Detection and Compensation of Covert Service-Degrading Intrusions in Cyber Physical Systems through Intelligent Adaptive Control

Faezeh Farivar  
Department of Mechatronics Engineering,  
Science and Research Branch, Islamic Azad University  
Tehran, Iran  
F.Farivar@srbiau.ac.ir

Soheila Barchinezhad  
School of Electrical and Computer Engineering  
University of Tehran  
Tehran, Iran  
S.Barchinezhad@ut.ac.ir

Mohammad Sayad Haghighi  
School of Electrical and Computer Engineering  
University of Tehran  
Tehran, Iran  
(corresponding author: sayad@ut.ac.ir)

Aireza Jolfaei  
School of Engineering and Information Technology  
Federation University  
Melbourne, Australia  
A.Jolfaei@federation.edu.au

Abstract—Cyber-Physical Systems (CPS) are playing important roles in the critical infrastructure now. A prominent family of CPSs are networked control systems in which the control and feedback signals are carried over computer networks like the Internet. Communication over insecure networks make system vulnerable to cyber attacks. In this article, we design an intrusion detection and compensation framework based on system/plant identification to fight covert attacks. We collect error statistics of the output estimation during the learning phase of system operation and after that, monitor the system behavior to see if it significantly deviates from the expected outputs. A compensating controller is further designed to intervene and replace the classic controller once the attack is detected. The proposed model is tested on a DC motor as the plant and is put against a deception signal amplification attack over the forward link. Simulation results show that the detection algorithm well detects the intrusion and the compensator is also successful in alleviating the attack effects.

Keywords—Cyber physical system, intrusion detection, covert, service degradation attack, cyber security, system identification.

I. INTRODUCTION

Cyber Physical Systems (CPS) are known to be an integral part of the future Industry 4.0. They are a composition of cyber and physical entities and mostly use computer networks to control and interconnect the physical components. There are many CPS examples ranging from autonomous vehicles [1,2] to electric power grids [3], medical robots [4], and water distribution networks [5].

The inclusion network in the CPS paradigm introduces new forms of threat as it brings its inherent vulnerabilities along. Attacks on cyber physical systems could cause damages, and in the case of Industry 4.0, even create a cascade of failures which further boost the attack effect. A prominent category of the CPS family is network control systems. The advantages of using internet to connect controllers and physical plants has increased number of this kind of CPSs. However, remote controlling and remote sensing make system vulnerable to the threats in cyber domain, especially when ad hoc or unprotected networks are used for this purpose [6].

Numerous cyber incidents in safety-critical infrastructures have been reported in the last decade which drew the attention of researchers in the CPS field. The disaster caused by Stuxnet in 2010 [7], the power plant shutdown by cyber attack in 2008 [8], and Georgia water treatment plant incident in 2013 [9] are just a few examples.

The new cyber-physical systems suffer from some vulnerabilities which traditional control systems do not. Therefore, appropriate detection and compensation techniques are needed. Classic Intrusion Detection Systems (IDS) were developed to detect anomalies in the cyber world, and usually over the network channel. They try to single out malicious packets using a set of distinguishing features. However, in CPS, a packet can carry a control signal which does not necessarily look harmful unless the system state is also included in the analysis.

The authors in [27] give a taxonomy of CPS threats and define a class which is so called the “covert” attacks. In one of such attacks, the control signal is intercepted over the forward link and manipulated by the adversary to create either human-invisible fluctuations in the plant operation or unnoticeable steady-state errors. Both of them target damaging the plant through Service Degradation (SD), one in a short period and the other over the long run. SD attack can be made more intelligent if adversary identifies controller and/or plant.

This paper proposes a novel Intrusion Detection System (IDS) which adopts the same system identification technique a smart adversary does, but to detect malicious interventions. First, during the initial immune phase of network operation, a Neural Network (NN) identifier is trained based on plant inputs and outputs. Then, it is used to predict the plant outputs for every control signal the controller generates. If the plant’s mathematical model is available, it can be equally used for the prediction purpose. This identifier will be a virtual model of the original system. The IDS module sits at the controller side and collects error samples during the normal operation of the system. These samples specify normal deviation patterns of the estimated outputs from the actual outputs.

After the immune/learning phase finishes, any unusual behavior which results in a significant deviation from the expected output raises the alarm. A compensating controller is then triggered by the IDS to get in the loop and intervene in order to mitigate the attack effect. We mainly focus on the stealth or covert attacks which aim to create overshoots [27].

In order to evaluate the efficacy of the proposed method, a DC motor is picked as the test plant and some simulations were conducted using Simulink of Matlab and TrueTime tools [10]. The results demonstrate that proposed method
successfully detects covert amplification attacks. The compensator, which is an intelligent learning controller, also manages to reduce the impact of attacks compared to when there was no compensation involved.

The rest of this paper is organized as follows: The related works are given in Section 2. Section 3 gives a picture of the problem, especially from the cyber perspective. The proposed intrusion detection and compensation algorithms are presented in Section 4. Simulation results on the test CPS are reported in Section 5. The conclusion is made in Section 6.

II. RELATED WORK

The proliferation of cyber-physical attacks became a reality and security of them has received an increasing attention in the literature, especially after Stuxnet [6]. In this section, some works related to this subject are presented.

General approaches have been proposed to detect intrusions in cyber-physical or IoT systems. For instance, in [18], an IDS, by using neural networks, was proposed to enhance the security of vehicular networks.

The authors in [19] presented a specification-based IDS for Home Area Networks (HAN). Their IDS targeted ZigBee technology since ZigBee is dominant in HANs. Normal behavior of the network was defined through selected specifications and deviations from the defined normal behavior was deemed to be a sign of intrusion. In a similar study, the authors in [20] developed an algorithm that can monitor power flows and detect anomalies. Their algorithm uses principal component analysis to separate regular and irregular flow data. Analysis of the information in this subspace determines whether the power system data has been compromised or not.

False data injection is a kind of deception attack that is launched as a man-in-the-middle (MITM) attack. Respectively, in [21] and [22], false data injection attacks in electric power grids and wireless sensor networks are investigated. These studies describe a kind of attacker on CPSs that has information about both the physical system and the controller, which can potentially make them covert. Two queuing models are proposed to simulate the stochastic process of packet delay and Denial of Service (DoS) attacks in [23], while in [24], the authors have studied a scenario in which the attacker performs zero-dynamics attacks on the system. They tried to reveal stealth attacks for linear time invariant systems via monitoring.

With a different approach, [25] presents three mechanisms for time-based intrusion detection in CPSs. It is suggested that the information obtained by timing analysis is used for intrusion detection. Bound checking of execution micro-timings is adopted by application(s) to detect intrusions as a self test procedure. Alternatively, the IDS can be implemented on the embedded system OS scheduler.

In an effort to classify the methods of securing cyber physical systems, the authors of [26] defined general frameworks and approaches that can be adopted to make CPSs more survivable. Among them there were solutions for robust networked control systems and fault tolerant control methods. However, this paper does not study any specific plant or attack scenario and merely provides general design guidelines.

A covert attack for the purpose of service degradation was proposed in [27]. The goal was to investigate how these attacks decrease the performance of networked control systems. The authors claimed that the two attacks they designed were able to affect the system hardware, in a covert way. In their scheme, the attacker identifies both the plant and

the controller first. Then, he determines what actions he can take to degrade or damage the system, either in a short period or over the long run. The short-time attack was designed to create a 50% overshoot on the system output and the long-term service degradation attack aimed creating a noticeable steady state error. Both attacks targeted damaging the plant hardware, but in different ways. The idea was tested on an unprotected networked DC motor.

III. THE CYBER ATTACK MODEL AND ASSUMPTIONS

We assume having a networked control system with possible deception (unnoticed modification) attacks on the forward link packets, similar to [27]. This implies that the control signals are not protected well or the forward link key has been compromised [29]. The attacker aims to manipulate the control packets at some point of time. However, it is assumed that the system starts from a safe phase during which it is attack-free. This is a necessity for IDS error learning and identification purpose. We consider the amplification Service Degradation (SD) attack. This is a covert attack in which the adversary intercepts the network channel, modifies it and sends it through towards the rightful recipient. The manipulation is normally multiplication of the control signals by a fixed number.

Reference [27] defines the intelligent version of this attack in which the attacker identifies the plant and the controller first and then applies a constant proportional gain in the forward link to create temporal overshoots as high as %50 of the nominal reference. Through repeated tries, the attacker aims to degrade the service and gradually damage the physical system or reduce the mean time between failures (MTBF). The difference between a DoS attack and a SD-Controlled attack is that the latter is not intended to disrupt the physical process in a short period.

In the subsequent sections, we will study the effect of this kind of intelligent attacks on the performance of a CPS. First, we design an intrusion detection system to capture such covert attacks. Under this adversarial model, it is assumed that the feedback link is safe. Therefore, when legitimate state changes happen at the plant side, the IDS will know and uses the corresponding reference error model learnt in the initial phase.

IV. INTRUSION DETECTION AND COMPENSATION

A. Intrusion Detection System

The most important issues in dealing with CPS attacks are accuracy in detection (low false positives) and timely reactions. The time between detection and reaction is critical to have a sustainable process. Early detections increase the chance of containing the attack. Fig. 1 shows the architecture.

![Fig. 1. The proposed IDS architecture for the networked control system.](image)
Detector. To design the IDS, it is assumed that there exists an initial secure period for the CPS during which an identifier learns the plant dynamics (e.g. the equivalent transfer function). Attacks are (potentially) launched after $T_{\text{attack}}$.

We design an intelligent identifier to model the CPS while it works normally in the safe phase. The identifier is an artificial neural network. The detection strategy proposed in this study is based on adaptive hard thresholds. IDS makes a decision based on the modelling error plus network sample-and-hold quantization and jitter errors ($Y_{\text{out}} - Y_{\text{in}}$). The detector gathers error data from the plant in each operation state during the secure operation phase. Assuming that the sum of errors have a Gaussian distribution in the system state $\psi_i$, the detection thresholds are defined to be $\mu_i \pm K \sigma_i$ where $K$ is a constant and $\mu_i$ and $\sigma_i$ are the Gaussian mean and standard deviation in $\psi_i$, respectively. Therefore, the false positive probability will roughly be $Q(K) = \text{erf} \left( \frac{K}{\sqrt{2}} \right)$. In a trade-off with false negatives, $K$ should be chosen so that the physical system operation is not affected much if a small attack is missed while significant deviations from the expected normal working condition ($Y_{\text{in}}$) are captured.

Compensator. In our solution, the compensator is a robust controller and gets involved when IDS triggers a signal indicating an attack. To have a stable plant (system), using a compensator is necessary. For the attacks of signal amplification kind, we need to damp the accuracy and $\theta \left( t \right) = \theta \left( t \right) - \theta \left( t \right)$ can be obtained as:

$$e(t) = \theta_{\text{ref}} - \theta$$  \hspace{1cm} (1)

where $\theta_{\text{ref}}$ is the desired speed of the speed and $\dot{\theta}$ is the speed of the DC motor.

The sliding surfaces are defined as in [29]:

$$S(t) = \left( \frac{d}{dt} + \delta \right)^{n-1} e(t)$$  \hspace{1cm} (2)

where $S(t) \in R$ and $\delta$ is a real positive constant parameter.

In case the physical plant is a DC motor, one can define the sliding surface as follows:

$$S(t) = \dot{e}(t) + \lambda e(t)$$  \hspace{1cm} (3)

where $\lambda$ is a real positive constant parameter and $e(t)$ stands for the difference between the desired and current rotational speeds (rpm). The RBF-NN controller is designed to establish a robust control system that adapts using SMC’s stability theory. It guarantees the existence of the sliding surface which converges to $S(t) = 0$.

RBF Neural Network. RBF-NN can be considered as one layer feed forward neural network with nonlinear elements. The RBF-NN does the mapping according to:

$$f(z) = \sum_{j=1}^{n} w_j G_j(z_j, m_j, \sigma_j)$$  \hspace{1cm} (4)

where $z = [z_1, z_2, \ldots, z_n]^T \in \mathbb{R}^n$ is the input, $G_j(z_j, m_j, \sigma_j) \in \mathbb{R}^n$, $j = 1, 2, \ldots, n$ are the Gaussian radial basis functions, $\sigma_j \in \mathbb{R}$ is the spread of the the $j$th Gaussian function, $m_j$ is the mean value of that function and $n$ is the number of neurons. Each Gaussian radial basis function can be written as:

$$G_j(z_j, c_j, \sigma_j) = \exp \left[ \frac{(z_j - c_j)^2}{2\sigma_j^2} \right]$$  \hspace{1cm} (5)

Notice that the optimal values may not be unique. In this study, $c$ and $\sigma$ are not trained. The input vector $z$ is $[S_r, S_t]$ and the RBF-NN output ($f$) is considered as the control input ($u$) sent to the plant via network.

Learning Algorithm. The RBF-NN parameters are adjusted using SMC Lyapunov stability theory. Let us define the Lyapunov function as:

$$V(t) = \frac{1}{2} S^2(t)$$  \hspace{1cm} (6)

Differentiating Eq. (6) with respect to time, we have:

$$\dot{V}(t) = S(t) \dot{S}(t)$$  \hspace{1cm} (7)

According to Lyapunov stability theory, the control system will asymptotically stabilize the error dynamics given in Eq. (3) and $S(t) \rightarrow 0$ if Eq. (7) is strictly negative. Applying an appropriate control input leads to achieving the mentioned condition. Let us define Eq. (7) as a cost function of RBF-NN controller. Hence, the cost function of RBF-NN controller is defined as follows:

$$E(t) = S(t) \dot{S}(t)$$  \hspace{1cm} (8)

Using Back-Propagation (BP) algorithm, the weighting vector of the RBF-NN is adjusted such that the cost function takes bigger negative values. This algorithm is written briefly as:

$$w_{\text{RBF}}(t + 1) = w_{\text{RBF}}(t) - \eta \frac{\partial E(t)}{\partial w_{\text{RBF}}(t)}$$  \hspace{1cm} (9)

where $\eta$ and $w$ represent the learning rate and tuning parameter of RBF-NNs, respectively. The gradient of $E$ with respect to the weighting vector $w$ can be obtained as:
\[
\frac{\partial E(t)}{\partial \omega_{\text{RBF}}(t)} = \frac{\partial}{\partial \omega_{\text{RBF}}(t)} \left( S(t) \frac{\partial S(t)}{\partial \omega_{\text{RBF}}(t)} \right) = S(t) \frac{\partial S(t)}{\partial \omega_{\text{RBF}}(t)} \tag{10}
\]

Eq. (10) can be rewritten as follows:
\[
\frac{\partial E(t)}{\partial \omega_{\text{RBF}}(t)} = S(t) \frac{\partial \tilde{S}(t)}{\partial \omega_{\text{RBF}}(t)} = S(t) \frac{\partial S(t)}{\partial \omega_{\text{RBF}}(t)} \tag{11}
\]

Using Eq. (5), (2) and (3), Eq. (11) is simplified as:
\[
\frac{\partial E(t)}{\partial \omega_{\text{RBF}}(t)} = S(t) G_{\text{RBF}}(z(t),c,\sigma) \tag{12}
\]

in which \( G \) is the value given in Eq. (5). Therefore, the adaptation law of RBF-NN is obtained as:
\[
w_{\text{RBF}}(t+1) = w_{\text{RBF}}(t) + \eta SG_{\text{RBF}}(z(t),c,\sigma) \tag{13}
\]

where \( \eta \) is the learning rate of the RBF-NN controller.

V. MODEL EVALUATION & SIMULATION RESULTS

To evaluate the proposed SD-controller, we will employ it in a test system whose plant is a DC motor. The speed control of the DC motor will be carried out through a wireless network similar to Fig. 1. We will dedicate the next subsection to the introduction of this plant.

A. Networked Speed Control of DC Motor

A DC motor is a popular actuator in control systems. The electric circuit of the armature and the free-body diagram of the rotor is depicted in Fig. 2. The voltage source \( V \) is the system or control input applied to the armature of the motor. The rotational speed of the shaft \( \dot{\theta} \) is the system output. It is assumed that the rotor and shaft are rigid. Also, it is assumed that the friction torque is proportional to shaft angular velocity. The system parameters are presented in Table. 1. The torque that a DC motor provides is proportional to the armature current and the strength of the magnetic field. In the model, it is assumed that the magnetic field is constant. Hence, the motor torque is only proportional to the armature current \( i \) by a constant parameter \( k_t \) as shown below:
\[
T = k_t i \tag{14}
\]

The back emf \( v_e \) is proportional to the angular velocity of the shaft by a constant parameter of \( K_e \) as follows:
\[
v_e = k_e \dot{\theta} \tag{15}
\]

Dynamic equations of the system using Newton’s law and Kirchoff’s law are as below,
\[
\frac{d^2 \dot{\theta}}{dt^2} = \frac{1}{J} (k_t i - b \frac{d \dot{\theta}}{dt}) \tag{16}
\]
\[
\frac{di}{dt} = \frac{1}{L} (-R_e + V - k_e \frac{d \dot{\theta}}{dt}) \tag{17}
\]

In the model, the motor torque and back emf constants are equal. Therefore, \( k_t \) represents both constants in Table. 1. The state space model of the DC motor according to Eq. (16) and Eq. (17) are presented as follows,

![Fig. 2. The electronic circuit of the DC motor.](image)

\[
\begin{bmatrix} \frac{d}{dt} \dot{\theta} \\ \frac{d}{dt} i \end{bmatrix} = \begin{bmatrix} -\frac{b}{J} & \frac{1}{J} \\ \frac{-k_e}{L} & \frac{-R_e}{L} \end{bmatrix} \begin{bmatrix} \dot{\theta} \\ i \end{bmatrix} + \begin{bmatrix} 0 \\ 1/L \end{bmatrix} V 
\]

\[
y = [1 \ 0] \begin{bmatrix} \dot{\theta} \\ i \end{bmatrix} \tag{18}
\]

The transfer function of the system is obtained as follows,
\[
H(s) = \frac{\dot{\theta}(s)}{V(s)} = \frac{2}{s+9.997}(s+2.003) \tag{19}
\]

Remark. To calculate Eq. (11), there is a term which depends on the system equations, which for DC motor becomes,
\[
\frac{\partial \dot{\theta}}{\partial i} = \frac{-K_t}{LJ} \tag{20}
\]

It is a constant value. However, in this study, we assume that the dynamic system equation is unknown. Thus, there is no information about this term and it is compensated by the learning rate \( \eta \) in Eq. (13).

B. Simulation Results

We test the intrusion detection and compensation system of ours on the DC motor described in the previous subsection. The simulator is MATLAB Simulink. We pick the Controlled Area Network (CAN) standard used in vehicles (including electric cars) to connect the controller and the plant. A modern automobile has as many as 70 Electric Control Units (ECU), some of which control transmission and electric power steering. The network rate is set to 240Kbps and the frame/message size is set to 80bits. This implies that the average rate of sending control signals as well as reading samples from the output sensor is 3000/s.

To control the speed of the DC motor under normal conditions, a lag compensator is designed such that the system output tracks a step function. The designed PID controller is as follows,
\[
C(s) = 30 \frac{s+1}{s+0.01} \tag{21}
\]

Similar to [27], a covert SD attack of gain-multiplication kind is launched on the forward link. However, instead of merely targeting an overshoot of 50%, we tried attacking the system with a range of gains, including one that could create

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J )</td>
<td>inertia moment of the rotor</td>
<td>0.01 kg.m²</td>
</tr>
<tr>
<td>( b )</td>
<td>motor viscous friction constant</td>
<td>0.1 N.m.s</td>
</tr>
<tr>
<td>( k_e )</td>
<td>electromotive force constant</td>
<td>0.01 V/rad/sec</td>
</tr>
<tr>
<td>( k_t )</td>
<td>motor torque constant</td>
<td>0.01 N.m/Amp</td>
</tr>
<tr>
<td>( R )</td>
<td>electric resistance</td>
<td>1 Ohm</td>
</tr>
<tr>
<td>( L )</td>
<td>electric inductance</td>
<td>0.5 H</td>
</tr>
</tbody>
</table>
such an overshoot. In the experiments, $\sigma$ is set to 5 and all the attacks are launched at $t = 5$. Fig. 3 shows the output of the system and the control signal in two scenarios. Each plot shows the case in which the IDS and compensator were not in place as well as the case they were in the circuit. In the normal operation, the motor speed rises sharply and gradually converges to one (i.e. the nominal speed) when a step like reference signal with the value of one is applied.
Fig. 3. SD attack with a [gain of attack=0.5] applied on two systems, one with IDS and one without. The top plot shows the system output as well as the IDS flag, and the bottom plot shows the corresponding control signals.

Fig. 4. SD attack with a [gain of attack=5] applied on two systems, one with IDS and one without. The top plot shows the system output as well as the IDS flag, and the bottom plot shows the corresponding control signals.

Fig. 5. SD attack with a [gain of attack=15] applied on two systems, one with IDS and one without. The top plot shows the system output as well as the IDS flag, and the bottom plot shows the corresponding control signals.

Fig. 6. SD attack with a [gain of attack=120] applied on two systems, one with IDS and one without. The top plot shows the system output as well as the IDS flag, and the bottom plot shows the corresponding control signals.
When there is no IDS or compensator, the attacker manages to bring the rotor speed (temporarily) down by 20% when he applies a gain of 0.5 on the forward link signals. This value is less with the IDS and compensator involved. Note that IDS has correctly detected the anomaly in the motor speed at $t = 5.07s$ and switched the controllers. The output gain of 0.5 for the IDS flag is merely for better demonstration of its output. Fig. 4 shows a similar attack but with a gain of 5. This time we have an overshoot of around 30% in the motor speed. However, with IDS and the subsequent compensation, it was reduced to less than 10%. Similarly, in Fig. 5, there was an overshoot of 50% when the attacker applied a gain of 15 on an unprotected system. But it was mitigated to a value as low as 17% in a setup with IDS and compensator.

The lower plots show the corresponding control signals in each case. As it can be seen, sometimes compensation of the attack required very high and very low spike-like reactions in the control/input voltage. This is unavoidable if the attack is as huge as the ones tested here. However, in practice, there might be a limit on the input range of the DC motor which limits the range of compensation.

Fig. 6 reports an interesting phenomenon. The excessive gain the attacker has applied (i.e. 120) made the system completely unstable. This is obvious from the oscillating output of the system. Such attacks can cause permanent damages. However, the system with IDS and compensator managed to keep the system stable. The overshoot is less than 100%, which is better than having an unstable system.

VI. CONCLUSION

In this paper, we studied the problem of detecting stealth deception intrusions in cyber physical systems. We proposed that an identifier learns the plant and predicts its outputs. An IDS compares the actual plant outputs with those of the identifier that received the same control inputs. Normal prediction error patterns are learnt during the learning phase. Anomalies in error are captured by statistical measures and prediction error patterns are learnt during the learning phase.

An attacker intercepts and switches the controllers. The output gain of 0.5 for a setup with IDS and compensator.

REFERENCES


