Optimal Defense and Control for Cyber-Physical Systems

Haifeng Niu and S. Jagannathan

***Abstract*—In this paper, we present a novel representation for cyber-physical systems wherein the states of the cyber system are incorporated into the physical system and vice Next, by using this representation, optimal strategies are derived for the defender and the attacker by using zero-sum game formulation and iterative Q-learning is utilized to obtain the Nash equilibrium. In addition, a Q-learning-based optimal controller is revisited for the physical system with the presence of uncertain dynamics resulting from the cyber system under attacks. The benefit of the learning strategy is that the approach can handle a variety of attacks provided they affect packet losses and delays. Simulation results on the yaw-channel control of the unmanned aerial vehicle (UAV), show that for the cyber system, both the defender and the attacker gain their largest payoff and for the physical system, the controller maintains the system stable.**

I. Introduction

With the growth in management and networking, the security in Cyber-physical systems (CPS) has received a lot of attention. As physical and cyber capabilities are becoming more and more intertwined, a framework which presents the representation of the physical system, the cyber system dynamics as well as their interrelationship is increasingly needed.

In general, two types of representation for analyzing the security of CPS have been found in the existing works: one that describe the effect on the cyber systems under some certain attacks [1-4] and the other study the influence brought by the cyber-attacks on the physical system [5-8]. The former works explores the behaviors of the malicious attackers and the defenders, attempt to formulate the cyber state changes under attacks, and offer appropriate strategies to bring the cyber states back to normal condition. For instance, the authors in [1] describe the Denial of Service (DoS) attacks by a continuous-time Markov chain and use the state-space method to compute the security metrics accurately. In [2] the optimal cyber defense is derived by modeling the action-pairs of the attacker and the defender as a zero-sum game. The authors of [3] give the definition of the measure of vulnerabilities in cyber-physical systems and introduce a security framework consisting of attack detection as well as mitigation strategies. The authors in [4] assess the cyber security level by deriving the probabilities of the malicious attacker and using those probabilities to create a transition model through a game-theoretic approach.

*Research supported in part by NSF I/UCRC on Intelligent Maintenance Systems. Haifeng Niu (hnhy6@mst.edu) and S. Jagannathan (sarangap@mst.edu) are with the Department of Electrical & Computer Engineering, Missouri University of Science and Technology (formerly University of Missouri–Rolla), Rolla, MO, 65401, USA.

In contrast, others [5-8] concentrate on modeling the physical system dynamics under attacks by modifying the classic state-space description such that the attacks can be included. For example, in [5] an additive term is embedded to the system state with the purpose of simulating the false data injection attack. Likewise, in [6], an extra term is added to characterize the deception attack. Unlike [6], the authors in [7] model the deception attackers with several objectives and propose a way to incorporate the stealthy deception attacks in both nonlinear and linear estimators. In [8], the authors multiply the control input by a coefficient in order to describe the influences brought by the DoS attacks.

Despite these developments, a significant effort is still needed due to weaknesses [9] observed. First, these representations can only characterize a certain type of attack as each attack affects the dynamics of the physical or cyber systems in various ways. It is worth to notice that the author in [9] introduces a unified framework which can detect different attacks however this work still has the following two drawbacks. Second, it is a challenge to implement the representation mentioned above if one assumes that the system dynamics with the presence of attacks are known while in practice the system dynamics become uncertain as shown in this paper. At last, these representations tend to neglect the interactions between the physical system controller and the cyber defense policy.

In this paper, we propose a mathematical representation for CPS, in which the activities of the cyber system have an impact on the physical system states and vice versa. The major contributions of this work include: 1) the proposal of a novel representation for CPS that captures their interrelationship; 2) the derivation of the optimal policies for the defender as well as the attacker through Q-learning approach; 3) the application of the optimal controller [10] for the physical system with uncertain dynamics effected by the cyber system; and 4) the demonstration of the proposed scheme on a small scale UAV helicopter under attacks.

One of the benefits with the learning approach is that it does not require the adversary model. As long as an attack has been launched a number of times during the learning phase, the system is able to learn the optimal defense strategy against it in terms of the predefined cost function

II. Proposed Representation for CPS

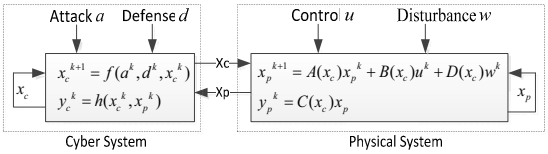

Fig. 1 Proposed representation for CPS.

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.98

As depicted in Fig. 1, the proposed scheme consists two representations: one for the cyber defense and the other for the optimal controller.

A. Cyber system

First we describe the cyber system with the presence of attacks by the following nonlinear discrete-time system

$$x_c(k+1) = f\left(a(k), d(k), x_c(k)\right), \quad (1)$$

where $x_c \in \mathbb{R}^{N_c}$ is the cyber system states where N_c is the dimension of the state vector for the cyber system. $d, a \in \mathbb{R}$ are the actions taken by the defender and attacker the respectively.

The cyber state x_c consists of a set of performance metrics such as throughput, latency, packet loss rate, and others. In some cases where the defender is concerned about the information security, a few other security metrics such as the changes of IP addresses or the number of unsuccessful verifications are included. Obviously, the cyber states are subject to the defense and attack strategies and we use the function f to describe such relationship.

In particular, we introduce a more concrete representation for the cyber system dynamics as

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

where $A_c = \left[a_0, a_1, ..., a_{N_a}\right]$ is a row vector including all N_a possible attacks where $a_i \in \{0,1\}$ denotes a type of attack wherein $a_i = 1$ implies that the i^{th} attack has been launched and $a_i = 0$ otherwise. Furthermore, we let $a_0 = 1$ if there is no attack currently being launched. The defense vector D_c can be explained in the same manner. Finally, we let $F = \left[f_{00}, f_{01}, ..., f_{0N_d}; ...; f_{N_a 0}, f_{N_a 1}, ..., f_{N_a N_d}\right]$ be a matrix of functions with each element $f_{ij} : \mathbb{R}^{N_c \times 1} \to \mathbb{R}^{N_c \times 1}$ describing the effect on the cyber states brought by the ongoing defense / attack pair (a_i, d_j). We assume that when two or more defense actions (and attacks) are launched simultaneously, the effect of each defense action (and attack) is independent.

As shown in Fig. 1, we use the following nonlinear equation to describe the output of the cyber system

$$y_c(k) = h\left(x_c(k), x_p(k)\right), \quad (3)$$

where $y_c \in \mathbb{R}$ is the cyber system output and $x_p \in \mathbb{R}^{N_p}$ is the physical system state where N_p is the dimension of the physical state vector. The output is a quantized metric that is used to indicate the health level of the cyber system. The function h is selected such that it should utilize the observed states to generate a precise prediction of the ongoing and potential attacks. One example of function h is written as

$$y_c(k) = x_c^T(k)\Lambda_c x_c(k) + x_p^T(k)\Lambda_p x_p(k), \quad (4)$$

where $y_c \in \mathbb{R}$; $\Lambda_c \in \mathbb{R}^{N_c \times N_c}$ and $\Lambda_p \in \mathbb{R}^{N_p \times N_p}$ stand for the weighting coefficient matrices of each state. This quadratic form maps the physical and cyber states vector onto a scalar that offers an approximate indication of the system health level.

One can evaluate the health level or even the type of attacks by analyzing the physical states as well as the cyber system state. It is important to introduce the cyber output because the states need to be interpreted in order to help the administrator make the correct defense decisions. Note that the physical system state is also included in the cyber system such that an accurate and comprehensive estimation of the system health level can be obtained.

B. Physical system

As presented in the right block of Fig. 1, we describe the physical system dynamics by a linear discrete system with the presence of disturbances:

$$x(k+1) = A(x_c)x(k) + B(x_c)u(k) + D(x_c)w(k), \; y^k(k) = Cx(k), \quad (5)$$

where u is the control input, w the disturbance input, y the output, and A, B, C and D denote the system matrices with appropriate dimensions.

It is important to emphasize that unlike the traditional linear discrete system, this system dynamics are functions of the cyber state x_c. That is to say, the cyber system state will also change the physical system dynamics. For example, a large network-induced packet loss or delay could degrade the system performance and even leads to instability.

Here, the cyber states, which are updated on the basis of the defense/attack decisions, change the physical system dynamics. Subsequently, it is necessary to adjust the control input in order to drive the physical system states back to the desired zone. The changes in both physical and cyber states, in turn influence the cyber output and further determine the defense/attack decisions. A summary of this interrelationship between the physical and the cyber systems is shown in Fig. 2.

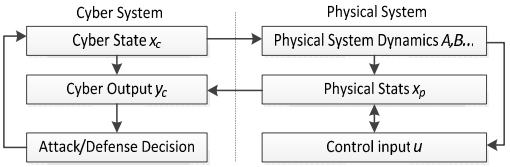

Fig. 2 Interrelationship between the physical and the cyber system.

III. Optimal Attack/Defense Policy

In this section, we begin with modeling the interactions between the attacker and the defender with a two-player zero-sum game. After giving the definition of the instant payoff and the discounted payoff function, we proceed with introducing a lemma which gives the solution of the optimal policy. Next, we give the definition of the Q-function and show in Theorem 1 that the Q-function eventually converges to the game value by using the Minimax-Q algorithm [11]. Consequently, the optimal strategies that give the defender and the attacker their largest discounted payoff can also be obtained.

Consider the cyber system described by (2) and output function in the form of (4). Then we can model the system as a Markov decision process where the cyber state at the next sampling instant, $x_c(k+1)$, is determined by the cyber state at

the current sampling instant, $x_c(k)$, together with the attack/defense action pair $(A_c(k), D_c(k))$. The attacker and the defender update their defense strategies based on the health level indicated by y_c, which is computed based on the observed state vector x_c and x_p.

Let Y denotes the set of all admissible values of y_c, then it is impractical to derive the optimal strategy for every y_c as Y contain infinity elements. Thus, we propose to divide Y into a number of subsets and the objective is to derive the optimal strategies for every subset rather than for every element. Each subset of Y corresponds to one level of health status, as illustrated in Fig. 3. The defender makes the decision on defending strategies based on the specific subset that y_c is in. Obviously, the more number of subsets Y is divided into, the more precise the model becomes. But more computation is involved as more subset of optimal strategies require to be computed.

Fig. 3 Each subset corresponds to a heath level.

Let $r(A_c(k), D_c(k), Y_i(k))$ denotes the instant payoff (cost or reward) at time instant k in subset $Y_i(k)$ for the action pair $(A_c(k), D_c(k))$. Let r_d and r_a be the instant payoff of the defender and the attack respectively. With the assumption of zero-sum game, we can have the relationship

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

Furthermore, define the instant reward as

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

which is a function of the cost of the physical state, cyber state, attack, and defense. Define the defense cost is as $\xi_d D_c(i)$ with $\xi_d = \left[\xi_{d,1}, \xi_{d,2}, ..., \xi_{d,N_d}\right]$ and $\xi_{d,i} \in \mathbb{R}^+$ being the corresponding cost for launching defense d_i. Likewise, $\xi_a = \left[\xi_{a,1}, \xi_{a,2}, ..., \xi_{a,N_a}\right]$ is the row vector describing the cost for launching attacks. Next, we only show the derivation of the optimal attack strategy since the optimal defense strategies can be derived in the same fashion.

After defining the instant payoff, now we are interested in the discounted payoff over multiple stages. Let $\Xi_D = \{D_c(1), D_c(2), ...D_c(k)...\}$ and $\Xi_A = \{A_c(1), A_c(2), ...A_c(k)...\}$ stand for the *policies* for the defense and attack respectively, where $A_c(k)$ and $D_c(k)$ are the actions at time instant k. A policy is a sequence of decisions over multiple stages that mathematically describes the player's plan for the game. Now define the expectation of the discounted cost function V within each subset Y_i as

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

where $\beta \in [0,1)$ denotes the discount factor. Subsequently, the attacker aims at deriving the optimal policy Ξ_A within each subset Y_i such that the expected discounted payoff V can be maximized. Likewise, the objective of the defender is to derive the correct defense policy Ξ_D within each Y_i in order to minimize V. In other words, we need to solve $\Xi_A = \arg\max_{\Xi'_A} V_a(\Xi'_A)$ and $\Xi_D = \arg\max_{\Xi' D} V_d(\Xi'_D)$. The following lemma 1 is introduced before we proceed to derive the optimal policies.

Lemma 1. [12] The policy (Ξ_A^*, Ξ_D^*) is optimal if the following fixed-point Bellman equation is satisfied

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

where p is the transition probability from current state Y_i to the next state Y_i' upon the action pair (A_c, D_c).

Now we use iterative Q-learning technique to search for the game value $V(\Xi_A^*, \Xi_D^*, Y_i)$ in (9). Define the Q-function for each subset Y_i as

$$Q(A_c, D_c, Y_i) = r_i + \beta\sum_{Y_i' \in Y} p(Y_i' \mid Y_i, A_c, D_c) V(\Xi_A, \Xi_D, Y_i') \quad (10)$$

where r_i is short for $r(A_c, D_c, Y_i)$. Accordingly, define the optimal action dependent value function Q^* of the game as

$$Q^*(A_c, D_c, Y_i) = r_i + \beta\sum_{Y_i' \in Y} p(Y_i' \mid Y_i, A_c, D_c) V(\Xi_A H^*, \Xi_D^*, Y_i'). \quad (11)$$

From (9) to (11), it can be concluded that if the action pair policy (Ξ_A, Ξ_D) is optimal, the optimal Q-function $Q^*(A_c, D_c, Y_i)$ is then equal to the game value $V(\Xi_A^*, \Xi_D^*, Y_i)$. That is to say,

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

The Minimax-Q algorithm proposed in [11] is used to derive $Q^*(A_c, D_c, Y_i)$ since it provides strong convergence guarantees according to the following theorem.

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 (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 update law given by

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

where $\partial(i) \in \mathbb{R}^+$ 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* [11] calculated by

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

where $\pi_a(A_c, Y_i)$ denotes the probability for the attacker to take attack action A_c given $y_c \in Y_i$. The proof of Theorem 1 is similar to the one shown in [11]. Note that In practice, however, Q* is considered as converged when ||Q*(k+1)-Q*(k)|| is less than a threshold.

IV. Optimal Controller Design

In this section, we give the optimal control input and show that the system can be stabilized only if the cyber states satisfy certain criterion. The derivation of the system dynamics and the Q-function update law are taken from the paper [10]. Consider the linear continuous system described as

$$\dot{x}(t) = Ax(t) + \gamma(t)Bu(t - \tau(t)); \quad y(t) = Cx(t)s, \quad (15)$$

where τ is the delay and $\gamma(t)$ is the $n \times n$ identity matrix if the control input is received at time t and null matrix if the control input is lost. Let T_s be the sampling time, the system can be discretized as

$$x_{k+1} = A_s x_k + \sum_{i=0}^{b} \gamma_{k-i} B_i^k u_{k-i}; \quad y_k = Cx_k, \quad (16)$$

where b is the maximum number of delayed control input during the sampling interval; $x_k = x(kT)$; $A_s = e^{AT}$; $B_0^k = \int_{\tau_0^k}^{T} e^{A(T-s)} ds B \cdot \mathbf{1}(T - \tau_0^k)$; for $i = 1, 2, ..., b$, $B_i^k = \int_{\tau_i^k - iT}^{\tau_{i-1}^k - (i-1)T} e^{A(T-s)} ds B \cdot \delta(T + \tau_{i-1}^k - \tau_i^k) \cdot \delta(\tau_i^k - iT)$; $\delta(x) = 1$ if $x \geq 0$ and $\delta(x) = 0$ if $x < 0$; and $\gamma_{k-i} = 1$ if u_{k-i} is received during $[kT_s, (k+1)T_s)$ and $\gamma_{k-i} = 0$ otherwise. Let the augmented state z_k be defined as: then the system dynamics become (17)

$$z_{k+1} = A_{zk} z_k + B_{zk} u_k, \; y_k^n = C_z z_k, \quad (17)$$

where ("$\mathbf{0}$"denotes the null vector with appropriate dimension)

$$A_{zk} = \begin{bmatrix} A_s & \gamma_{k-1}B_1^k & \cdots & \cdots & \gamma_{k-b}B_b^k \\ 0 & \cdots & \cdots & \cdots & 0 \\ \mathbf{0} & & diag\{I_m ... I_m\} & & \mathbf{0} \end{bmatrix}, B_{zk} = \begin{bmatrix} \gamma_k B_0^k \\ I_m \\ \mathbf{0} \end{bmatrix},$$

$C_z = diag\{C, I_m, I_m, ..., I_l\}$;

$y_k^n = [y_k^T \; u_{k-1}^T \cdots u_{k-b}^T \; w_{k-1}^T \cdots w_{k-b}^T]^T$ where I_m, I_l are $m \times m$ and $l \times l$ identity matrices. The 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 $R_z = R / b$ and $S_z = diag\{S, R/b, ..., R/b\}$. The cost function is given as $J_k = \underset{\tau,\gamma}{E}\left(z_k^T P_k z_k\right)$ where $P_k \geq 0$. Define the Q-function as

$$Q(z_k, u_k) = \underset{\tau,\gamma}{E}\left(r(z_k, u_k) + J_{k+1}\right) = \begin{bmatrix} z_k^T, u_k^T \end{bmatrix} \underset{\tau,\gamma}{E}(H_k) \begin{bmatrix} z_k^T, u_k^T \end{bmatrix}^T, \quad (18)$$

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 as

$$\bar{H}_k = \begin{bmatrix} \bar{H}_k^{zz} & \bar{H}_k^{zu} \\ \bar{H}_k^{uz} & \bar{H}_k^{uu} \end{bmatrix} = \begin{bmatrix} S_z + \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} A_{zk}\right) & \underset{\tau,\gamma}{E}\left(A_{zk}^T P_{k+1} B_{zk}\right) \\ \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} A_{zk}\right) & R_z + \underset{\tau,\gamma}{E}\left(B_{zk}^T P_{k+1} B_{zk}\right) \end{bmatrix}. \quad (19)$$

Therefore, the optimal control gain can be represented in terms of $\bar{H}_k$ as $K_k = \left(\bar{H}_k^{uu}\right)^{-1} \bar{H}_k^{uz}$. Furthermore, 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. Define the residual as $e_{hk+1} = \hat{J}_{k+1} - \hat{J}_k + r(z_k, u_k)$, then dynamics of the residue becomes

$$e_{hk+1} = r(z_k, u_k) + \hat{\bar{h}}_{k+1}^T \Delta W_k, \text{ where } \Delta W_k = \bar{w}_{k+1} - \bar{w}_k. \quad (20)$$

Now define the auxiliary residual error vector as $\Xi_{hk} = \Gamma_{k-1} + \hat{\bar{h}}_k^T \Omega_{k-1}$ where $\Omega_{k-1} = \begin{bmatrix} \Delta W_{k-1} & \cdots & \Delta W_{k-1-j} \end{bmatrix}$ and $\Gamma_{k-1} = [r(z_{k-1}, u_{k-1}) \quad r(z_{k-2}, u_{k-2}) \quad \cdots \quad r(z_{k-1-i}, u_{k-1-j})]$.

Then we have $\Xi_{hk+1} = \Gamma_k + \hat{\bar{h}}_{k+1}^T \Omega_k$. Target matrix $\bar{H}_k$ is updated according to

$$\hat{\bar{h}}_{k+1} = \Omega_k \left(\Omega_k^T \Omega_k\right)^{-1} \left(\alpha_h \Xi_{hk}^T - \Gamma_k^T\right). \quad (21)$$

At last, we give the sufficient condition for the cyber state that need to satisfy such that the system is stochastically stable. The linear time-varying discrete-time system can be represented as $z_{k+1} = A_{zk}^* z_k$ [10]. Subsequently, the expectation of A_{zk}^* can be written as

$$E\left(A_{zk}^{**}\right) = \begin{bmatrix} E\left(A_s - \gamma_k B_0^k K\right) & E\left(\gamma_{k-b} B_b^k\right) & \\ -K & 0 & \\ & & Diag\{I_m, I_m ... I_m\} \end{bmatrix}$$

.

V. An Illustrative Example

The proposed scheme is verified on a small-scale UAV helicopter with a remote controller. The objective of the attacker is to maximize the payoff, which functions of the network packet loss and delay, such that the physical yaw channel becomes unstable.

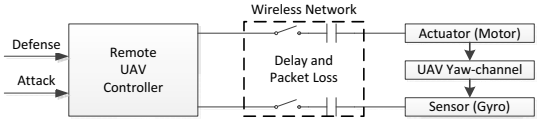

Fig. 4 Diagram of UAV with remote controller.

A. Physical System Representation

An accurate model of the yaw-channel is given in [13] as

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

where $x = \begin{bmatrix} x_1, x_2, x_3, x_4 \end{bmatrix}^T$ are the first to the fourth derivatives of the yaw rotation rate; y is the yaw rotation rate which can be measured by sensors like gyros; and

$$A=\begin{bmatrix} -2.66 & 21.94 & 3.83 & 6.05 \\ -31.03 & -3.52 & 17.10 & -3.09 \\ 6.11 & -6.96 & -9.76 & -96.38 \\ 17.17 & 25.73 & 37.18 & -33.08 \end{bmatrix}, B=\begin{bmatrix} 0.63 \\ 6.22 \\ -29.20 \\ -14.64 \end{bmatrix}, C^T=\begin{bmatrix} 15.32 \\ -10.32 \\ 0.73 \\ -4.73 \end{bmatrix}.$$

The sampling time is 100ms with the total simulation time of 200 steps.

B. Cyber System Representation

As illustrated in Fig. 4, we suppose the UAV is controlled by a base station via a wireless network that is vulnerable to various cyber-attacks. The time delays τ and packet losses κ are chosen as the cyber state vector, i.e., $x_c=[\kappa,\tau]^T$. Moreover, smurf attack and slow read attack are selected in the simulation. Smurf attack is a type of denial of service (DoS) attack which exploits the unprotected networks by generating significant traffic load. Slow read attack aims to congest the servers' connection pool by sending multiple legitimate application-layer requests and reading the response slowly. Based on their characteristics, we model the packet loss rate and delay to increase linearly under slow read attacks and exponentially under smurf attacks. Moreover, let d_1 and d_2 denote the corresponding actions that are able to defend the smurf attack and the slow read attack, respectively. Similarly, it is assumed that when the correct defense strategy is launched, the packet losses and the network delay decrease linearly.

The output of the cyber states is defined as a quadratic function. Next, we divide the output Y into totally four subsets $Y=Y_0\bigcup Y_1\bigcup Y_2\bigcup Y_3$ where Y_0, Y_1, Y_2, Y_3 correspond to the "healthy", "acceptable", "critical", and "compromised" health level respectively. Furthermore, we let the instant reward function be in the form of (7) with $\xi_d=[0,\xi_{d,1},\xi_{d,2}]$ and $\xi_a=[0,\xi_{a,1},\xi_{a,2}]$. That is to say, the costs for "no defenses", "launching defense d_1", and "launching defense are 0, $\xi_{d,1}$, and $\xi_{d,2}$, respectively.

Note that we make the subset Y_0 be the subset with "healthy" condition by configuring the cost of launching the defense very close to the upper bound of Y_1. Consequently, once the cyber output is in subset Y_0, the defender is unlikely to launch the defense because the cost is larger than the payoff. On the other hand, subset Y_1 is made as the "acceptable" subset in which the defender tends to launch the defense in order to avoid the cyber output going into subset Y_2, the "critical" subset. Similarly, if the cyber output falls into subset Y_2, there is a high chance that the defender needs to take actions to avoid the output going into Y_3, the "compromised" subset.

C. Simulation Results

Two scenarios have been considered in the simulation, after deriving the optimal defense/attack policies. In the first scenario, the defender launches the defense policy according to the derived optimal probability distribution. In the second scenario, by contrast, the defender chooses the defense strategies at random.

1) Results of the optimal attack/defense policies derivation

After about 1800 iterations, the Q-values for all actions converge to some fixed values. The percentages of the Q-values for each action pair are listed in TABLE I.

TABLE I. PERCENTAGES OF EACH ACTION IN ALL SUBSET

	Defender			Attacker		
	d_0	d_1	d_2	a_0	a_1	a_2
Y_0	0.71	0.09	0.20	0.02	0.58	0.34
Y_1	0.11	0.25	0.64	0.53	0.08	0.39
Y_2	0.04	0.37	0.59	0.69	0.13	0.18
Y_3	0.03	0.40	0.57	0.71	0.13	0.16

One can conclude from TABLE I that the attacker shall load a_2 more frequently when $y_c\in Y_0$ because it increases the packet losses and delay more quickly. On the other hand, the defender shall load no action, which verifies our earlier analysis where Y_0 is configured as the "acceptable" subset. With the increase of y_c, the attacker shall slow-down in order to avoid too much exposure to the defender, as is verified by the Q-value distributions in Y_1 in TABLE I. Correspondingly, the defender begins launching the defense more frequently in this acceptable subset. Once the output y_c falls into subset Y_2, the attacker shall halt attacking actions and wait for the system recover to subset Y_1 where he/she gains the greatest expected payoff. Note that we design the cyber system as a secure one by setting the recovery speed when appropriate defense is launched much faster than the degrading speed when the system is under attacks. Consequently, the attacker obtains the largest payoff only when the output y_c is degraded enough yet not being detected by the defender.

2) Scenario I: the defender loads the optimal policy

The proposed scheme and analysis has been verified by the following scenario. We begin with the cyber states initialized to zero and stop the simulation after 1000 iterations. During the iteration the defender and the attacker determine which subset the output y_c is in and take actions based on the probabilities given by TABLE I.

Fig. 5 shows the evolution of the output y_c, from which one can conclude that the cyber output y_c stays in the "acceptable" subset for the most of times, goes to the "critical" subset occasionally, and never falls into the "compromised" subset. This verifies that the attacker obtains the largest payoff by choosing the optimal policies meanwhile the defender keeps the health condition out of the "critical" level.

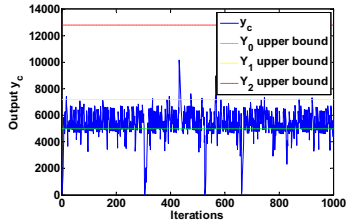

Fig. 5 Evolution of the output.

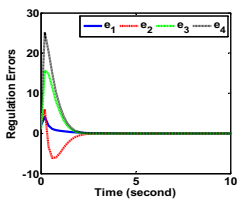

Fig. 6 Regulation errors.

Fig. 6 shows the simulation results of the regulation errors for the physical plant. Since the packet losses and delay are small enough, one can see that the regulation errors converge to zero therefore the closed-loop system is stable.

3) Scenario II: the defender selects a random policy

In this scenario, the defense action is chosen at random. Consequently, in some cases the attacker manages to compromise the system and the cyber states exceeds the limit as verified in Fig. 7 in which the network delay is plotted. Because of the large delay and packet loss, the system becomes unstable. Fig. 8 shows the regulation errors in scenario II in which the regulation errors do not converge.

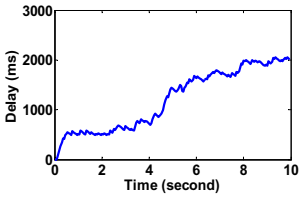

Fig. 7 Delay in Case II.

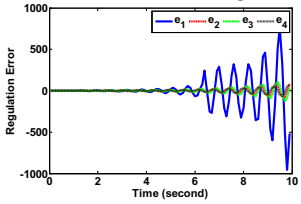

Fig. 8 Regulation errors in Case II.

Therefore, this simulation result verifies that the decisions made on the cyber system influence the stability of the physical system. The physical system can be stabilized when the cyber defender applies the optimal defense policy. If the cyber states become abnormal such that the packet losses and the delay are large enough, actions needs to be taken on the cyber side to bring the states back to normal otherwise the physical system has to be halted to avoid further damages.

VI. Conclusions and Future Work

This paper proposed a comprehensive representation which is capable of capturing the interrelationship between the physical and cyber systems. As a result, the states in the cyber system affect the controller design for the physical systems and vice versa. Making use of this representation, the optimal defense and attack policies are derived which yield the largest payoff. An optimal controller is revisited to stabilize the physical plant with the presence of uncertainties brought by the cyber stats. Since the proposed scheme is in a general form, it can be applied in a variety of industrial applications including autonomous systems. As future work, we consider analyzing the impact of various attacks on the network performance.

REFERENCES

[1] H. Baumman and W. Sandmann, Markovian Modeling and Security Measure Analysis for Networks under Flooding DoS Attacks, Distributed and Network-Based Processing (PDP), 2012 20th Euromicro International Conference on, pp. 298-302, 2012.

[2] Q. Zhu and T. Basar, Robust and Resilient Control Design for Cyber-Physical Systems with an Application to Power Systems, In Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, pp. 4066-4071, 2011.

[3] C. Ten, G. Manimaran, and C. Liu, Cybersecurity for Critical Infrastructures: Attack and Defense Modeling, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol. 40, no. 4, pp. 853-865, July, 2010.

[4] K. Sallhammar, B. Helvik, and S. Knapskog, Towards a Stochastic Model for Integrated Security and Dependability Evaluation, in Availability, Reliability and Security, ARES 2006, the First International Conference on, pp. 1-8, 2006.

[5] L. Liu, M. Esmalifalak, Q. Ding, V. Emesih, and Z. Han, Detecting False Data Injection Attacks on Power Grid by Sparse Optimization, Smart Grid, IEEE Transactions on, vol. 5, no. 2, pp. 612-621, 2014.

[6] C. Kwon, W. Liu, and I. Hwang, Security Analysis for Cyber-Physical Systems Against Stealthy Deception Attacks, in American Control Conference (ACC), pp. 3344-3349, 2013.

[7] A. Teixeira, S. Amin, H. Sandberg, K. Johansson, and S. Sastry, Cyber Security Analysis of State Estimators in Electric Power Systems, In Decision and Control (CDC), 49th IEEE Conference on, pp. 5991-5998, Dec. 2010.

[8] S. Amin, A. Cárdenas, and S. Sastry, Safe and secure networked control systems under denial-of-service attacks, in Hybrid System Computer Control, vol. 5469, pp. 31–45, April 2009.

[9] F. Pasqualetti, F. Dorfler, and F. Bullo, Attack Detection and Identification in Cyber-Physical Systems, Automatic Control, IEEE Transactions on, vol. 58, no. 11, pp. 2715-2729, 2013.

[10] H. Xu, S. Jagannathan, and F. Lewis, Stochastic Optimal Control of Unknown Linear Networked Control System in the Presence of Random Delays and Packet Losses, Automatica, vol. 48, pp. 1017-1030, 2012.

[11] M. Littman, Markov Games as a Framework for Multi-agent Reinforcement Learning, in Proceedings of the eleventh international conference on machine learning, vol. 157, pp. 157-163, 1994.

[12] M. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, New York, 1994.

[13] G. Cai, B. Chen, K. Peng, M. Dong, and T. Lee, Modeling and Control of the Yaw Channel of a UAV Helicopter, Industrial Electronics, IEEE Transactions on, vol. 55, no. 9, pp. 3426-3434, 2008.