Energy-Based Attack Detection in Networked Control Systems

Emeka Eyisi
United Technologies Research Center
East Hartford, CT, USA
eyisiep@utrc.utc.com

Xenofon Koutsoukos
Institute for Software Integrated Systems
EECS Department
Vanderbilt University
Nashville, TN, USA
xenofon.koutsoukos@vanderbilt.edu

ABSTRACT

The increased prevalence of attacks on Cyber-Physical Systems (CPS) as well as the safety-critical nature of these systems, has resulted in increased concerns regarding the security of CPS. In an effort towards the security of CPS, we consider the detection of attacks based on the fundamental notion of a system's energy. We propose a discrete-time Energy-Based Attack Detection mechanism for networked cyber-physical systems that are dissipative or passive in nature. We present analytical results to show that the detection mechanism is effective in detecting a class of attack models in networked control systems (NCS). Finally, using simulations we illustrate the effectiveness of the proposed approach in detecting attacks.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection (e.g., firewalls)*; H.1.1 [**Models and Principles**]: Systems and Information Theory—*General Systems Theory*

General Terms

Algorithms; Design; Security; Theory

Keywords

Energy-based detection; Networked Control Systems; Attacks

1. INTRODUCTION

The increased autonomy of CPS, together with the introduction of communication networks, has increased the security vulnerabilities of CPS infrastructure to malicious cyber attacks. Within the past few years, there has been a surge in attacks on CPS infrastructures. This increased prevalence of attacks has resulted in increased concerns regarding the security of these systems. Due to the safety-critical nature of CPS, failure or disruption of normal operation can potentially lead to serious harm to the physical system under control and to the people and other infrastructures that depend on it. Hence, securing these systems in order to ensure resilient operation is of utmost importance. Some of the well-known examples of attacks on CPS include the W32.Stuxnet worm attack that maliciously infected an Iranian Nuclear facility, taking control and heavily disrupting its normal operation according to the attacker's design [6], the cyber attacks on power transmission networks operated by Supervisory Control and Data Acquisition (SCADA) Systems [12], as well as attacks that infiltrated critical systems including medical devices [13] and waste water treatment plants [1].

In securing CPS infrastructures, the reliable detection of attacks is very important and also fundamental to the design of compensation and reconfiguration mechanisms for mitigating the impact of attacks. The presence of the network increases the complexity of the detection of attacks. Hence, effective and yet efficient novel approaches are needed to enable the early detection of attacks in CPS. A majority of the existing detection approaches are typically from the cyber-security community. As highlighted in [3], the traditional approach often used in information/cyber-security neglects the knowledge of the physical process under control in the detection of attacks. Contrary to the traditional cyber-security approach, newer approaches in the CPS community, instead of creating models of network traffic or software behavior, leverage the knowledge of the physical process in designing effective mechanisms in order to facilitate the detection of attacks. The idea is that by understanding the interactions of the control system with the physical world, it would be possible to develop systematic frameworks to detect attacks and secure CPS in general.

In this work, we utilize the energy of physical systems in order to define precise detectability conditions for certain attack models and vulnerabilities. The concept of energy is very important in the behavior of dynamical systems. Compared to traditional detection approaches such as observer-based detection [14], there are only a handful of work whereby the concept of energy or passivity is used in model-based detection. In [7], the authors proposed a fault detection and isolation method for port-Hamiltonian systems to detect variations in the parameters of system components. The work in [4] proposed an energy balance scheme for fault detection for continuous-time passive systems. The author performed fault detection by checking when the energy balance is perturbed indicating the presence of faults. An energy balance fault detection approach was also applied for sensor fault detection in steel galvanizing process [16]. In [18], a passivity-based fault detection method was introduced based on evaluating the traditional passivity-based inequality. In this work, a fault is said to have occurred whenever the inequality is not satisfied. Most of these works in energy-based detection have been focused on reliability as it pertains to the protection of physical components against faults. Additionally, existing work does not consider the introduction of a communication

HiCoNS'14, April 15–17, 2014, Berlin, Germany.
Copyright 2014 ACM 978-1-4503-2652-0/14/04 ...$15.00.
http://dx.doi.org/10.1145/2566468.2566472 .

network and do not address the detection of intentional malicious cyber attacks against CPS.

Using the intuitive notion of energy, we propose an attack detection mechanism for CPS. The proposed approach is complementary to other detection mechanisms such as observer-based detection. The underlying idea is that the presence of attacks disturbs the energy balance of the physical system by dissipating or injecting additional energy. We define the notion of detectability of an attack by its effect on system's energy. The proposed approach provides the additional benefit of detecting when and to what magnitude a system's energy property is impacted due to the occurrence of an attack. In particular, we focus on dissipative CPS, which include a large class of existing systems. We present the use of energy-balance in the detection of attacks in NCS. We present a general characterization of attacks on the energy of a dynamical system. In addition, we demonstrate the impact of specific attack models on the stability guarantees of NCS. Finally, we demonstrate our approach using a case study on the velocity tracking control of a single joint of a robotic arm over a network.

The rest of the paper is organized as follows: Section 2 presents a brief background and underlying definitions used in paper. Section 3 presents the networked control system model, the attack models and formally states the problem that is addressed in this paper. The energy-based attack detection approach, the analytical results on the detection mechanism and the characterization of passive and non-passive attacks are presented in Section 4. Section 5 presents an example case study using simulations to evaluate the proposed approach. The paper is concluded in Section 6.

2. BACKGROUND

We define some fundamental concepts which are important in the description of the proposed approach.

DEFINITION 1. *A dynamic Linear Time-invariant (LTI) system, $\mathcal{H}$, is minimal if it is both controllable and observable.*

The notion of dissipativity and passivity of a system presented in this work follow the behavior-based approach given by Willems in [17] which involves associating the system to a non-negative definite storage function $V(x)$ and a supply function, W. We provide the following definition of dissipativity.

DEFINITION 2. *[2] [9] A discrete-time system, $\mathcal{H}$, is said to be* ***dissipative*** *with respect to the* ***supply function*** *$W(u(k), y(k))$ if there exists a positive definite function $V(x_k)$ or V_k, called* ***storage function****, satisfying $V(0) = 0$ such that $\forall x_0 \in X$, $\forall k \geq k_0$, and all $u \in \mathbb{R}^n$ and with, $V_k = \frac{1}{2}x_k^T P x_k$*

$$V_{k+1} - V_0 \leq \sum_{k=0}^{N-1} W(u(k), y(k)) \qquad (1)$$

DEFINITION 3. *[10] [17] A dynamic system, $\mathcal{H}$, is said to be* ***QSR-dissipative*** *if it is dissipative with respect to the supply rate,W given as*

$$W(u, y) = y^T Q y + 2y^T S u + u^T R u \qquad (2)$$

where Q, S, R are matrices of appropriate dimensions with Q and R symmetric. By choosing different values for Q, S, R, special cases of dissipativeness can be derived [8]. Special cases of QSR dissipative systems are as follows. If the system $\mathcal{H}$ is QSR-dissipative then it is

1. Passive *if Q=0, S=$\frac{1}{2}I$, $R = 0$*
2. Strictly input passive (SIP) *if $Q = 0$, $S = \frac{1}{2}I$, $R = -\delta I$*
3. Strictly output passive (SOP) *if $Q = -\epsilon I$, S=$\frac{1}{2}I$, $R = 0$*
4. Very strictly passive (VSP) *if Q=$-\epsilon I$, $S = 0$, R=$-\delta I$*

where ϵ and δ are positive scalars.

LEMMA 1. *[8][**Generalized Positive Real Lemma**] Let $G(z)$ be a transfer function description and $M(z) = R + G^H(z)S + S^T G(z) + G^H(z)QG(z)$, with $G^H(z)$ denoting the hermitian transpose of G(z). Let $\mathcal{H}$ be a minimal realization of G(z). Then $\forall z$ s.t. $\|z\| \geq 1$, $M(z) \geq 0$ if and only if there exist a real symmetric positive definite matrix P and real matrices L and W such that*

$$A^T PA - P = C^T QC - L^T L \qquad (3)$$

$$A^T PB = C^T QD + C^T S - LW \qquad (4)$$

$$B^T PB = R + D^T S + S^T D + D^T QD + C^T S - W^T W \qquad (5)$$

3. SYSTEM MODEL AND PROBLEM

In this section, we describe the components of networked control system and the attack models considered in this work. Subsequently, we formulate the attack detection problem and describe the underlying assumptions. The notations used in the following sections are standard. Let $\mathbb{R}^n$ denote the Euclidean space of dimension n, I denotes the identity matrix of appropriate dimensions. For a matrix $P \in \mathbb{R}^{nxn}$, its transpose is denoted by P^T. For a symmetric matrix, P, where $P = P^T$, $\quad P > 0$ denotes it is positive definite.

3.1 Networked Control System Model

We consider a networked control system as depicted in Figure 1. The main components of the NCS are the physical plant, the controller, the wave transformation (a static local controller), and the communication network. The data exchange between the plant and the controller is done over a communication network.

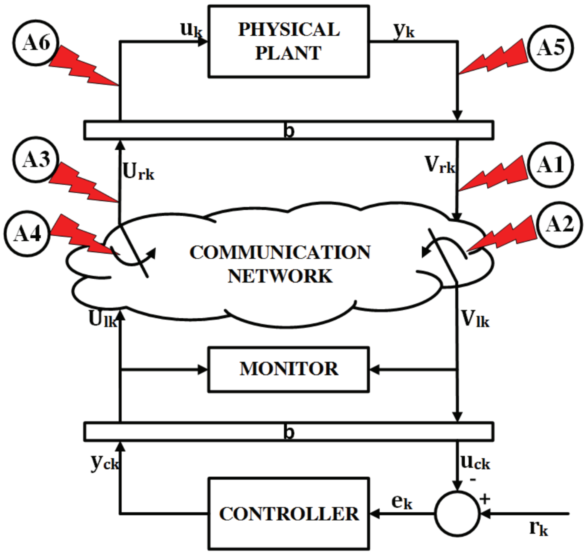


Figure 1: Networked Control System

(a) Physical Plant Model: We model the physical plant as a discrete linear time-invariant system. This version of the plant neglects system nonlinearities and presence of noise in the dynamics and measurement signals. We consider the physical plant which can be represented in the state space form as follows:

$$\mathcal{H}_p : \begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k + Du_k \end{cases} \qquad (6)$$

where $x_k \in \mathcal{X}$ represents the state variables, $u_k \in \mathcal{U}$ represents the control inputs to the plant and $y_k \in \mathcal{Y}$ represents the plant outputs obtained by sensors at sampling instant $k \in \mathbb{Z}$.

(b) Controller Model: The controller modifies the behavior of the physical plant through the application of a control input command in order to achieve a desired objective or satisfy a performance requirement. The controller can be represented in discrete-time state-space form as follows:

$$\mathcal{H}_c : \begin{cases} z_{k+1} = A_c z_k + B_c e_k \\ y_{c_k} = C_c z_k + D_c e_k \end{cases} \tag{7}$$

where $z_k \in \mathbb{R}^q$ represents the controller states, y_{c_k} is the control command, $e_k = r_k - u_{c_k}$ is the error between the reference, r_k and the received plant output, u_{c_k}, with the matrices A_c, B_c, C_c and D_c of appropriate dimensions. It is assumed that the controller is designed under the nominal conditions, i.e. without attacks, to achieve the desired performance objective.

(c) Wave Transformation: In order to preserve the power content of information exchanged over a network, the sensor and control signals are transformed into wave variables which are then sent over the network. From Figure 1, the wave transformation is denoted by the blocks, **b**. The wave variables pair (U_{r_k}, V_{r_k}) on the plant side as well as the pair (U_{l_k}, V_{l_k}) on the controller side of the network can be described by the following expressions:

$$U_{r_k} = \frac{1}{\sqrt{2b}}(y_k + bu_k) \tag{8}$$

$$V_{r_k} = \frac{1}{\sqrt{2b}}(y_k - bu_k) \tag{9}$$

$$U_{l_k} = \frac{1}{\sqrt{2b}}(u_{c_k} + by_{c_k}) \tag{10}$$

$$V_{l_k} = \frac{1}{\sqrt{2b}}(u_{c_k} - by_{c_k}) \tag{11}$$

where $b \in \mathbb{R}_0^+$. From Figure 1, one can observe that under ideal network conditions, $V_{l_k} = V_{r_k}$ and $U_{r_k} = U_{l_k}$.

3.2 Attack Model

Figure 1 depicts the feasible cyber-attacks as a result of the vulnerabilities of the networked control system. While the attacks denoted as **A1**-**A4** model attacks on the information exchanged over the communication network, the attacks denoted as **A5** and **A6** models attacks on sensors and actuators respectively. Similar well-known attack types have be proposed in [15] [11]. For each attack type, $\mathcal{A}_k$, let $\mathcal{T}_a = k_s, ..., k_e$ denote the attack duration with the attack starting from k_s and ending at k_e. We consider two main classes of attacks, integrity attacks and denial-of-service attacks. These attack types are described as follows:

(a) Integrity attacks: In an integrity attack, an adversary deceives a compromised component of the NCS into believing that a received false data is valid or true. The underlying assumption is that all attacks lie within a predetermined range since attacks leading to signals that exceed such a range can be easily detected. The integrity attacks represented as **A1**, **A3**, **A5** and **A6** in Figure 1 can be further categorized into the following:

(i) Min/Max attacks: These attacks involve the adversary modifying the content of compromised signals to their respective minimum or maximum values. We model min/max attacks on the exchanged wave variables as well as the min/max attacks on the sensors and actuators. The attacks on the exchanged variables essentially exploit the vulnerabilities as a result of the communication while the attacks on the sensors and actuators exploit the vulnerabilities of the computing interfaces to these components which may or may not be colocated. We consider them separately since each component's interaction with the overall NCS is different and hence it is important to understand the impact of an attack on each component on the correct operation of the overall NCS.

(1) Min/Max attacks on exchanged wave variables: For attacks on the wave variable, V_{r_k}, sent from the plant we have,

$$\tilde{V}_{r_k}^{min} = \begin{cases} V_{r_k} & \forall k \notin \mathcal{T}_a \\ V_{r_{min}} & \forall k \in \mathcal{T}_a \end{cases} \tag{12}$$

$$\tilde{V}_{r_k}^{max} = \begin{cases} V_{r_k} & \forall k \notin \mathcal{T}_a \\ V_{r_{max}} & \forall k \in \mathcal{T}_a \end{cases} \tag{13}$$

Similar attacks can be launched against the wave variable, U_{r_k}, sent from the controller.

(2) Min/Max attacks on sensors and actuators: For attacks on the sensor signal, y_k, we have,

$$\tilde{y}_k^{min} = \begin{cases} y_k & \forall k \notin \mathcal{T}_a \\ y_{min} & \forall k \in \mathcal{T}_a \end{cases} \tag{14}$$

$$\tilde{y}_k^{max} = \begin{cases} y_k & \forall k \notin \mathcal{T}_a \\ y_{max} & \forall k \in \mathcal{T}_a \end{cases} \tag{15}$$

Similar attacks could be launched against the actuator signal, u_k.

(ii) Additive attacks: This attack involves introducing an additional offset/bias, $\alpha \neq 0$ to the actual exchanged information. We model additive attacks on the exchanged wave variables as well as additive attacks on the sensors and actuators.

(1) Additive attacks on exchanged wave variables: For attacks on the wave variable, V_{r_k}, sent from the plant we have,

$$\tilde{V}_{r_k}^{a} = \begin{cases} V_{r_k} & \forall k \notin \mathcal{T}_a \\ V_{r_k} + \alpha_k & \forall k \in \mathcal{T}_a \quad and \quad V_{r_k} + \alpha_k \in \mathcal{V} \\ V_{r_{min}} & \forall k \in \mathcal{T}_a \quad and \quad V_{r_k} + \alpha_k < V_{r_{min}} \\ V_{r_{max}} & \forall k \in \mathcal{T}_a \quad and \quad V_{r_k} + \alpha_k > V_{r_{max}} \end{cases} \tag{16}$$

(2) Additive attacks on sensors and actuators: For attacks on the sensor signal, y_k, we have,

$$\tilde{y}_k^{a} = \begin{cases} y_k & \forall k \notin \mathcal{T}_a \\ y_k + \alpha_k & \forall k \in \mathcal{T}_a \quad and \quad y_k + \alpha_k \in \mathcal{Y} \\ y_{min} & \forall k \in \mathcal{T}_a \quad and \quad y_k + \alpha_k < y_{min} \\ y_{max} & \forall k \in \mathcal{T}_a \quad and \quad y_k + \alpha_k > y_{max} \end{cases} \tag{17}$$

(iii) Min/Max energy attacks: Considering that the proposed approach is based on energy, an attacker's objective could be to apply the largest impact damage on the system based on the knowledge of the system's energy. We model two types of energy-based attacks based on their intended impact on the system.

(1) Max energy attack: In this case, we model attacks that attempt to dissipate maximum amount of energy i.e. the energy of the system becomes positive. This type of attack can be seen as an attacker's attempt to degrade system performance without destabilizing the system in regards to energy. In this attack type, for each time step, the attacker chooses a value for the compromised wave variable such that the total dissipated energy is maximized without exceeding the predetermined limits of the wave variable. The max energy attack can be captured as follows:

$$\begin{aligned} \underset{V_{r_k}}{\text{maximize}} \quad & E_T \\ \text{subject to} \quad & V_{r_k} \in [V_{r_{min}}, V_{r_{max}}] \end{aligned}$$

(2) Min Energy Attack: Similar, to the max energy attack, in this case we model attacks that attempts to inject the largest amount of

energy which from the system's perspective portrays the system as generating additional energy i.e. the energy of the system becomes negative. This attack type can be seen as an attacker's attempt to both degrade the performance of the system and potentially destabilize the system. In the model of this attack, at each time step the attacker chooses the compromised wave variable such that the energy is minimized without exceeding the predetermined limits of the wave variable. The min energy attack can be captured as follows:

$$\begin{aligned} &\underset{V_{r_k}}{\text{minimize}} && E_T \\ &\text{subject to} && V_{r_k} \in [V_{r_{min}}, V_{r_{max}}] \end{aligned}$$

(b) Denial-of-Service (DoS) attacks: DoS attacks, denoted as **A2** and **A4** in Figure 1, prevent signals from reaching the intended destination. In NCS, it involves the disruption of the availability of information exchanged between the plant and the controller. DoS attacks are typically carried out by jamming the communication channel, changing the routing protocol or saturating the receiver with useless signals. The attacker's main objective is usually to degrade the performance of the NCS as well as to potentially destabilize the physical system. The DoS attack can be modeled as a form of the additive attack as follows:

$$\tilde{V}_{r_k}^{DoS} = V_{r_k} + \alpha V_{r_k} \quad \begin{cases} \alpha = 0 & \forall k \notin \mathcal{T}_a \\ \alpha = -1 & \forall k \in \mathcal{T}_a \end{cases} \tag{18}$$

3.3 Problem Statement

Consider the networked control system as shown in Figure 1, under possible cyber attacks as indicated by the attacks **A1**-**A6** due to the vulnerabilities of NCS. We define what is meant by an energy-based monitor and detectability of attacks in this framework.

DEFINITION 4. *An energy-based monitor is a deterministic algorithm, $\Phi : \Lambda \mapsto \Psi$, with knowledge of the plant dynamics and access to discrete-time measurements and control inputs. The output of a monitor is $\Psi = \{\psi_1, \psi_2\}$, with $\psi_1 \in \{True, False\}$, and $\psi_2 \in$ {Passive, Non-Passive}*

DEFINITION 5. *An attack is detectable if in the presence of the attack, $\mathcal{A}_k$, ψ_1 =True and ψ_2 = Passive or Non-Passive.*

The following problem is of interest:
{**1. Detection Problem**} *Design an algorithm, Φ, for an energy-based monitor which can quantify or estimate the energy of the system, E_T, such that in the presence of an attack and with the knowledge of the plant, the controller and exchanged wave variables the following holds:*

$$\Psi = \{\psi_1, \psi_2\} = \begin{cases} \{\text{True, Passive}\} & \forall E_T > 0 \\ \{\text{True, Non-Passive}\} & \forall E_T < 0\} \end{cases} \tag{19}$$

In the following sections, we propose a solution to the above problem. We assume the following about the NCS.
Assumption 1: The plant and controller are dissipative by design, both with a sampling period, T$_s$. The assumption of dissipativity for both the plant and controller is to ensure stability guarantees in the nominal case.
Assumption 2: The components of the NCS including the physical plant, the sensor, actuator, controller and attack monitor are time-synchronized. This ensures that all the components of the NCS are progressing in lock step in regards to time.
Assumption 3: Whenever the input buffers are empty, null packets are processed. This assumption is used to preserve passivity in the nominal sense in order to avert the typical hold-last sample approach which is known to be non-passive. Other approaches for handling missed packets can be sought in this case as well with no loss of generality.
Assumption 4: It is assumed that the controller and monitoring system for the plant are co-located together. The idea is that the controller is assumed to be trustworthy while the plant's trustworthiness is not known or guaranteed. In the case that the trustworthiness of the controller is not known or guaranteed, an additional monitor can be co-located with the plant.
Assumption 5: The attacker has knowledge of the plant and controller. In this assumption, we consider that the attacker is smart in the sense that he/she can attempt to use knowledge of the system to introduce attacks that cannot be easily detected with a simple bad data detector.
Assumption 6: For our initial analysis, we assume an ideal communication network, hence do not consider the usual communication network effects such as time-delays and packet losses but rather we focus on malicious attacks on the cyber-physical infrastructure. In this regard, we assume that any anomaly in the behavior of the overall system is due to an attack. This assumption will be relaxed later to include network effects in our approach.

4. ENERGY-BASED ATTACK DETECTION

In this section, we derive the energy balance for the networked control system in terms of the input-output wave variables, U_{r_k} and V_{r_k}. Next, we provide a generalized characterization of attacks based on the derived energy balance. We then evaluate the impact of the attack models presented in Section 3.2. Finally, we consider the case where the states of the system are not measurable, in which case we introduce the use of an observer to estimate the states.

4.1 Discrete-Time Energy Balance

We present the energy-based attack detection mechanism for the networked control system in Figure 1. We first present the derivation of general energy balance in terms of the plant's input, u_k and output, y_k, and then we refine the derivation to represent the energy balance system in terms of the wave variables exchanged over the network.

PROPOSITION 1. *Consider the discrete-time physical plant, $\mathcal{H}_p$, with a minimal realization (controllable and observable) defined in (6). If $\mathcal{H}_p$ is QSR dissipative then it satisfies the energy balance, E_T given by*

$$E_T = E_{su} - E_{st} - E_d = 0 \tag{20}$$

where E_{su} is the supplied energy, E_{st} is the stored energy and E_d is the dissipated energy.

PROOF. Recall the storage function, V_k, defined as $\frac{1}{2}x_k^T P x_k$. The change in the storage function, ΔV is given by

$$\Delta V = V_{k+1} - V_k = \frac{1}{2}x_{k+1}^T P x_{k+1} - \frac{1}{2}x_k^T P x_k$$

substituting x_{k+1} from (6), we have

$$\begin{aligned} \Delta V &= \frac{1}{2}((x_k^T A^T + u_k^T B^T)P(Ax_k + Bu_k) - x_k^T P x_k) \\ &= \frac{1}{2}(x_k^T(A^T PA - P)x_k + x_k^T A^T PBu_k + u_k^T B^T PAx_k \\ &\quad + u_k^T B^T PBu_k) \end{aligned} \tag{21}$$

From the Generalized KYP lemma described in lemma 1, we can substitute (3), (4) and (5) into equation (21), then we have

$$\begin{aligned}\Delta V = \frac{1}{2}(x_k^T(C^TQC - LL^T)x_k \\ + x_k^T(C^TQD + C^TS - LW)u_k \\ + u_k^T(D^TQC + S^TC - W^TL^T)x_k \\ + u_k^T(R + D^TS + S^TD + D^TQD - W^TW)u_k)\end{aligned} \qquad (22)$$

After some manipulation and simplification, we have

$$\begin{aligned}\Delta V =& \frac{1}{2}((Cx_k + Du_k)^TQ(Cx_k + Du_k) \\ &+ 2(Cx_k + Du_k)^TSu_k + u_k^TRu_k) \\ &-\frac{1}{2}(x_k^TLL^Tx_k + x_k^TLWu_k + u_k^TW^TL^Tx_k \\ &+ u_k^TW^TWu_k)\end{aligned}$$

From (6), noting that $y_k = Cx_k + Du_k$, we can now write

$$\begin{aligned}\Delta V = \frac{1}{2}(y_k^TQy_k + 2y_k^TSu_k + u_k^TRu_k) \\ -\frac{1}{2}(x_k^TLL^Tx_k + x_k^TLWu_k + u_k^TW^TL^Tx_k \\ + u_k^TW^TWu_k)\end{aligned}$$

Summing over the time interval from $k = 0$ to $k = N$ with a sampling time of T_s. The total energy equation becomes

$$\begin{aligned}&\frac{T_s}{2}\sum_{k=0}^{N}(y_k^TQy_k + 2y_k^TSu_k + u_k^TRu_k) - T_s(V_{k+1} - V_{k_0}) \\ &-\frac{T_s}{2}\sum_{k=0}^{N}(x_k^TLL^Tx_k + x_k^TLWu_k + u_k^TW^TL^Tx_k \\ &+ u_k^TW^TWu_k) = 0\end{aligned} \qquad (23)$$

where

$$\begin{aligned}E_{su} &= \frac{T_s}{2}\sum_{k=0}^{N}(y_k^TQy_k + 2y_k^TSu_k + u_k^TRu_k) \\ E_{st} &= T_s(V_{k+1} - V_{k_0}) \\ E_d &= \frac{T_s}{2}\sum_{k=0}^{N}(x_k^TLL^Tx_k + x_k^TLWu_k + u_k^TW^TL^Tx_k \\ &\quad + u_k^TW^TWu_k)\end{aligned}$$

Hence, $\quad E_T = E_{su} - E_{st} - E_d = 0 \quad \square$

The system under consideration is a networked system and the components of the system communicate over a packet-switched network. Hence, it is appropriate to directly relate the energy based equation to the transmitted and received components over the network. We now provide the energy balance in terms of the exchanged wave variables.

PROPOSITION 2. *Given the system $\mathcal{H}_p$ with the energy balance as defined in (23) and the wave transformation provided in (8)-(11). The resulting energy balance of the system in wave domain is*

$$E_{T_{wv}} = E_{su_{wv}} - E_{st_{wv}} - E_{d_{wv}} = 0 \qquad (24)$$

where $E_{T_{wv}}$ is the total of the system, $E_{su_{wv}}$ is the supplied energy, $E_{st_{wv}}$ is the stored energy and $E_{d_{wv}}$ is the dissipated energy.

PROOF. From equations (9) and (10), and also assuming **b**$= 1$, solving for the plant output, y_k and input, u_k, we have

$$y_k = \frac{1}{\sqrt{2}}(U_{r_k} + V_{r_k}) \qquad (25)$$

$$u_k = \frac{1}{\sqrt{2}}(U_{r_k} - V_{r_k}) \qquad (26)$$

After some manipulations and simplification, the plant dynamics can be expressed in terms of the input wave variable, U_{r_k} and output wave variable, V_{r_k}. The resulting system, $\mathcal{H}_{p_{wv}}$ can be described as

$$\mathcal{H}_{p_{wv}} : \begin{cases} x_{k+1} = \bar{A}x_k + \bar{B}U_{r_k} \\ V_{r_k} = \bar{C}x_k + \bar{D}U_{r_k} \end{cases} \qquad (27)$$

with

$$\bar{A} = A - B(D+I)^{-1}C; \bar{B} = \frac{B}{\sqrt{2}}(I - (D+I)^{-1}(D-I))$$
$$\bar{C} = \sqrt{2}(D+I)^{-1}C; \bar{D} = (D+I)^{-1}(D-I)$$

Recall the total energy expression given in (23), by substitution, the energy balance in the wave domain becomes

$$\begin{aligned}&\frac{T_s}{2}\sum_{i=0}^{N}(V_{r_k}^T\bar{Q}V_{r_k} + 2V_{r_k}^T\bar{S}U_{r_k} + U_{r_k}^T\bar{R}U_{r_k}) - T_s(V_{k+1} - V_0) \\ &\quad -\frac{T_s}{2}\sum_{i=0}^{N}(x_k^T\overline{LL^T}x_k + x_k^T\overline{LW}U_{r_k} + U_{r_k}^T\overline{W^TL^T}x_k \\ &\quad + U_{r_k}^T\overline{W^TW}U_{r_k}) = 0\end{aligned} \qquad (28)$$

where

$$\bar{Q} = (\frac{Q - 2S + R}{2}); \bar{S} = (\frac{Q - R}{2}); \bar{R} = (\frac{Q + 2S + R}{2}); \qquad (29)$$

$$\begin{aligned}\overline{LL^T} &= (LL^T - \frac{LW\bar{C}}{\sqrt{2}} - \frac{\bar{C}W^TL^T}{\sqrt{2}} + \bar{C}^TW^TW\bar{C}) \\ \overline{LW} &= (\frac{LW}{\sqrt{2}} - \frac{LW\bar{D}}{\sqrt{2}} - \frac{\bar{C}^TW^TW}{2} + \frac{\bar{C}^TW^TW\bar{D}}{2}) \\ \overline{W^TL^T} &= (\frac{W^TL^T}{\sqrt{2}} - \frac{\bar{D}^TW^TL^T}{\sqrt{2}} - \frac{W^TW\bar{C}}{2} + \frac{\bar{D}^TW^TW\bar{C}}{2}) \\ \overline{W^TW} &= (\frac{W^TW - W^TW\bar{D} - \bar{D}W^TW + \bar{D}^TW^TW\bar{D}}{2})\end{aligned}$$

With

$$\begin{aligned}E_{su_{wv}} &= \frac{T_s}{2}\sum_{i=0}^{N}(V_{r_k}^T\bar{Q}V_{r_k} + 2V_{r_k}^T\bar{S}U_{r_k} + U_{r_k}^T\bar{R}U_{r_k}) \\ E_{st_{wv}} &= T_s(V_{k+1} - V_0) \\ E_{d_{wv}} &= \frac{T_s}{2}\sum_{i=0}^{N}(x_k^T\overline{LL^T}x_k + x_k^T\overline{LW}U_{r_k} + U_{r_k}^T\overline{W^TL^T}x_k \\ &\quad + U_{r_k}^T\overline{W^TW}U_{r_k})\end{aligned}$$

Hence, $E_{T_{wv}} = E_{su_{wv}} - E_{st_{wv}} - E_{d_{wv}} = 0 \quad \square$

4.2 Energy Balance in the Presence of Attacks

In this section, we provide a generalized characterization of the total energy in the presence of attacks.

THEOREM 1. *Consider the networked control system depicted in Figure 1, under cyber-attack, $\mathcal{A}_k$, where by the attacker can remove or modify the exchanged wave variables U_{r_k} and V_{r_k}. Since the plant is assumed to be linear and time-invariant, the modified variables due to an attack can be modeled as*

$$\tilde{U}_{r_k} = U_{r_k} + U_{a_k}; \tilde{V}_{r_k} = V_{r_k} + V_{a_k}; \tilde{x}_k = x_k + x_{a_k} \quad (30)$$

The total energy of the plant, $\tilde{E}_{T_{wv}}$, in the presence of attack is

$$\tilde{E}_{T_{wv}} = E_{T_a} \neq 0 \quad (31)$$

PROOF. Based on the attack-modified input-output relations, the energy for the system becomes

$$\begin{aligned}\tilde{E}_T &= \frac{T_s}{2}\sum_{k=0}^{N}(\tilde{V}_{r_k}^T\bar{Q}\tilde{V}_{r_k} + 2\tilde{V}_{r_k}^T\bar{S}\tilde{U}_{r_k} + \tilde{U}_{r_k}^T\bar{R}\tilde{U}_{r_k})\\ &- \frac{T_s}{2}\sum_{k=0}^{N}(\tilde{x}_k L\bar{L}^T\tilde{x}_k + \tilde{x}_k L\bar{W}\tilde{U}_{r_k} + \tilde{U}_{r_k}W^{\bar{T}}L^T\tilde{x}_k\\ &+ \tilde{U}_{r_k}W^{\bar{T}}W\tilde{U}_{r_k})\\ &- T_s(\frac{1}{2}\tilde{x}_{k+1}P\tilde{x}_{k+1} - \frac{1}{2}\tilde{x}_0 P\tilde{x}_0) \quad (32)\end{aligned}$$

Next, we simplify the above total energy based on the individual energy components which include supplied, stored and dissipated energies. For the new supplied energy we have,

$$\tilde{E}_{su} = \frac{T_s}{2}\sum_{k=0}^{N}(\tilde{V}_{r_k}^T\bar{Q}\tilde{V}_{r_k} + 2\tilde{V}_{r_k}^T\bar{S}\tilde{U}_{r_k} + \tilde{U}_{r_k}^T\bar{R}\tilde{U}_{r_k}) \quad (33)$$

substituting (30) in (33), we have

$$\begin{aligned}\tilde{E}_{su} &= \frac{T_s}{2}\sum_{k=0}^{N}((V_{r_k} + V_{a_k})^T\bar{Q}(V_{r_k} + V_{a_k})\\ &+ 2(V_{r_k} + V_{a_k})^T\bar{S}(U_{r_k} + U_{a_k})\\ &+ (U_{r_k} + U_{a_k})^T\bar{R}(U_{r_k} + U_{a_k}))\\ &= \frac{T_s}{2}\sum_{k=0}^{N}(V_{r_k}^T\bar{Q}V_{r_k} + 2V_{r_k}^T\bar{S}U_{r_k} + U_{r_k}^T\bar{R}U_{r_k})\\ &+ \frac{T_s}{2}\sum_{k=0}^{N}(V_{a_k}^T\bar{Q}V_{a_k} + 2V_{r_k}^T\bar{Q}V_{a_k} + 2V_{r_k}^T\bar{S}U_{a_k})\\ &+ \frac{T_s}{2}\sum_{k=0}^{N}(2V_{a_k}^T\bar{S}U_{r_k} + 2V_{a_k}^T\bar{S}U_{a_k} + 2U_{r_k}^T\bar{R}U_{a_k}\\ &+ U_{a_k}^T\bar{R}U_{a_k}) \quad (34)\end{aligned}$$

From (34) above, it can be seen that,

$$\tilde{E}_{su} = E_{su_{wv}} + E_{su_a} \quad (35)$$

Next, the new stored energy component becomes,

$$\tilde{E}_{st} = T_s(\frac{1}{2}\tilde{x}_{k+1}^T P\tilde{x}_{k+1} - \frac{1}{2}\tilde{x}_0^T P\tilde{x}_0) \quad (36)$$

substituting (30) in (36), we have

$$\begin{aligned}\tilde{E}_{st} &= T_s(\frac{1}{2}(x_{k+1} + x_{a_{k+1}})^T P(x_{k+1} + x_{a_{k+1}})\\ &- \frac{1}{2}(x_0 + x_{a_0})^T P(x_0 + x_{a_0}))\\ &= \frac{T_s}{2}((x_{k+1}^T Px_{k+1} - x_0^T Px_0)\\ &+ (x_{a_{k+1}}^T Px_{a_{k+1}} - x_{a_0}^T Px_{a_0} + 2x_{k+1}^T Px_{a_{k+1}} - x_0^T Px_{a_0})) \quad (37)\end{aligned}$$

From (37) above, it can be seen that,

$$\tilde{E}_{st} = E_{st_{wv}} + E_{st_a} \quad (38)$$

Finally, the new dissipated energy component becomes

$$\begin{aligned}\tilde{E}_d &= \frac{T_s}{2}\sum_{k=0}^{N}(\tilde{x}_k L\bar{L}^T\tilde{x}_k + \tilde{x}_k^T L\bar{W}\tilde{U}_{rk} + \tilde{U}_{rk}W^{\bar{T}}L^T\tilde{x}_k\\ &+ \tilde{U}_{rk}W^{\bar{T}}W\tilde{U}_{rk}) \quad (39)\end{aligned}$$

substituting (30) in (39), we have

$$\begin{aligned}\tilde{E}_d &= \frac{T_s}{2}\sum_{k=0}^{N}(x_k^T L\bar{L}^T x_k + x_k^T L\bar{L}^T x_{a_k} + x_{a_k}^T L\bar{L}^T x_k\\ &+ x_{a_k}^T L\bar{L}^T x_{a_k} + x_k^T L\bar{W}U_{rk} + x_k^T L\bar{W}U_{a_k} + x_{a_k}^T L\bar{W}U_{rk}\\ &+ x_{a_k}^T L\bar{W}U_{a_k} + U_{rk}^T W^{\bar{T}}L^T x_k + U_{rk}^T W^{\bar{T}}L^T x_{a_k}\\ &+ U_{a_k}^T W^{\bar{T}}L^T x_k + U_{a_k}^T W^{\bar{T}}L^T x_{a_k} + U_{rk}W^{\bar{T}}WU_{rk}\\ &+ U_{rk}W^{\bar{T}}WU_{a_k} + U_{a_k}W^{\bar{T}}WU_{rk} + U_{a_k}W^{\bar{T}}WU_{a_k})\end{aligned} \quad (40)$$

From (40) above, it can be seen that,

$$\tilde{E}_d = E_{d_{wv}} + E_{d_a} \quad (41)$$

Hence, from (35), (38) and (41), the total energy, $\tilde{E}_T$ in the presence of attack(s), then becomes

$$\begin{aligned}\tilde{E}_T &= \tilde{E}_{su} - \tilde{E}_{st} - \tilde{E}_d\\ &= E_{su_{wv}} + E_{su_a} - E_{st_{wv}} - E_{st_a} - E_{d_{wv}} - E_{d_a}\\ &= E_{T_{wv}} + E_{T_a} \quad (42)\end{aligned}$$

From (24), we have

$$\tilde{E}_T = E_{T_{wv}} + E_{T_a} = E_{T_a} \quad (43)$$

□

COROLLARY 1. *In the absence of any detectable attack, $\mathcal{A}_k$, the total energy of the system, $\tilde{E}_{T_{wv}}$ is*

$$\tilde{E}_{T_{wv}} = E_{T_{wv}} = 0 \quad (44)$$

PROOF. This result follows directly from the system total energy property described in Theorem 2 and the results in Theorem 1 in the presence attacks. □

REMARK 1. *The detection algorithm for the monitor is evaluated based on the information received at the controller. Considering the fact that the controller is considered trustworthy, the effects of attacks on the wave variable, U_{l_k} which is received as U_{r_k} by the plant will be reflected on wave variable V_{l_k}, which is V_{r_k} sent from the plant side of the network. Recall the expression in (9) relating the actuator signal and sensor signal , to the wave variable,*

$$V_{r_k} = \frac{1}{\sqrt{2b}}(y_k - bu_k)$$

It is straight forward to see that attacks on either the sensor or actuator will be reflected on the wave variable V_{r_k}, which is subsequently received at the controller as V_{l_k}.

COROLLARY 2. *An attack, $\mathcal{A}_k$, is characterized as a passive attack if the presence of the attack results in $\tilde{E}_{T_{wv}} > 0$.*

From the definition of passivity in (2), $\tilde{E}_{T_{wv}} > 0$ implies that the supplied energy for the attack system is larger than the dissipated and stored energies.

COROLLARY 3. *An attack, $\mathcal{A}_k$, is characterized as a non-passive attack if the presence of the attack results in $\tilde{E}_{T_{wv}} < 0$.*

This essentially implies that the supplied energy of the attacked system is less than the dissipated and store energies. Therefore, the system generates additional internal energy which results in a non-passive behavior. This implies that the overall stability of the networked control system is no longer guaranteed.

The energy-based attack detector can be summarized by Algorithm 1. Figure 2 also shows the block diagram for the energy based monitor. The inputs to the algorithm, also denoted in Figure 2, are the wave variables, V_{r_k} and U_{r_k} and the plant's state x_k. The output of the algorithm is Ψ, which provides information on whether an attack has occurred and the impact of the attack on the overall system. The blocks supplied energy, stored energy and dissipated energy in Figure 2 corresponds to the computation of the supplied energy, stored energy and dissipated energy respectively as indicated by lines 1-4 in Algorithm 1. Figure 2 shows the block diagram for the designed energy based monitor for attacks in the case of measurable plant states.

Algorithm 1: Energy-Based Attack Detection

```
   Input: $V_{r_k}$,$U_{r_k}$,$x_k$
   Output: $\Psi$
1  Compute the supplied energy, $E_{su_{wv}}$
2  Compute the stored energy, $E_{st_{wv}}$
3  Compute the dissipated energy, $E_{d_{wv}}$
4  Compute the total energy, $E_{T_k} = E_{su_{wv}} - E_{st_{wv}} - E_{d_{wv}}$
5  if $E_{T_k} \neq 0$ then
6      $\psi_1$ =True
7      if $E_{T_k} > 0$ then
8          $\psi_2$ =Passive
9      else
10         $\psi_2$ =Non-Passive
11 else
12     $\psi_1$ =False
13 $\Psi$ ={$\psi_1$,$\psi_2$}
14 return $\Psi$
```

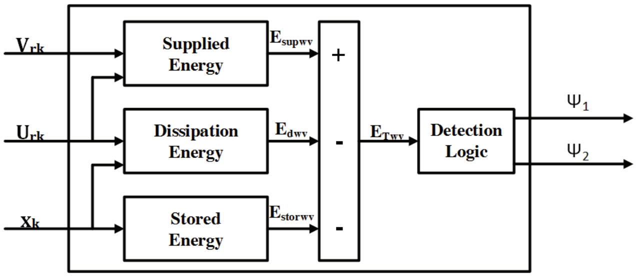


Figure 2: Energy-Based Monitor

REMARK 2. *Thus far, the characterization of attacks are based on the notion that in the absence of attacks, the nominal energy balance of the monitored system should be zero. In a more realistic setting, this assumption can be relaxed in order to integrate the potential effects of the network communication as a result of delays or packet loss. In the presence of network effects and possibly other system uncertainties, instead of the energy balance being zero, a notion of a maximal value of energy due to network effects and system uncertainties are considered. Based on this notion, a threshold boundary, E_{th}, is defined. The characterization of attacks are then evaluated based on the impact of the attacks that results in computed energy that lies outside the boundary. This maximal energy value can be obtained empirically through simulations and by imposing worst-case network conditions for the NCS.*

4.3 Characterization of Attack Models

In order to illustrate the impact of attack models on the physical system, we evaluate the effect of classical attacks on the total energy of the system. For brevity, we focus on the attacks **A1** and **A2**, although similar approach can be used to evaluate the effects of attacks **A3**-**A6**. Also, due to space limitations we consider the cases where the dissipative plant is passive. The proofs for the presented results as well as the case for strictly-output passive plant is provided in [5].

Assuming there are no attacks on U_{r_k}, the impact of the attacks on V_{r_k} is reflected on only the supplied energy resulting in the component,

$$\begin{aligned}\tilde{E}_{T_{wv}} &= E_{T_{wva}} = E_{su_{wva}} \\ &= \frac{T_s}{2}\sum_{k=0}^{N}(2V_{r_k}^T\bar{Q}V_{a_k} + V_{a_k}^T\bar{Q}V_{a_k} + 2V_{a_k}^T\bar{S}U_{r_k}) \quad (45)\end{aligned}$$

PROPOSITION 3. *Consider the passivity-based network control system depicted in Figure 1, under a max integrity attack, $\mathcal{A}_k$, if the system dynamics $\mathcal{H}_p$ is passive, then*

$$\mathcal{A}_k : \begin{cases} \textit{Passive} \quad \textit{if } \sum\limits_{k=0}^{N} V_{r_{max}}^T V_{r_{max}} < \sum\limits_{k=0}^{N} V_{r_k}^T V_{r_k} \\ \textit{Non-Passive} \ \textit{if } \sum\limits_{k=0}^{N} V_{r_{max}}^T V_{r_{max}} > \sum\limits_{k=0}^{N} V_{r_k}^T V_{r_k} \end{cases} \quad (46)$$

PROPOSITION 4. *Consider the passivity-based network control system depicted in Figure 1, under a min integrity attack, $\mathcal{A}_k$, if the system dynamics $\mathcal{H}_p$ is passive, then*

$$\mathcal{A}_k : \begin{cases} \textit{Passive} \quad \textit{if } \sum\limits_{k=0}^{N} V_{r_{min}}^T V_{r_{min}} < \sum\limits_{k=0}^{N} V_{r_k}^T V_{r_k} \\ \textit{Non-Passive} \ \textit{if } \sum\limits_{k=0}^{N} V_{r_{min}}^T V_{r_{min}} > \sum\limits_{k=0}^{N} V_{r_k}^T V_{r_k} \end{cases} \quad (47)$$

PROPOSITION 5. *Consider the passivity-based network control system depicted in Figure 1, under an additive integrity attack, $\mathcal{A}_k$, if the system dynamics $\mathcal{H}_p$ is passive, then*

$$\mathcal{A}_k : \begin{cases} \textit{Passive} \quad \textit{if } \sum\limits_{k=0}^{N} 2V_{r_k}^T \alpha_k < -\sum\limits_{k=0}^{N} \alpha_k^T \alpha_k \\ \textit{Non-Passive} \ \textit{if } \sum\limits_{k=0}^{N} 2V_{r_k}^T \alpha_k > -\sum\limits_{k=0}^{N} \alpha_k^T \alpha_k \end{cases} \quad (48)$$

PROPOSITION 6. *Consider the passivity-based network control system depicted in Figure 1, under a denial-of-service attack, $\mathcal{A}_k$, if the system dynamics $\mathcal{H}_p$ is passive, then*

$$\mathcal{A}_k : \textit{Passive with } E_{su_{wva}} = \frac{T_s}{2}\sum_{k=0}^{N}\frac{V_{r_k}^T V_{r_k}}{2} > 0 \forall V_{r_k} \neq 0 \quad (49)$$

REMARK 3. *The result obtained for the characterization of DoS attacks is similar to the analysis of packet losses due to unreliability of network in the literature. While packet losses are due to unreliable network, DoS is as a result of intentional and malicious attacks by an adversary.*

4.4 The case of unmeasurable states

In the case unmeasurable plant states, a Luenberger observer of the form in (50) is introduced to reconstruct an estimate of the plant states.

$$\mathcal{H}_{obs}:\begin{cases}\hat{x}_{k+1}=A\hat{x}_k+BU_{r_k}-L(V_{r_k}-C\hat{x}_k-DU_{r_k})\\ \hat{V}_{r_k}=C\hat{x}_k+DU_{r_k}\end{cases} \tag{50}$$

where L is the observer gain. Recall that the plant system is assumed to be observable. This means there exists an observability matrix L such that the estimated state $\hat{x}_k$ of the Luenberger observer asymptotically converges to the true state x_k.

PROPOSITION 7. *Given the system $\mathcal{H}_p$ with the energy balance described in Theorem 2. In the case whereby the states are unmeasurable assuming a Luenberger observer, $\mathcal{H}_{obsv}$ as given in (50) is integrated to estimate the states, $\hat{x}_k$. Then, the resulting equivalent total energy of the system in wave domain, in the absence of attacks can be described by*

$$E_{T_{wv}}=E_{T_{wvo}}=-E_{st_{oe}}-E_{d_{oe}} \tag{51}$$

Due to limited space, the proof is presented in [5].

Similar to Algorithm 1, in the case of unmeasurable states, the energy-based detection with the integration of an observer can be summarized by Algorithm 2 below.

Algorithm 2: Energy-Based Attack Detection in the case of unmeasurable states

```
Input: $V_{r_k}$,$U_{r_k}$
Output: $\Psi$
Estimate the states, $\hat{x}_k$
Compute the supplied energy, $E_{su_{wvo}}$
Compute the stored energy, $E_{st_{wvo}}$
Compute the dissipated energy, $E_{d_{wvo}}$
Compute the total energy,
    $E_{T_{wvo}}=E_{su_{wvo}}-E_{st_{wvo}}-E_{d_{wvo}}$
if $E_{T_{wvo}}\neq 0$ then
    $\psi_1$ =True
    if $T_{wvo}>0$ then
        $\psi_2$ =Passive
    else
        $\psi_2$ =Non-Passive
else
    $\psi_1$ =False
$\Psi$ ={$\psi_1$,$\psi_2$}
return $\Psi$
```

5. EVALUATION

In this section, we evaluate the proposed energy-based detection mechanism using simulations. The system under consideration is a control system composed of a plant and controller that exchange information over a network in order to cooperatively achieve a specified objective.

5.1 Simulation Setup

The case study involves the velocity control of a single joint robotic arm over a communication network. It is assumed that the only information the networked controller receives from the plant is the wave variable, V_{r_k}, which becomes V_{l_k} at the controller side of the network. Hence the detection mechanism with an integrated observer, as described in Section 4.4, is used in this evaluation. The NCS is considered to passive by design.

The simulation of the NCS is performed in Matlab/Simulink. The plant, controller, energy-based monitor, scattering transformation, attack models and communication are implemented using a combination of Matlab scripts and blocks from the Simulink library. The dynamics of the plant is described by the discrete-time state-space representation as defined in (6) with a sampling time of $T_s = 0.01s$. The parameters for the plant are A=0.9952, B=0.0625, C=0.1214, D=0.0251. The controller for the robot is a Proportional-Integral (PI) controller and similar to the plant is represented by the discrete time state space representation defined in (7) with the sampling time $T_s = 0.01s$. The parameters for the controller are A$_c$ =1, B$_c$ =0.0625, C$_c$ =0.1, D$_c$ =0.6385. The main objective of the controller is to modify the behavior of the plant in order to track a reference velocity trajectory, r_k over a communication network.

5.2 Scenarios

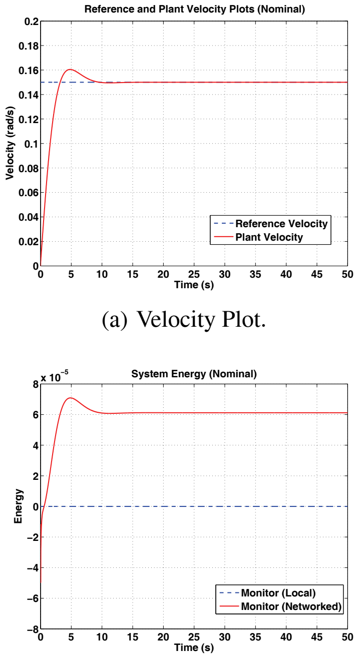
First, we present the control of the plant in the nominal case when there are no attacks. We also present the effects of the network on the system's energy. Next, we evaluate the behavior of the system under attack and the ability of the proposed approach to detect the attacks. In the experiments with attacks, the simulated attacks are injected from the duration, $t = 15s$ to $t = 20s$. Due to space limitations, we only present results for some of the attack models, results for other cases are presented in [5].

(1) Nominal Case: In this scenario, the NCS operates nominally while achieving the tracking objective. Figure 3a depicts the reference velocity of $0.15rad/s$ as well as the plant velocity clearly showing that the plant is able to track the velocity as desired. Figure 3b shows the energy balance of the monitored plant computed based on the approach described in Section 4.2. In order to illustrate the effect of communication network on the energy-balance, we also co-located the energy-based detector at the plant side of the network to essentially perform the same total energy computation. The only difference being the delay experience by the monitor co-located with the controller. From Figure 3b, the energy-balance computed by the local monitor is essentially zero as expected but the balance computed by the networked monitor has an offset as a result of the communication network. Hence, this offset value or a more conservative value can be used to characterize the threshold energy, E_{th}, which will be non-zero due to network effects.

(2) Integrity Attacks

(a) Min Attack on V_{rk}: In this scenario, we assume the NCS channel from the plant to the controller, transmitting V_{rk}, is compromised and during the attack duration, the attacker replaces the true or actual signal exchanged with the value of $V_{r_{min}}$. Figure 4a shows that the presence of the attack results in the degraded reference tracking performance. Figure 4b depicts the total energy computation clearly indicating the presence of the attack. Additionally, one can observe that the min integrity attack can be characterize as a passive attack as it leads to increase in the computed total energy which indicates the dissipation of energy. Based on the computed energy, passivity of the overall NCS is still guaranteed.

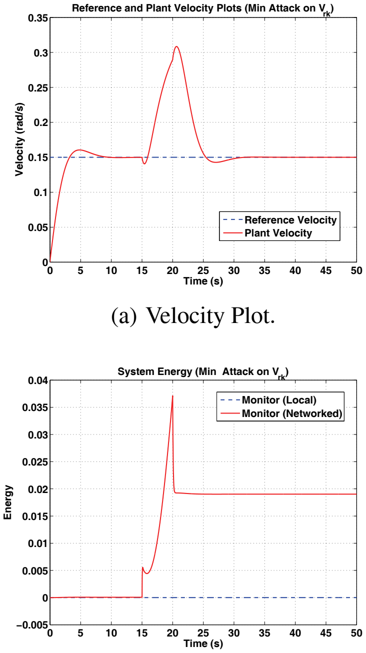
(b) Max Energy Attack on V_{rk}: In this scenario, we again assume the NCS channel from the plant to the controller, transmitting V_{rk}, is compromised and during the attack duration, the attacker replaces the true or actual signal exchanged with the value of V_{r_a} that maximizes the energy dissipated at that time step. Figure 5a shows that the presence of the attack results in the degraded reference tracking performance. Figure 4b depicts the total energy com-



(a) Velocity Plot.

(b) Energy Balance Plot.

Figure 3: Nominal Case.

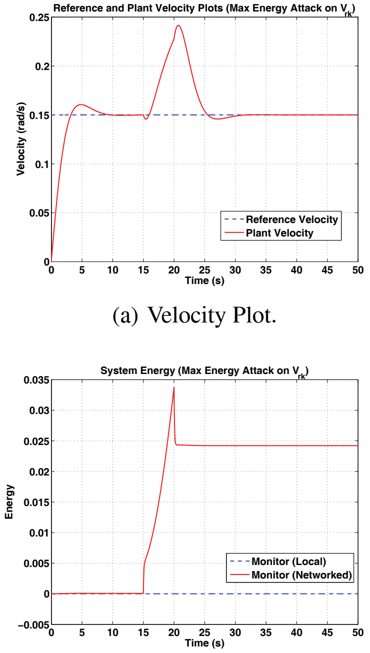


(a) Velocity Plot.

(b) Energy Balance Plot.

Figure 4: Min Attack on V_{rk}.

putation clearly indicating the presence of the attack. Additionally, one can observe that as expected the max energy attack results in an increase in the computed total energy. Based on the computed energy, passivity of the overall NCS is still guaranteed.



(a) Velocity Plot.

(b) Energy Balance Plot.

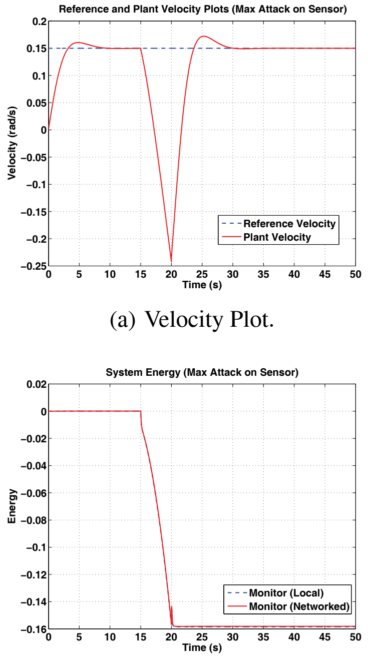
Figure 5: Max Energy Attack on V_{rk}.

(c) Max Attack on the Sensor, y_k: In this scenario, we assume the sensor signal from the plant, y_k, is compromised and during the attack duration, the attacker replaces the true or actual signal exchanged with the value of y_{max}. Figure 6a shows that the presence of the attack results in the degraded reference tracking performance. Figure 6b depicts the total energy computation clearly indicating the presence of the attack. From the energy plots, one can observe that as expected the max integrity attack on the sensor perturbs both the local and networked energy monitors, clearly indicating that the attack is perpetuated locally. From Figure 6b, the max attack on the actuator results in a decrease in the system energy indicating the injection of excess energy and hence can be characterize as a non-passive attack. Based on the computed energy, the injected max attack on the sensor leads to the violation of passivity of the overall NCS, in addition to the observed significant degradation in performance.

(3) Denial-of-service attack on V_{rk}: In this scenario, we introduce a denial-of-service attack on the NCS. During the attack duration, the attacker erases or discard the information exchanged over the network based on a simulated Bernoulli random variable, the probability of erasure for this evaluation was set at 0.2. Figure 7 shows that the presence of the attack clearly degrades the tracking performance. From Figure 7, one can observe that the denial-of-service attack can be characterized as passive attack as it leads to a positive total computed energy and this is in line with the results in Section 4.3 defining DoS attacks as always passive. Hence, passivity is always maintained but the performance in tracking is deteriorated.

6. CONCLUSION

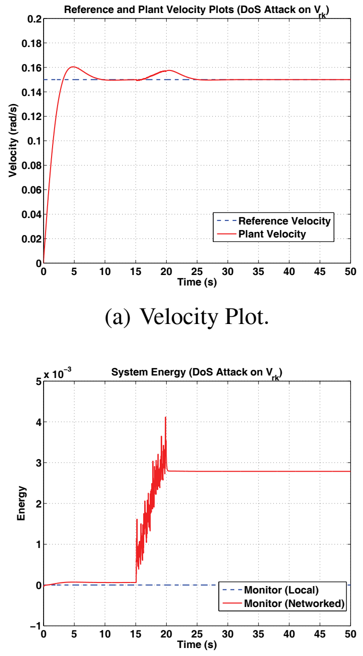
Due to increased attacks on CPS, there is an increased effort towards approaches to detect and secure CPS from cyber attacks. We present an energy-based attack detector for a class of CPS that are considered dissipative. We provided analytical results to show the detector can successfully detect attacks. Using well-known attack models we characterize attacks that can be considered either passive or non-passive based on their impact on the evaluated system



(a) Velocity Plot.

(b) Energy Balance Plot.

Figure 6: Max Attack on Sensor, y_k.



(a) Velocity Plot.

(b) Energy Balance Plot.

Figure 7: DoS Attack on V_{rk}.

energy. We quantitatively evaluate the performance of the proposed mechanism using simulations and experiments on a networked single joint robotic arm with the introduction of artificially simulated attacks. The results show that the proposed detection mechanism is effective in detecting attacks based on the energy balance of a system.

7. ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation (CNS-1238959, CNS-1035655) and NIST (70NANB13H169).

8. REFERENCES

[1] M. Abrams and J. Weiss. Malicious control system cyber security attack case study: maroochy water services. 2008.

[2] C. I. Byrnes and W. Lin. Losslessness, feedback equivalence, and the global stabilization of discrete-time nonlinear systems. *IEEE Trans. on Aut. Control,*, 39(1):83–98, 1994.

[3] A. A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Proc. of the 3rd Conf. on Hot topics in security*, pages 1–6, 2008.

[4] W. Chen, S. Ding, A. Q. Khan, and M. Abid. Energy based fault detection for dissipative systems. In *Conf. on Control and Fault-Tolerant Systems*, pages 517–521, 2010.

[5] E. Eyisi and X. Koutsoukos. Energy-Based Attack Detection in Networked Control Systems. Technical Report ISIS-13-107.

[6] N. Falliere, L. Murchu, and C. E. W32.stuxnet dossier. 2011.

[7] C. Fantuzzi and C. Secchi. Energetic approach to parametric fault detection and isolation. In *American Control Conf.*, volume 6, pages 5034–5039, 2004.

[8] G. C. Goodwin and K. S. Sin. *Adaptive filtering prediction and control*. Courier Dover Publications, 2013.

[9] W. M. Haddad and V. Chellaboina. *Nonlinear dynamical systems and control: a Lyapunov-based approach*. Princeton University Press, 2011.

[10] D. Hill and P. Moylan. The stability of nonlinear dissipative systems. *IEEE Trans. on Aut. Control*, 21(5):708–711, 1976.

[11] Y.-L. Huang, A. A. Cárdenas, S. Amin, Z.-S. Lin, H.-Y. Tsai, and S. Sastry. Understanding the physical and economic consequences of attacks on control systems. *Int. J. of Critical Infrastructure Protection*, 2(3):73–83, 2009.

[12] T. W. S. Journal. Electricity grid in u.s. penetrated by spies. A1, April 2009.

[13] C. Li, A. Raghunathan, and N. Jha. Hijacking an insulin pump: Security attacks and defenses for a diabetes therapy system. In *Int. Conf. In e-Health Networking Applications and Services*, pages 150 –156, June 2011.

[14] R. Patton and J. Chen. Observer-based fault detection and isolation: Robustness and applications. *Control Engineering Practice*, 5(5):671 – 682, 1997.

[15] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson. Attack models and scenarios for networked control systems. In *Proc. of the 1st Int. Conf. on High Confidence Net. Sys.*, pages 55–64, 2012.

[16] D. Theilliol, H. Noura, D. Sauter, and F. Hamelin. Sensor fault diagnosis based on energy balance evaluation: Application to a metal processing. *ISA Trans.*, 45(4):603–610, 2006.

[17] J. C. Willems. Dissipative dynamical systems part ii: Linear systems with quadratic supply rates. *Arch. for Rat. Mechanics and Analysis*, 45(5):352–393, 1972.

[18] H. Yang, V. Cocquempot, and B. Jiang. Fault tolerance analysis for switched systems via global passivity. *IEEE Trans. on Cir. and Sys. II: Express Briefs*, 55(12):1279–1283, 2008.