C x_k, \tag{17}$$

where b is the maximum number of delayed control input during the sampling interval; $x_k = x(kT)$; $A_s = e^{AT}$; $B_0^k = \int_{\tau_0^k}^{T} e^{A(T-s)} ds B \cdot \mathbf{1}(T - \tau_0^k)$; $B_i^k = \int_{\tau_i^k - iT}^{\tau_{i-1}^k - (i-1)T} e^{A(T-s)} ds B \cdot \delta(T + \tau_{i-1}^k - \tau_i^k) \cdot \delta(\tau_i^k - iT)$; $D_i^k = \int_{\tau_i^k - iT}^{\tau_{i-1}^k - (i-1)T} e^{A(T-s)} ds D \cdot \delta(T + \tau_{i-1}^k - \tau_i^k) \cdot \delta(\tau_i^k - iT) \; \forall i = 1, 2, ..., b$; $\delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$; and $\gamma_{k-i} = \begin{cases} 1, & \text{if } u_{k-i} \text{ was received during } [kT_s, (k+1)T_s) \\ 0, & \text{if } u_{k-i} \text{ was lost during } [kT_s, (k+1)T_s) \end{cases}$. Let the augmented state z_k be defined as: $z_k = [x_k^T \; u_{k-1}^T \cdots u_{k-b}^T]^T$, then the system dynamics become[14]

$$z_{k+1} = A_{zk} z_k + B_{zk} u_k, \quad y_k^n = C_z z_k, \tag{18}$$

where the system matrices are a function of the unknown random delays, and packet losses or the cyber state vector which are given by[14]

$$A_{zk} = \begin{bmatrix} A_s & \gamma_{k-1} B_1^k & \cdots & \gamma_{k-i} B_i^k & \cdots & \gamma_{k-b} B_b^k \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & I_m & \cdots & \cdots & 0 & 0 \\ \vdots & 0 & I_m & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & I_m & 0 \end{bmatrix},$$

$$B_{zk} = \begin{bmatrix} \gamma_k B_0^k \\ I_m \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; C_z = \begin{bmatrix} C & & & & \\ & I_m & & & \\ & & I_m & & \\ & & & \ddots & \\ & & & & I_l \end{bmatrix},$$

and $y_k^n = [y_k^T \; u_{k-1}^T \cdots u_{k-b}^T \; w_{k-1}^T \cdots w_{k-b}^T]^T$, where I_m, I_l are $m \times m$ and $l \times l$ identity matrices. The objective is to minimize the cost function $J_k = \underset{\tau,\gamma}{E}\left(\sum_{m=k}^{\infty} (x_m^T S x_m + u_m^T R u_m) \right)$, where S and R are symmetric positive semi-definite and symmetric positive definite constant matrices respectively. Applying the augmented state vector, the cost function can be represented as $J_k = \underset{\tau,\gamma}{E}\left(\sum_{m=k}^{\infty} (z_m^T S_z z_m + u_m^T R_z u_m) \right)$, where $S_z = diag\{S, R/b, ..., R/b\}$ and $R_z = R/b$. The cost function is also known to be quadratic and is given as $J_k = \underset{\tau,\gamma}{E}(z_k^T P_k z_k)$ where $P_k \geq 0$. Now define the Q-function as

$$\begin{aligned} Q(z_k, u_k) &= \underset{\tau,\gamma}{E}(r(z_k, u_k) + J_{k+1}) \\ &= \underset{\tau,\gamma}{E}\left(\begin{bmatrix} z_k^T, u_k^T \end{bmatrix} H_k \begin{bmatrix} z_k^T, u_k^T \end{bmatrix}^T \right) = \begin{bmatrix} z_k^T, u_k^T \end{bmatrix} \underset{\tau,\gamma}{E}(H_k) \begin{bmatrix} z_k^T, u_k^T \end{bmatrix}^T. \end{aligned} \tag{19}$$

where $r(z_k, u_k) = z_m^T S_z z_m + u_m^T R_z u_m$. Therefore $\underset{\tau,\gamma}{E}(H_k)$ can be expressed in terms of the system matrices as

$$\begin{aligned} \bar{H}_k &= \underset{\tau,\gamma}{E}(H_k) = \begin{bmatrix} \bar{H}_k^{zz} & \bar{H}_k^{zu} \\ \bar{H}_k^{uz} & \bar{H}_k^{uu} \end{bmatrix} \\ &= \begin{bmatrix} S_z + \underset{\tau,\gamma}{E}(A_{zk}^T P_{k+1} A_{zk}) & \underset{\tau,\gamma}{E}(A_{zk}^T P_{k+1} B_{zk}) \\ \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} A_{zk}) & R_z + \underset{\tau,\gamma}{E}(B_{zk}^T P_{k+1} B_{zk}) \end{bmatrix}. \end{aligned} \tag{20}$$

Consequently, the optimal control gain is represented in terms of $\bar{H}_k$ as $K_k = (\bar{H}_k^{uu})^{-1} \bar{H}_k^{uz}$. Moreover, with the linear in the unknown parameters (LIP) assumption, the Q-function can be written as $Q(z_k, u_k) = w_k^T \bar{H}_k w_k = \bar{h}_k^T \bar{w}_k$, where $\bar{h}_k = vec(\bar{H}_k)$, $w_k = [z_k^T, u^T(z_k)]^T$, and $\bar{w}_k = (w_{k1}^2, ..., w_{k1} w_{kq}, w_{k2}^2, ..., w_{kq-1} w_{kq}, w_{kq}^2)$ is the Kronecker product quadratic polynomial basis vector. Therefore, the Q-function can be estimated as $\hat{Q}(z_k, u_k) = \hat{\bar{h}}_k^T \bar{w}_k$, in which $\hat{\bar{h}}$ is the estimate value of the target parameter vector $\bar{h}$.

Now define the residual or temporal difference error as $e_{hk+1} = \hat{J}_{k+1} - \hat{J}_k + r(z_k, u_k)$; then we can rewrite the residual dynamics as

$$e_{hk+1} = r(z_k, u_k) + \hat{\bar{h}}_{k+1}^T \Delta W_k \quad \text{were} \quad \Delta W_k = \bar{w}_{k+1} - \bar{w}_k. \tag{21}$$

Next, we define an auxiliary residual error vector as $\Xi_{hk} = \Gamma_{k-1} + \hat{\bar{h}}_k^T \Omega_{k-1}$ where

$$\Gamma_{k-1} = [\, r(z_{k-1}, u_{k-1}) \quad r(z_{k-2}, u_{k-2}) \quad \cdots \quad r(z_{k-1-i}, u_{k-1-j}) \,]$$
$$\text{and} \quad \Omega_{k-1} = [\, \Delta W_{k-1} \quad \Delta W_{k-2} \quad \cdots \quad \Delta W_{k-1-j} \,].$$

Similarly, the dynamics of the auxiliary vector are derived as: $\Xi_{hk+1} = \Gamma_k + \hat{\bar{h}}_{k+1}^T \Omega_k$. The update law of the target matrix $\bar{H}_k$ is given by

$$\hat{\bar{h}}_{k+1} = \Omega_k (\Omega_k^T \Omega_k)^{-1} (\alpha_h \Xi_{hk}^T - \Gamma_k^T). \tag{22}$$

It was shown by Littman that,17 with the update law of equation (22), there exists a positive constant α_h satisfying $0 < \alpha_h < 1$ such that both the state vectors z_k and the

adaptive parameter estimator errors are asymptotically stable in the mean.

Finally, we show the sufficient condition on the cyber state in term of the delay and packet loss that need to satisfy in order to maintain the system to be stochastically stable. Consider the systems with slowly-varying parameters, since the initial stabilizing control and disturbance inputs are given, the linear discrete-time system can be represented as $z_{k+1} = A^*_{zk} z_k$.[23] Applying the linear transformation, the expectation of A^*_{zk} can be written as

$$A^*_{zk} = \begin{bmatrix} A_s - \gamma_k B_0^k K & \gamma_{k-1} B_1^k & \cdots & \cdots & \gamma_{k-b} B_b^k \\ -K & 0 & \cdots & \cdots & 0 \\ 0 & I_m & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & I_m & 0 \end{bmatrix}$$

$$\Rightarrow E(A^{**}_{zk}) = \begin{bmatrix} E(A_s - \gamma_k B_0^k K) & E(\gamma_{k-b} B_b^k) & 0 & \cdots & 0 \\ -K & 0 & 0 & \cdots & 0 \\ 0 & 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & I_m \end{bmatrix}.$$

According to the definition of stability for stochastic linear time-varying system,[24] if eigenvalues of $E(A^{**}_{zk})$ are within a unit radius n-dimensional sphere (or disc) for all instants, then the system is stable. Since the eigenvalues of the right bottom block of $E(A^{**}_{zk})$ are ones, the left upper block has to satisfy the condition $\lambda_i[\underset{\tau,\gamma}{E}(A_s - \gamma_k B_0^k K)] < 1$ for any i and k, and $\lambda(M)$ denotes the eigenvalue of the matrix M. Since K and L are the initial fixed stabilizing control and disturbance input gains for the linear discrete-time system, we have

$$\lambda_i(A_s - B_s K) = \lambda_i^s < 1 \text{ with } B_s = \int_0^T e^{A(T-s)} ds B. \quad (23)$$

Then $\underset{\tau,\gamma}{E}(A_s - \gamma_k B_0^k K)$ can be represented as

$$\underset{\tau,\gamma}{E}(A_s - \gamma_k B_0^k K) = A_s - \underset{\gamma}{E}(\gamma_k)\underset{\tau}{E}(B_0^k)K$$
$$= [I - \min\{\Psi_1, \Psi_2\}]A_s + \min\{\Psi_1, \Psi_2\}A_s - \Psi_1 B_s K, \quad (24)$$

where $\Psi_1 = \underset{\gamma}{E}(\gamma_k)\underset{\tau}{E}(\int_{\tau_0^k}^T e^{A(T-s)} ds)/\int_0^T e^{A(T-s)} ds$ and $\Psi_2 = \underset{\gamma}{E}(\gamma_k)\underset{\tau}{E}(\int_{\tau_0^k}^T e^{A(T-s)} ds)/\int_{\tau_0^k}^T e^{A(T-s)} ds$.

Combining equation (23) with equation (24), we have

$$\lambda_i[\underset{\tau,\gamma}{E}(A_s - \gamma_k B_0^k K)] < (1 - \min\{\Psi_1, \Psi_2\}) \times \lambda_i(A_s) + \min\{\Psi_1, \Psi_2\}\lambda_i^s.$$

Therefore, in order to maintain stability, the expected values of the delays and packet losses should satisfy

$$\min\{\Psi_1, \Psi_2\} > 1 - [1 - \min\{\Psi_1, \Psi_2\}\lambda_i^s]/\lambda_i(A_s), \quad (25)$$

where Ψ_1 and Ψ_2 are functions of the delay and packet losses defined by equation (24). When this inequality is not satisfied, the cyber system needs to launch an appropriate defense to reduce the delay and packet losses in order to prevent instability; otherwise the physical system needs to be halted as it becomes unstable.

5. An illustrative example

In this illustrative example, the proposed framework is verified on a small-scale UAV helicopter with remote controller. The objective of the controller design is to stabilize the yaw rotation rate with the presence of two types of cyber-attacks. The attacker aims to maximize the payoff, which are given in terms of the network delay and packet losses in this case, such that the yaw channel becomes unstable. The defender, on the other hand, aims to limit the delay and packet losses under a certain threshold. We will show that on the cyber side, both the attacker and the defender gain their greatest payoff while on the physical system side, the optimal controller is able to maintain the yaw rate stable when the cyber state vector expressed as delay and packet loss meets the derived condition.

5.1. Physical system setup

In this illustrative example, we consider the control of the yaw rotation of a small-scale UAV helicopter. A yaw rotation, as illustrated in Figure 5, is a movement around the yaw axis of a rigid body that changes the direction it is pointing.[26] The yaw rotation control is one of the most challenging tasks in controlling small-scale UAVs because even a small control input or disturbance can cause the vibration of the light-weight body.[26] Since it has been verified that the yaw-channel dynamics for small-scale

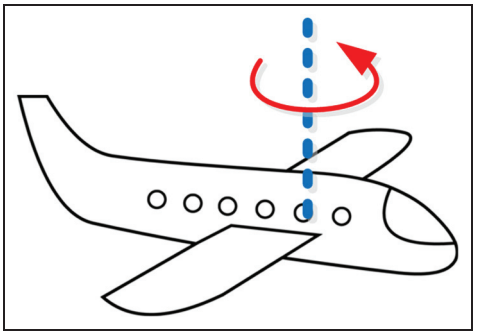

Figure 5. Illustration of a yaw rotation.

helicopters can be physically decoupled from other channels,[27,28] it is reasonable to assume that the yaw-channel dynamic is a single-input–single-output system. Furthermore, after applying the prediction-error method,[29] an accurate fourth-order model is proposed as[26]

$$\dot{x} = Ax + Bu; \quad y = Cx,$$

where $x = [x_1, x_2, x_3, x_4]^T$ consists of the first to the fourth derivatives of the yaw rotation rate, y is the yaw rotation rate that can be measured by a gyro, and

$$A = \begin{bmatrix} -2.66 & 21.94 & 3.83 & 6.05 \\ -31.03 & -3.52 & 17.10 & -3.09 \\ 6.11 & -6.96 & -9.76 & -96.38 \\ 17.17 & 25.73 & 37.18 & -33.08 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.63 \\ 6.22 \\ -29.20 \\ -14.64 \end{bmatrix}, C = [\,15.32 \quad -10.32 \quad 0.73 \quad -4.73\,].$$

The other parameters of the physical system are introduced as follows. The total simulation time is 200 steps with the sampling time of 100 ms and the positive constant α_h equals to 10^{-6}. In the first 50 steps, zero-mean exploration noises with variance of 0.006 and 0.003 are added for the odd and even steps respectively, in order to meet the persistency of excitation (PE) condition. The objective of the controller is to stabilize the yaw rotation rate$y$by driving the state vector$x$ to zero.

5.2. Cyber system setup

As illustrated in Figure 6, we suppose that the UAV is controlled by a base station through a wireless network that suffers from cyber-attacks. As stated earlier, we choose packet losses κ and time delays τas the cyber state vector in order to evaluate the effect on the network induced by the attack/defense activities, i.e. $x_c = [\kappa, \tau]^T$. Furthermore, smurf attack and slow read attack are considered.[30–32]

Smurf attack is an example of amplification distributed denial of service (DDoS) attack that exploits the unprotected networks to generate significant traffic load on the victim network.[30,31] Slow read attack, on the other hand, tries to exhaust the server's connection pool by sends legitimate application layer request but reads the response slowly.[32] Based on these characteristics, we model the delay and packet loss rate to increase exponentially under the smurf attack and linearly under the slow read attack, which are illustrated in Figure 7(a) and (b). Furthermore, the corresponding strategies that are capable of defending smurf attack and slow read attack are denoted as d_1 and d_2, respectively. We assume that when the appropriate defense strategy is loaded, the packet loss rate and the time delay decrease in a linear manner, which are illustrated in Figure 7(c) and (d). In addition, the delay and packet loss rate are modeled to decrease slowly and linearly once the attack is stopped regardless of the action of the defender. For simplicity, we mainly focus on the case where only one attack and one defense are active at a sampling instant. However, it is also briefly shown that the proposed representation can be easily expanded to apply multiple attacks and defenses.

The cyber output is defined as $y_c = x_c^T(k)\Lambda_c x_c(k) + \rho \cdot \delta\big(\|x_p\| - x_{pt}\big)$, where $\Lambda_c = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, x_{pt} is the threshold of the physical states, and $\delta(\cdot)$ is defined in Section 4. According to this definition, when the physical states are within the threshold, the cyber output is a quadratic function of the cyber state vector only.

Next, as presented in the flowchart in Figure 4, we divide the cyber output Y into four subsets, i.e. $Y = Y_0 \cup Y_1 \cup Y_2 \cup Y_3$ where Y_0, Y_1, Y_2, and Y_3 correspond to the ''healthy,''''sensitive,''''dangerous,'' and ''failed'' condition, respectively. Moreover, we define the instant reward in the form of (7) with $\xi_d = [0, \xi_{d,1}, \xi_{d,2}]$ and $\xi_a = [0, \xi_{a,1}, \xi_{a,2}]$. In other words, the costs for ''not

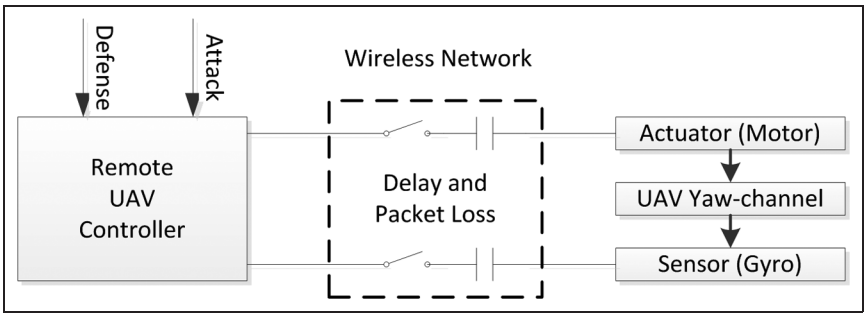

Figure 6. Diagram of the UAV with remote controller.

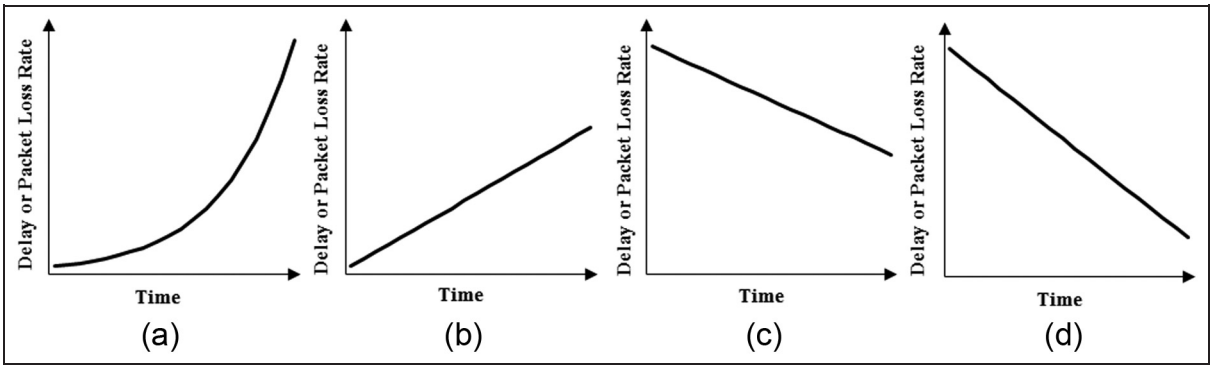

Figure 7. Models of delay/packet loss rate under (a) smurf attack, no defense; (b) slow read attack, no defense; (c) smurf attack with the corresponding defense; (d) slow read attack with the corresponding defense.

Table 1. Summary of system information used in the illustrative example.

Attacks Defenses	$A_c = [a_0, a_1, a_2]$, where a_0 denotes no attacks, a_1 denotes smurf attack, and a_2 denotes slow read attack. $D_c = [d_0, d_1, d_2]^T$, where d_0 denotes no defenses, d_1 denotes the defense against smurf attack, and d_2 denotes the defense against slow read attack.
Cyber states	$x_c = [\kappa, \tau]^T$, where κ is the packet loss rate and τ is the delay.
System dynamics	$x_c(k+1) = a_0 d_0(x_c(k) - \Delta_0) + a_0 d_1(x_c(k) - \Delta_0) + a_0 d_2(x_c(k) - \Delta_0) + a_1 d_0(\xi \cdot x_c(k)) + a_1 d_1(x_c(k) - \Delta_1) + a_1 d_2(\xi \cdot x_c(k)) + a_2 d_0(x_c(k) + \Delta_2) + a_2 d_1(x_c(k) + \Delta_2) + a_2 d_2(x_c(k) - \Delta_3)$ where $\Delta_0, \Delta_1, \Delta_2, \Delta_3 \in \mathbb{R}_+^{2\times 1}$ characterize the packet loss rate/delay linearly decrease or increase rate; $\xi > 1$ characterizes the exponentially increasing rate.
Cyber output	$y_c = x_c^T(k) \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} x_c(k) + \rho \cdot \delta(\|x_p\| - x_{pt})$, where $\lambda_1, \lambda_2, \rho, y_c \in \mathbb{R}^+$.
Subsets of cyber output	$Y = Y_0 \cup Y_1 \cup Y_2 \cup Y_3$
Payoff	$r(A_c(k), D_c(k), Y_i(k)) = x_c^T(k) \Lambda_c x_c(k) + \xi_d D_c(k) - \xi_a A_c^T(k)$, where $\xi_d = [0, \xi_{d,1}, \xi_{d,2}]$ and $\xi_a = [0, \xi_{a,1}, \xi_{a,2}]$.

launching any defenses,'', ''launching defense d_1,'' and ''launching defense d_2'' are 0, $\xi_{d,1}$, and $\xi_{d,2}$, respectively.

It is important to note that we make Y_0 be the region with ''healthy'' condition by setting the cost for launching the defense close to the upper values of Y_1. As a result, if the cyber output falls into subset Y_0, the defender tends not to launch the defense as it costs more than the payoff brought by the state. Subset Y_1, on the other hand, is modeled as the ''sensitive'' region where the defender is more likely to launch the defense to avoid the output going into subset Y_2, which is the ''dangerous'' state in this model. Likewise, if the output falls into region Y_2, there is a very high chance that the defenses needs to be launched to avoid the system going into Y_3, which is the ''failed'' region.

The system information for this particular example is summarized as in Table 1. The simulation is performed with the algorithm described in Figure 4 and numerical values shown in Table 2.

Table 2. Numerical values used in the simulation.

$\alpha(k) = 1/k$; $\beta = 0.5$; $N_a = N_d = 3$; $\xi = 1.2$; $\lambda_1 = \lambda_2 = 1$;
$\Delta_0 = [1; 1.1]$, $\Delta_1 = [50; 48]$, $\Delta_2 = [3; 2.9]$; $\xi_d = [0, 5000, 4500]$; $\xi_a = [0, 1500, 1000]$;
$Y_0 = [0, 5000)$, $Y_1 = [5000, 7200)$, $Y_2 = [7200, 12800)$, $Y_3 = [12800, \infty)$.

5.3. Simulation results

In the simulation, the optimal defense/attack policies for the cyber system and the optimal controller are derived in the presence of delay and packet losses. Since the delay and packet losses are generated from the cyber system, they are determined directly by the policy launched by the defender. After deriving the optimal defense/attack policies, two scenarios are considered in the simulation. In the first scenario, we let the defender launch the cyber defense

Table 3. Percentages for each action in the region.

	Attacker			Defender		
	a_0 No attacks	a_1 Smurf Attack	a_2 Slow read attack	d_0 No defense	d_1 Defending smurf attack	d_2 Defending slow read attack
Y_0	0.02	0.58	0.34	0.71	0.09	0.20
Y_1	0.53	0.08	0.39	0.11	0.25	0.64
Y_2	0.69	0.13	0.18	0.04	0.37	0.59
Y_3	0.71	0.13	0.16	0.03	0.40	0.57

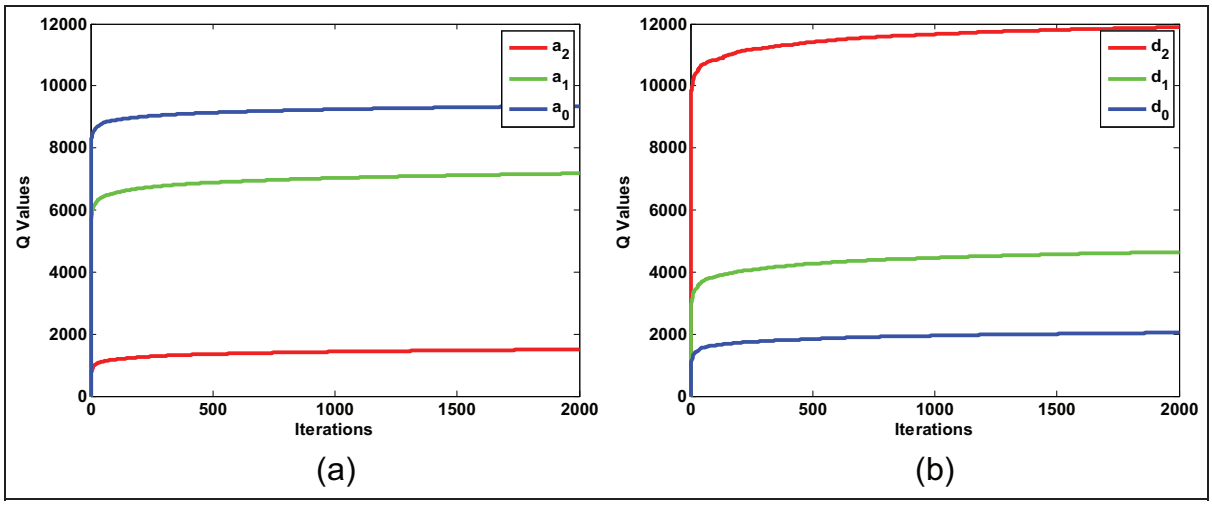

Figure 8. Q-values in region Y_1 for (a) the attacker; (b) the defender.

policy based on the probability distribution given by the derived optimal policy. By contrast, in the second scenario, the defender selects the defense actions at random.

5.3.1. Results of deriving the optimal attack/defense policies. First, we shall show the simulation results of deriving the optimal attack/defense. After about 2000 iterations, the Q-values for all action pairs converge to fixed values. To avoid redundancy, we only show the Q-values for the attacker and the defender in region Y_1 in Figure 8(a) and (b), respectively. From Figure 8 it can be concluded that the expected discounted payoff for the attacker in region Y_1 is higher if he chooses action a_0 rather than a_1 and a_2. Likewise, the expected discounted payoff values suggest the defender in region Y_i to load action$d_2$more frequently than d_0 and d_1. Furthermore, the percentages of the Q-values for each action in the regions are computed and listed in Table 3.

It can be concluded from Table 3 that when $y_c \in Y_0$, the attacker shall take action a_2 more often as it increases the delay and packet losses in a faster way. The defender, on the other hand, shall take no actions, which corresponds to our previous analysis that Y_0 is the region with ''acceptable'' health condition. With the increase in y_c, the attacker shall slow down the speed to avoid being detected by the defender, as one can conclude from the Q-value distributions in region Y_1 in the table. Correspondingly, the defender starts loading the defense more often in this sensitive region. If the attacker manages to drive y_c into region Y_2 or even Y_3, he shall stop attacking and let the system recover and go back to region Y_1 where he obtains the largest expected payoff. It is important to note that we deliberately design the system as a secure one by letting the recovery speed of the cyber states when appropriate defense is loaded much faster than the degrading speed when the system is under attacks. As a result, the attacker gains the greatest payoff only when $y_c$is large enough yet not to the degree of being detected by the defender.

The proposed model and analysis is verified through the following simulation. We start the system with the cyber state initialized to zero and stop after 1000 iterations. During iteration, the attacker and defender will (i) determine which region y_c is in and take actions according

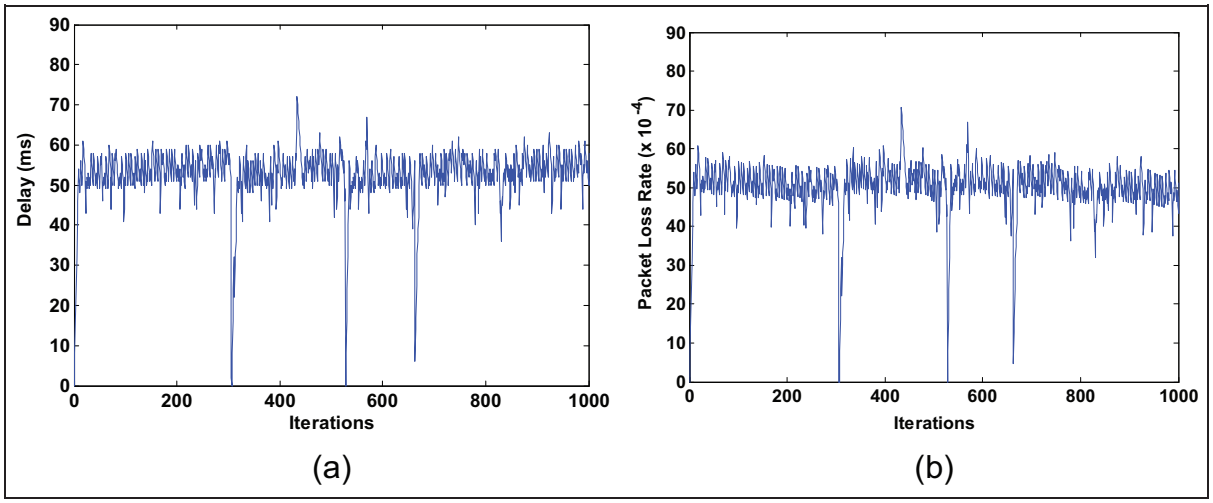

Figure 9. Evolution of the states (a): delay; (b): packet loss rate.

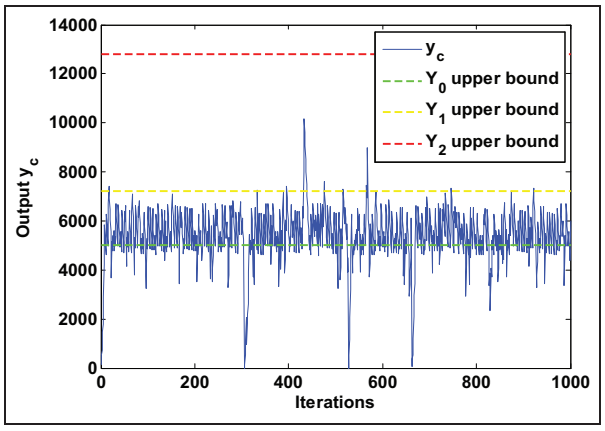

Figure 10. Evolution of the output.

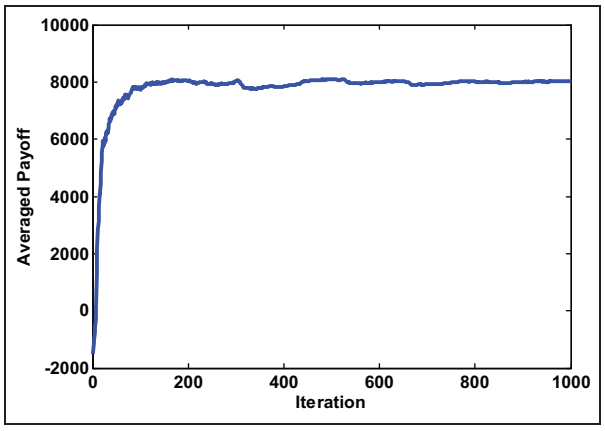

Figure 11. Evolution of average payoff.

to the probabilities given by Table 3; (ii) update the states; and (iii) calculate the accumulated payoff. The evolution of the states is shown in Figure 9.

From Figure 9 it can be concluded that after a rapid increase at the beginning, the delay and the packet loss rate remains relatively stable so that the attacker gains the largest expected payoff in terms of the delay and packet losses. This is achieved by loading much more a_0 (no attacks) than a_1 (smurf attack) and a_2 (slow read attack), as suggested by the probabilities in Table 3. Due to the stochastic property of this game, we observe that occasionally, the attacker loads the ''inappropriate'' attack (a_1) and detected by the defender, resulting in a significant drop in the states. Figure 10 shows the evolution of the output, where one can conclude that as previously analyzed, the output stays in the ''acceptable'' region at most times, goes to the ''dangerous'' region occasionally, and never reaches the ''failed'' region. The averaged payoff for the attacker is shown in Figure 11, from which we can see that after about 100 iterations, the averaged payoff tends to be stable at around 8000, which is the greatest averaged payoff for the attacker. This example shows that by applying the optimal policies the attacker is able to obtain the greatest payoff meanwhile the defender is able to keep the health condition under the ''dangerous'' level.

In addition, the simulation is repeated for the case where the two attacks/defenses can be loaded simultaneously. As a result, a table similar to Table 3 is obtained except that two extra columns are added, which are the probability distributions of simultaneously loading two attacks$(a_1 + a_2)$ and two defenses $(d_1 + d_2)$. To verify the results, we use the method mentioned earlier, in which we observe the output y_c by letting the attacker and defender select their action based on the derived probability

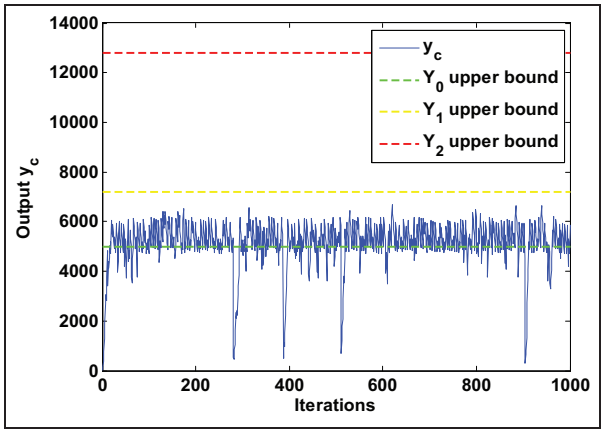

Figure 12. Evolution of the output, when two attacks/defenses can be loaded simultaneously.

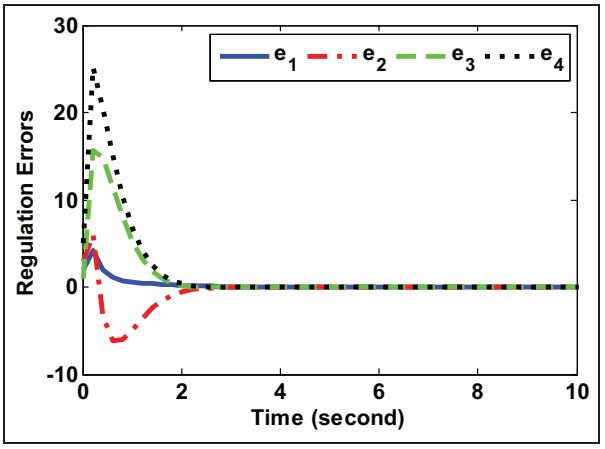

Figure 13. Regulation errors in Scenario I where the cyber defense is optimal.

distributions. The results are shown in Figure 12, in which one can conclude that the output stays in the "acceptable" region at most times and never goes to the "dangerous" or the "failed" region. This results agree with our previously analysis and verify that the proposed representation can be used in the case where multiple attacks can be loaded simultaneously.

5.3.2. Scenario I: defender chooses the optimal policy. In this scenario, we let the defender launch the defense policy based on the probability distribution given by the derived optimal policy. As a result, the delay and packet losses have been limited to relatively low values so that the system always stays out of the failed region, which is as verified in Figure 9(a). Consequently, equation (25) is satisfied in this scenario. The simulation results of the regulation errors for the physical system are shown in Figure 13, where the state regulation errors converge to zero thus forcing the closed-loop system being stable. Therefore, we show that on the cyber side, both the attacker and the defender gains their greatest payoff while on the physical side, the optimal controller is able to maintain the plant stable when the cyber state vector meets the derived criterion.

5.3.3. Scenario II: defender chooses a random policy. In the second scenario, the cyber defense is selected at random rather than based on the optimal probability distribution given in Table 3. As a result, the attacker manages to compromise the system in some cases and the cyber states go far beyond the limit, as verified in Figure 14 in which the time delay is plotted. Consequently, equation (25) cannot be satisfied and thus the system becomes unstable. The regulation errors in this scenario are plotted in Figure 15, where it can be seen that the errors do not converge.

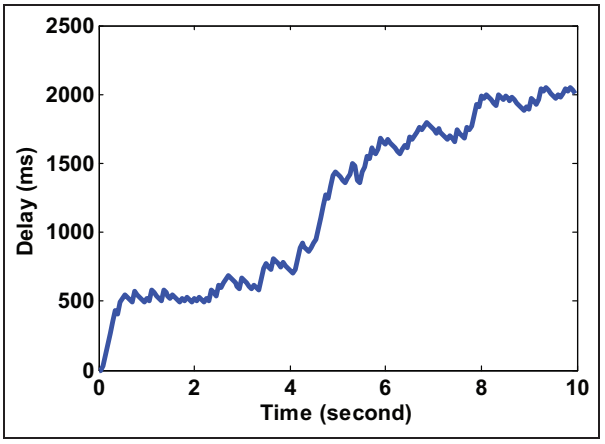

Figure 14. Delay in Scenario II where the cyber defense is randomly selected.

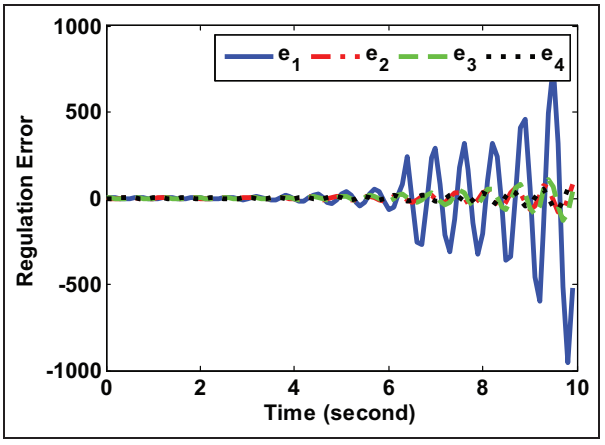

Figure 15. Regulation errors in Scenario II where the cyber defense is randomly selected.

In summary, the simulation results verify that that the decisions made on the cyber system have an effect on the convergence of the physical system. The system is stable when applying the optimal control in the physical plant and optimal defense policy in the cyber system. If the states go abnormal such that equation (25) is not satisfied, appropriate actions needs to be launched on the cyber system to bring them back to normal or the physical plant has to be shut down to avoid further damages.

6. Conclusions and future work

With the increasing meshing among the cyber-connected elements with the physical entities, the representation for such cyber-physical system becomes more complicated. In this paper, we have proposed a representation that captures the interrelationship between the cyber and physical systems such that the states in the physical system affect the decision made on the cyber systems and vice versa. Based on this representation, the optimal defense and attacks are given to gain the greatest payoff. An optimal controller from the literature is revisited to maintain the stability of the physical system in the presence of the uncertainties induced by the cyber state vector. Since the proposed representation is in a general form, it can be used in a variety of applications including autonomous systems. In particular, the cyber defender is able to make thorough decisions by selecting appropriate cyber state vector and output and customizing the payoff function that is of interest. Meanwhile, there are some recent works focusing on modelling and controlling for multi-agent networks or cyber-physical systems.[33–35] For example, the work by Xue et al. characterizes a binary notion of security and characterizes security levels in terms of the graph matrix and its spectrum,[33] which is complementary to control-theoretic modeling of attacks in cyber-networks and networked control systems. Based on these works, as future work, we can consider studying the impact of different attacks on the network performance to generate a more accurate model for the cyber system dynamics.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Zhou Y and Baras JS. CPS modeling integration hub and design space exploration with application to microrobotics. *Lect Notes Control Inf Sci* 2013; 449: 23–42.
2. Baumman H and Sandmann W. Markovian modeling and security measure analysis for networks under flooding DoS attacks. In: *20th Euromicro international conference on the parallel, distributed and network-based processing*, Garching, Germany, 15–17 February 2012.
3. Zhu Q and Basar T. Robust and resilient control design for cyber-physical systems with an application to power systems. In: *50th IEEE conference on decision and control and European control conference*, Orlando, FL, 12–15 December 2011.
4. Ten CW, Manimaran G and Liu CC. Cybersecurity for critical infrastructures: attack and defense modeling. *IEEE Trans Syst Man Cybern Part A Syst Humans* 2010; 40: 853–865.
5. Sallhammar K, Helvik BE and Knapskog SJ. Towards a stochastic model for integrated security and dependability evaluation. In: *IEEE conference on availability, reliability and security*, Vienna, Austria, 20–22 April 2006.
6. Aenes A, Salhammar K, Haslum K, et al. Real-time risk assessment with network sensors and intrusion detection systems. In: *International conference on computational intelligence and security*, Xi'an, China, 15–19 December 2005.
7. Kwon C, Liu W and Hwang I. Security analysis for cyber-physical systems against stealthy deception attacks. In: *American control conference (ACC)*, Washington, DC, 17–19 June 2013.
8. Liu L, Esmalifalak M, Ding Q, et al. Detecting false data injection attacks on power grid by sparse optimization. *IEEE Trans Smart Grid* 2014; 5: 612–621.
9. Teixeira A, Amin S, Sandberg H, et al. Cyber security analysis of state estimators in electric power systems. In: *IEEE conference on decision control*, Atlanta, GA, 15–17 December 2010.
10. Fawzi H, Tabuada P and Diggavi S. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Trans Autom Control* 2014; 59: 1454–1467.
11. Amin S, Cárdenas A and Sastry S. Safe and secure networked control systems under denial-of-service attacks. *Hybrid Syst Comput Control* 2009; 5469: 31–45.
12. Zhu M and Martínez S. Stackelberg game analysis of correlated attacks in cyber-physical systems. In: *American control conference*, San Francisco, CA, 29 June–1 July 2011.
13. Pasqualetti F, Dorfler F and Bullo F. Attack detection and identification in cyber-physical systems. *IEEE Trans Autom Control* 2013; 58: 2715–2729.
14. Xu H, Jagannathan S and Lewis FL. Stochastic optimal control of unknown linear networked control system in the presence of random delays and packet losses. *Automatica* 2012; 48: 1017–1030.
15. Nguyen HL and Nguyen UT. Study of different types of attacks on multicast in mobile ad hoc networks. In: *IEEE international conference on networking, international conference on systems, and international conference on mobile communications and learning technologies (ICNICONSMCL)*, Mauritius, 23–29 April 2006.
16. Lagoudakis M and Parr R. Value function approximation in zero-sum Markov games. In: *Proceedings of the eighteenth conference on uncertainty in artificial intelligence*, Alberta, Canada, 1–4 August 2002.
17. Littman ML. Markov games as a framework for multi-agent reinforcement learning. In: *Proceedings of the eleventh*

international conference on machine learning, New Brunswick, NJ, 10–13 July 1994.
18. Basar T and Olsder GJ. *Dynamic noncooperative game theory*. 2nd ed. Society for Industrial and Applied Mathematics, 1995.
19. Raghaven T, Ferguson T, Parthasarathy T, et al. *Stochastic games and related topics*. Springer, 1990.
20. Puterman ML. *Markov decision processes: discrete stochastic dynamic programming*. New York: John Wiley & Sons, 1994.
21. Littman M. Friend-or-foe Q-learning in general-sum Markov games. In: *Proceedings of eighteenth international conference on machine learning*, Williamstown, MA, 28 June 28–1 July 2001.
22. Li H, Chow MY and Sun Z. Optimal stabilizing gain selection for networked control systems with time delays and packet losses. *IEEE Trans Control Syst Technol* 2009; 17: 1154–1162.
23. Xu H, Jagannathan S and Lewis FL. Stochastic optimal design for unknown linear discrete-time systems zero-sum games in input–output form under communication constraints. *Asian J Control* 2014; 16: 1263–1276.
24. Kreisselmeier G. Adaptive control of a class of slowly time-varying plants. *Syst Control Lett* 1986; 8: 97–103.
25. Yu J, Su Z, Wang M, et al. Control of yaw and pitch maneuvers of a multilink dolphin robot. *IEEE Trans Rob* 2012; 28: 318–329.
26. Cai G, Chen BM, Peng K, et al. Modeling and control of the yaw channel of a UAV helicopter. *IEEE Trans Ind Electron* 2008; 55: 3426–3434.
27. Mettler B. *Identification modeling and characteristics of miniature rotorcraft*. Norwell, MA: Kluwer, 2003.
28. Shim DH, Kim HJ and Sastry S. Control system design for rotorcraft-based unmanned aerial vehicle using time-domain system identification. In: *IEEE conference on control application*, Anchorage, AK, 25–27 September 2000.
29. Ljung L. *System identification: theory for the user*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
30. Kumar S. Smurf-based distributed denial of service (DDoS) attack amplification in internet. In: *Second international conference on internet monitoring and protection, 2007 (ICIMP 2007)*, Silicon Valley, CA, 1–6 July 2007, pp. 25–25.
31. Zargar GR and Kabiri P. Identification of effective network features to detect smurf attacks. In: *IEEE student conference on research and development (SCOReD)*, UPM Serdang, Malaysia, 16–18 November 2009.
32. Cai S, Liu Y and Gong W. Client-controlled slow TCP and denial of service. In: *43rd IEEE conference on decision and control*, Bahamas, 14–17 December 2004.
33. Xue M, Wang W and Roy S. Security concepts for the dynamics of autonomous vehicle networks. *Automatica* 2014; 50: 852–857.
34. Chen CW and Roy S. State detection from local measurements in network synchronization processes. *Int J Control* 2013; 86: 1634–1645.
35. Roy S, Xue M and Das SK. Security and discoverability of spread dynamics in cyber-physical networks. *IEEE Trans Parallel Distrib Syst* 2012; 23: 1694–1707.

Author biographies

Haifeng Niu is a PhD student at the Missouri University of Science and Technology, Rolla, MO.

Dr S Jagannathan is a Rutledge-Emerson Endowed Chair Professor of Electrical and Computer Engineering at Missouri University of Science and Technology, Rolla, MO. He is the IEEE CSS Tech Committee Chair on Intelligent Control. He is a Fellow of the Institute of Measurement and Control, UK.