Critical infrastructures must continuously operate safely and reliably, despite a variety of potential system disturbances. Given their strict operating requirements, such systems are automated and controlled in real time by several digital controllers receiving measurements from sensors and transmitting control signals to actuators. Since these physical systems are often spatially distributed, there is a need for information technology (IT) infrastructures enabling the timely data flow between the system components. These networked control systems are ubiquitous in modern societies [1]. Examples include the electric power network, intelligent transport systems, and industrial processes.

Networked control systems are vulnerable to cyberthreats through the use of open communication networks and heterogeneous IT components. Because networked control systems are often operated through supervisory control and data acquisition (SCADA) systems and the measurement and control data in these systems are commonly transmitted through unprotected communication channels, the networked control system is vulnerable to several threats [2]. As illustrative examples, we mention the cyberattacks on power transmission networks operated by SCADA systems [3] and the Stuxnet malware that allegedly infected an industrial control system and disrupted its operation [4], [5].

Unlike other IT systems where cybersecurity mainly involves the protection of data, cyberattacks on networked control systems may influence physical processes through feedback actuation. Therefore, networked control-system security needs to consider threats at both the cyber and physical layers.

Control theory has developed frameworks to handle disturbances and faults [6], [7], and these tools can be used to detect and attenuate the consequences of cyberattacks on networked control systems. However, there are substantial conceptual and technical differences between a fault-tolerant and a secure control framework. Faults are commonly considered to be physical events that affect the system behavior. Simultaneous events are assumed to be noncolluding, in the sense that events do not act in a coordinated way. On the other hand, cyberattacks may be performed over a significant number of attack points in a coordinated fashion; see, for instance, [8]–[10]. Moreover, faults are constrained by the physical dynamics and do not have an intent or objective to fulfill, as opposed to cyberattacks that do have a malicious intent and are not directly constrained by the dynamics of the physical process. Ensuring security may involve addressing a large number of threats, thus requiring the use of risk management methods [11] to prioritize the threats to be mitigated.
The need for novel methods to enhance the cybersecurity of networked control systems has motivated several research directions recently. The general problem of networked control systems under cyberattack is discussed and formalized in [12] and [13]. Various attack scenarios are evaluated on real and simulated benchmark systems in [13] and [14], respectively. For electric power networks, false-data injection attacks have been analyzed in detail in terms of vulnerability quantification [15], attack impact [16], [17], detection schemes [18], and attack evaluation on a realistic test bed [8]. Specific classes of attacks have been analyzed for dynamic control systems, such as replay attacks [19], stealthy false-data injection attacks [9], [10], [13], and denial-of-service attacks [20]. Quantification of tolerable errors in sensor and actuator data and their mitigation is discussed in [21], while [22]–[24] studied the network-wide data dissemination under malicious links, and [25] proposed a game-theoretic approach to cross-layer security under cascading failures.

This article presents some of the recent approaches to address cybersecurity of networked control systems under the unified perspective of risk management. First the architecture and modeling assumptions of the networked control system and adversary are introduced, following the work in [13]. Specifically, we describe the models and assumptions used for the plant, communication network, controller, and anomaly detector. Moreover, important concepts regarding the system’s operation are defined, such as the system’s nominal behavior and safe sets. The adversary’s model, goals, and constraints are also discussed. After describing three fundamental security properties of IT systems, the adversary model is defined in terms of available resources to violate the aforementioned properties, knowledge of the system’s model, and a given attack policy. The attack policy is designed according to the adversary’s aims: to produce the maximum impact on the physical plant while remaining stealthy.

To tackle the existing threats, a defense methodology based on the risk management framework is presented. The notion of risk is defined in terms of a threat’s scenario, impact, and likelihood, and the risk management framework is described. In this article, emphasis is given to the assessment and treatment of risk. In particular, recent quantitative tools developed in [26] and [27] for analyzing the risk of threats of static and dynamic systems are presented. The risk assessment method from [26] is tailored to quantify the likelihood of threats on a static electric power system, while the approach in [27] addresses dynamic systems and analyzes both the likelihood and impact of threats. The proposed risk assessment methods attempt to quantify the risk of different hypothetical attack scenarios for the present configuration and model of the system. As such, these methods are not executed based on real-time data of the system. The outcome from the risk assessment methods may be used for risk treatment, which is also discussed in this article and related to previous work for static and dynamic systems, [18], [28], [29] and [19], [30], respectively.

The outline of this article is as follows. First, the models for the networked control system and adversary are discussed in “Networked Control Systems Under Attacks,” together with the risk management framework. The article proceeds by presenting the risk analysis methods in “Risk Analysis for Stealthy Deception Attacks.” First, the method for static systems is described in detail and illustrated for large-scale electric power systems. Risk treatment methodologies for electric power systems are also discussed. Next, the risk assessment method for the dynamic case considering impact and likelihood is briefly presented and illustrated on a wireless quadruple-tank test bed. Possible risk treatment approaches are also discussed and illustrated. A summary of the article and concluding remarks are presented in the last section.

**NETWORKED CONTROL SYSTEM UNDER ATTACKS**

Networked control systems are spatially distributed systems where the physical plant is operated by digital controllers that receive measurements from spatially distributed sensors and transmit control signals to spatially distributed actuators through a communication network; see [1] and references therein. A typical networked control system structure has the four main components given in Figure 1: the physical plant, communication network, digital feedback controller, and digital anomaly detector. A model of each component is described below. Since this article concerns mainly threats arising from the cyber side of the networked control system, only discrete-time models are discussed in this article, where \( k \in \mathbb{N} \) is the integer time index. Relevant work on the security of networked control system using continuous-time models may be found in [9] and [10].

The plant operation is supported by a communication network through which the sensor measurements and...
actuator data are transmitted. On the plant side, the measurement data correspond to $y_t \in \mathbb{R}^{n_y}$, while $\hat{u}_t \in \mathbb{R}^{n_u}$ represents the actuator data. On the controller side, the sensor and actuator data are denoted by $\hat{y}_t \in \mathbb{R}^{n_y}$ and $u_t \in \mathbb{R}^{n_u}$, respectively. The dynamical model of the plant is given by

$$
\begin{align*}
&x_{t+1} = \Phi(x_t, \hat{u}_t, d_t), \\
&y_t = \Gamma(x_t, \hat{u}_t, d_t),
\end{align*}
$$

where $x_t \in \mathbb{R}^{n_x}$ denotes the plant’s state, and the unknown input $d_t \in \mathbb{R}^{n_d}$ models possible disturbances or faults affecting the system. Mismatches between the transmitted signals, $y_t$ and $u_t$, and the received ones, $\hat{y}_t$ and $\hat{u}_t$, may be due to the communication network, as in the case of delays or packet drops. Since the focus of this article is on the security of the control system with respect to malicious adversaries, the communication network is assumed to be ideal and process and measurement noises are neglected. Under this assumption, the mismatches between the transmitted signals and the received ones are caused by the adversary’s actions. Similarly, physical attacks performed by malicious adversaries are modeled by the unknown input $d_t$. The nominal behavior of the system under no attack is defined as follows.

**Definition 1**
A networked control system is said to have **nominal behavior** if $\hat{u}_t = u_t, y_t = \hat{y}_t$, and $d_t = 0$. Otherwise, the system has **abnormal behavior**.

Several physical systems have tight operating constraints that, if not satisfied, might result in physical damage to the system, for example, power systems, where electrical power flows along transmission lines cannot exceed physical limits. In this work, the concept of safe sets is used to characterize the safety constraints. Consider the time interval $[0, N]$ and define the vector $x = [x_0, \ldots, x_N]^T \in \mathbb{R}^{n_x(N+1)}$. Given the set $\mathcal{S}_r \subseteq \mathbb{R}^{n_x(N+1)}$, safety is defined as follows.

**Definition 2**
The system is said to be **safe** over the time interval $[0, N]$ if the state trajectory $x$ is contained in the **safe set** $\mathcal{S}_r$.

Returning to the power system example, let $x_t$ be the state of the power system and denote the output $y_t = C_t x_t$ as the instantaneous power flow measured on a given transmission line. Due to physical limits, the cable cannot sustain an arbitrarily large instantaneous power. With the appropriate scaling of $y_t$, such an operating constraint can be defined in terms of the safe set $\mathcal{S}_r = \{x : \max_i |C_t x_i| \leq 1\}$.

To comply with performance requirements in the presence of unknown disturbances, the physical plant is assumed to be controlled by an appropriate feedback controller [6], which computes the control signal $u_t$ given the measurement signal $\hat{y}_t$ received through the communication network. The output feedback controller can be written in a state-space form as

$$
\begin{align*}
&z_{t+1} = f_z(z_t, \hat{y}_t), \\
&u_t = g_z(z_t, \hat{y}_t),
\end{align*}
$$

where the states of the controller are labeled as $z_t \in \mathbb{R}^{n_z}$. Given the plant model, the controller is supposed to be designed so that acceptable performance is achieved under nominal behavior.

The anomaly detector, which is collocated with the controller, monitors the system to detect possible deviations from the nominal behavior. It has access to only $\hat{y}_t$ and $u_t$. Several approaches to detecting malfunctions in control systems are available in the fault diagnosis literature [7], [31]. Other schemes tailored to detecting sparse adversarial attacks have also been proposed [32], [33]. A common approach is the observer-based fault detection filter

$$
\begin{align*}
&s_{k+1} = f_s(s_k, u_k, \hat{y}_{k+1}), \\
&y_k = g_s(s_k, u_k, \hat{y}_{k+1}),
\end{align*}
$$

where $s_k \in \mathbb{R}^{n_s}$ is the anomaly detector’s state. Based on the plant and controller models, the control signal $u_k$ and the received measurements $\hat{y}_k$ the fault detection filter computes the residue $r_k \in \mathbb{R}^{n_r}$. The residue signal is evaluated to detect and locate existing anomalies, as depicted in Figure 1.

The anomaly detector (3) is designed such that

1. under nominal behavior of the system ($u_t = \hat{u}_t, y_t = \hat{y}_t$), the expected value of $r_k$ converges asymptotically to a neighborhood of the origin
2. the residue is sensitive to the anomalies so that an abnormal behavior of the system results in a nonzero residue signal.

The aim of an anomaly detector is to evaluate the residue to detect anomalies with high probability while keeping the rate of false alarms due to uncertainties below a certain level. Given the aforementioned design specifications of the anomaly detector, various residue evaluation techniques described in [34] may be used to detect anomalies. For instance, the anomaly detector may be designed to trigger an alarm when a given norm of the residue signal exceeds a certain bound over the time interval $[k_0, k_f]$

$$
\| r \|_{\infty} = \left( \sum_{i=k_0}^{k_f} |r_i|^{p} \right)^{1/p} > \delta,
$$

where $r = [r_{k_0} \ldots r_{k_f}]^T$ is the residue signal over the time interval $[k_0, k_f]$. $\| r \|_{\infty}$ denotes the $\ell_\infty$-norm of $r$, and $\delta > 0$ ensures a desired false-alarm rate with respect to uncertainties. Different methods to compute the threshold $\delta$ and the corresponding false-alarm rate may be found in [31].

**Adversary Model**
Due to the tight coupling between the cyber and physical domains, the control system behavior depends on the state and properties of the IT infrastructure. To model and understand how a cyber adversary may affect the networked control system operation requires knowing how IT systems are vulnerable to adversaries. Computer security literature identifies three fundamental properties of information and services in IT systems, namely confidentiality, integrity, and availability, often denoted as CIA [35]. They
can be violated by disclosure, deception, and denial-of-service attacks, respectively. For examples of attacks violating these properties in networked control systems, see “The CIA in Networked Control Systems.”

Disclosure attacks enable the adversary to gather sequences of data \( I_k \) from the calculated control actions \( u_k \) and the real measurements \( y_k \). As such, the physical dynamics of the system are not affected by this type of attack. Instead, these attacks gather intelligence that may enable more complex attacks, such as replay attacks [36]. On the other hand, deception attacks modify the control actions \( u_k \) and sensor measurements \( y_k \) from their calculated or real values to the corrupted signals \( \tilde{u}_k \) and \( \tilde{y}_k \), respectively. The deception attacks are modeled as

\[
\tilde{u}_k = u_k + \Delta u_k, \\
\tilde{y}_k = y_k + \Delta y_k,
\]

where the vectors \( \Delta u_k \) and \( \Delta y_k \) represent the data corruption to the respective data channels, as depicted in Figure 1. The data corruption vectors may have sparsity patterns according to the adversary’s resources, namely the communication channels that can be corrupted. Similarly, denial-of-service attacks may also affect the transmitted data by preventing it from reaching the desired destination. Attacks that may affect the system behavior directly and through feedback are classified as disruption attacks [13]. From the preceding discussion, we conclude that physical, deception, and denial-of-service attacks are classified as disruption attacks. The data channels and physical actuators required to perform specific disclosure and disruption attacks are denoted as disclosure and disruption resources, respectively.

In addition to the disclosure and disruption resources required to stage a given attack, the adversary’s resources can also include knowledge of the system model. Different attack scenarios can be qualitatively categorized in terms of the required resources in the attack space, as illustrated in Figure 2. A given point in the attack space represents an instance of the adversary model in Figure 3 where each of the adversary resources is mapped to a specific axis of the attack space. The attack policy mapping the model knowledge \( \mathcal{K} \) and disclosed data gathered until time \( k, I_k \), to the attack vector \( a_k \in \mathbb{R}^n \) is denoted as \( a_k = g(\mathcal{K}, I_k) \).

For each attack scenario, the attack policy is designed according to the adversary’s intent, namely the attack goals and constraints. In particular, the attack scenarios in this article consider adversaries whose goal is to drive the state trajectory \( x \) of the physical system to an unsafe set while remaining stealthy, as illustrated in Figure 4. Therefore the attack goals are stated in terms of the attack impact on the system operation, while the constraints are related to the attack detectability.

The physical impact of an attack can be evaluated by assessing whether or not the state of the system remained in the safe set during and after the attack. The attack is considered successful if the state is driven out of the safe set. Attack constraints imply that attacks are constrained to remain stealthy. Denoting \( a = [a_i, \ldots, a_i] \) as the attack signal in the time interval \([k_0, k_f]\), and recalling that the residue signal is a function of the attack signal, a stealthy attack is defined as follows.

**Definition 3**

The attack signal \( a \) is *stealthy* over the time interval \([k_0, k_f]\) if the magnitude of the residue signal is smaller than the detection threshold, so that no alarm is triggered.

Below, it is assumed that the disruptive attack component consists of only data deception attacks and thus at time \( k \) the attack vector \( a_k = [\Delta u_k, \Delta y_k]^T \).

**Defense Methodology**

This subsection describes a common methodology to enhance a system’s cybersecurity, namely the risk management framework [35], [37], [38]. The main objective of risk
management is to assess and minimize the risk of threats, where the notion of risk is defined as follows [39].

**Definition 4**
Consider a given attack threat scenario, the corresponding impact to the system, and the likelihood of such scenario. The risk of the system is denoted as the set of triplets Risk = \{ (Scenario, Impact, Likelihood) \}.

The risk of different threat scenarios may be summarized in a two-dimensional risk matrix [38], where each dimension corresponds to the likelihood and impact of threats, respectively. Additionally, the risk of different threats may be compared

---

**The CIA in Networked Control Systems**

Three fundamental properties of information and services in IT systems are mentioned in the computer security literature [35] using the acronym CIA: confidentiality, integrity, and availability. Confidentiality concerns the concealment of data, ensuring it remains known only to the authorized parties. Integrity relates to the trustworthiness of data, meaning there is no unauthorized change to the information between the source and destination. Availability considers the timely access to information or system functionalities.

Figure S1 illustrates cyberattacks that violate each security property. In all three cases, the plant is sending the measurement vector \( \mathbf{y}_k = [2, 13]' \) to the controller through a communication network. This is a private message, hence only the plant and the controller should know the message contents.

In Figure S1(a), the adversary is able to eavesdrop on the communication, thus getting access to the contents of the message. Therefore confidentiality attribute was violated. Another scenario occurs in Figure S1(b), where the adversary succeeds in sending the false measurement vector \( \mathbf{y}_k = \mathbf{y}_k + \Delta \mathbf{y}_k \) to the controller, as if it was the plant sending it. Here data integrity is violated. In the final example, illustrated in Figure S1(c), the message sent by the plant is actually blocked and does not reach the controller. In this instance, data availability was compromised.

Whereas in IT systems the impact of such cyberattacks remains in the cyber realm, in networked control systems the impact may carry dire consequences for the physical side. Next, a specific example illustrating an attack scenario is presented, and its consequence to the physical system is discussed.

Consider a remotely controlled power generator, with \( \theta \) and \( \omega \) denoting its phase-angle and frequency deviation, respectively. Considering the single-machine infinite-bus model [68], the generator dynamics are described in continuous time by the normalized swing equation

\[
\omega(t) = \dot{\theta}(t), \\
M_\omega\dot{\omega}(t) = -D_\omega\omega(t) - P_I(t) + u(t),
\]

where \( u(t) \) is the normalized mechanical power provided to the generator and \( M \) and \( D \) are the inertia and damping coefficients, respectively. The term \( P_I(t) = b \sin(\theta(t)) \) corresponds to the electric power flow from the generator to the bus, where \( b \) is the susceptance parameter of the transmission line. Linearizing the model at \( \omega = \theta = 0 \) with \( M = D = b = 1 \), with the sampling period \( T_s = 1 \) s, and defining the discrete-time state \( \mathbf{x}_k = [\theta_k \ \omega_k]' \), the discrete-time model is

\[
\mathbf{x}_{k+1} = \begin{bmatrix} 0.66 & 0.53 \\ -0.53 & 0.13 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} 0.34 \\ 0.53 \end{bmatrix} \mathbf{u}_k,
\]

\[
\mathbf{y}_k = C_s \mathbf{x}_k,
\]

where \( C_s \) and \( \mathbf{u}_k \) is the control signal received on the plant side. Additionally, the system is safe if the frequency deviation \( \omega \) is small and the power flow \( P_I \) does not exceed the line ratings. In particular, the system is said to be safe if \( \|\omega_k\| \leq 0.05 \) and \( \|P_I\| = |\theta_k| \leq 0.1 \) for all \( k \). Defining the diagonal matrix \( T = \text{diag}(10, 20) \) and \( \mathbf{x} = [\mathbf{x}_0 \ldots \mathbf{x}_N]' \) as the state trajectory over the time interval \([0, N]\), the corresponding safe set is given by

\[
S_s = \{ \mathbf{x} : \|T \mathbf{x}_k\|_1 \leq 1, \forall k \in [0, N] \}.
\]

The anomaly detector corresponds to the state observer

\[
\mathbf{z}_{k+1} = (A_s - LC_s) \mathbf{z}_k + B_s \mathbf{u}_k + L \mathbf{y}_k,
\]

\[
\mathbf{r}_s = \mathbf{y}_k - C_s \mathbf{z}_k,
\]
through increasing functions of the threat’s impact and likelihood. As an example, Figure 5 illustrates a medium- and a high-risk threat with similar impacts but different likelihoods.

The risk management cycle, depicted in Figure 6, is composed of risk analysis, risk treatment, and risk monitoring. Risk analysis identifies threats and assesses the respective likelihood and impact on the system. Threats may be identified based on historical and/or empirical data of cyberattacks and known vulnerabilities in the system [38]. The likelihood of a given threat depends on the components compromised by the adversary in a given attack scenario and their respective vulnerability. Quantitative methods

where \( \tilde{y}_o \) is the measurement received at the anomaly detector side, \( z_k \) is the estimate of \( x_k \), the residue \( r_k \) is the output estimation error, and

\[
L = \begin{bmatrix} 0.36 & 0.27 \\ -0.31 & 0.08 \end{bmatrix}
\]

is the observer gain matrix design such that \( A_x - LC_x \) is stable. Denoting \( r = [r_1 \ldots r_N]^T \), an alarm is triggered by the anomaly detector when \( \| r \| > \delta = 0.01 \). The output feedback controller is given by

\[
z_{k+1} = (A_x - B_x K - LC_x) z_k + L \tilde{y}_k,
\]

\[
u_k = -K z_k,
\]

where \( K = [0.0556 \ 0.3306] \) is the controller gain matrix ensuring that \( A_x - B_x K \) is stable.

Consider an attack scenario where the adversary knows the exact model of the plant and is able to compromise the integrity of the control signal \( u_k \), that is, the mechanical power supplied to the generator, and the power flow measurement \( y_k = P_k = \theta_k \). Defining \( \tilde{u}_k = u_k + \Delta u_k \), \( \tilde{y}_k = y_k + \Delta y_k \), and the attack vector \( a_k = [\Delta u_k \ \Delta y_k]^T \) the plant under attack is described by

\[
x_{k+1} = A_x x_k + B_x u_k + B_a a_k,
\]

\[
\tilde{y}_k = C_x x_k + D_a a_k,
\]

(S1)

with

\[
B_a = \begin{bmatrix} 0.34 \\ 0.53 \end{bmatrix}, \quad D_a = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.
\]

The adversary attempts to drive the system to an unsafe state while remaining stealthy. To that end, the adversary injects an increasing piece-wise constant signal into the control input, making the generator produce more power, and thus increasing the power flow \( P \) along the transmission line. At the same time, the power flow increase is hidden from the controller and anomaly detector by tampering with the power flow measurement. More specifically, the attack policy is chosen as sequential instances of the zero-dynamics policy [13] during \( N \) time instants, where the attack vector is constructed as

\[
k_k = \left[ \frac{k+1}{N} \right],
\]

\[
a_k = k\lambda^k g,
\]

(S2)

with \( \lambda \in \mathbb{C} \) and \( g \neq 0 \) being the invariant zero and the corresponding input-direction satisfying

\[
\begin{bmatrix} \lambda I - A_x & -B_x \\ C_x & D_a \end{bmatrix} \begin{bmatrix} x_k \\ g \end{bmatrix} = 0.
\]

In the considered attack scenario, system (S1) has the zero \( \lambda = 1 \) on the unit circle, input direction \( g = \epsilon [-11]^T \), and \( x_k = \epsilon [-20 0]^T \) for \( \epsilon \neq 0 \). Notice that an input of the form \( a_k = \lambda^k g \) is blocked from the output at steady state by the zero \( \lambda = 1 \), yielding \( \lim_{x \to -\infty} \tilde{y}_k = 0 \).

For the generator’s closed-loop system with \( x_0 = z_0 = 0 \), choosing \( N = 10 \) and \( \epsilon = 0.01 \) and applying the attack policy (S2) results in the signals depicted in Figure S2. Observe that, at the end of the first instance, \( r_0 = 0 \), \( \theta_k = 0.01 \), and \( \omega_k = 0 \). Therefore, the second attack instance during the time interval \([N+1, 2N]\) begins with \( r_0 = 0 \) and also yields \( \|r_k\| \leq 0.0071 \) for \( k \in [N+1, 2N] \), as shown in Figure S2. Furthermore, the final value of the power flow is increased to \( P_{10} = \theta_{10} = 0.02 \). In fact, the attack policy (S2) yields \( \|r_k\| \leq 0.0071 \) for all \( k \) and \( P_{10} = \theta_{10} = 0.01 k \) for \( k = 1, 2, \ldots \), as depicted in Figure S2.

Thus, the adversary is able to drive the system to outside the safe set at \( k = 100 \) while remaining stealthy.

---

**Figure S2** Simulation results from the attack policy (S2) with \( N = 10 \), \( \lambda = 1 \), and \( g = [-0.01 \ 0.01] ^ T \). Plot (c) depicts the state trajectory under attack, the injected false-data is shown in (a), and (b) illustrates the corresponding residue signal and detection threshold.
Actions minimizing the risk of threats are determined within the risk treatment step. The different actions can be classified as prevention, detection, and mitigation. Prevention aims at decreasing the likelihood of attacks by reducing the vulnerability of the system components, for instance by encrypting the communication channels, using firewalls, and intelligent routing algorithms [28]. On the other hand, detection is an approach in which the system is continuously monitored for anomalies caused by adversary actions. Examples of detection schemes include antivirus software, network traffic analysis [41], and fault detection algorithms [31]. Once an anomaly or attack is detected, mitigation actions may be taken to disrupt and neutralize the attack, thus reducing its impact. The attack may be neutralized by replacing the compromised components or using redundant components.

The effectiveness of the defensive actions and the evolution of risk over time is evaluated throughout the risk monitoring stage. Risk monitoring continuously assesses the known and newly discovered vulnerabilities of the system, as well as the deployment of the threat mitigation actions.

Given the importance of risk analysis and risk treatment in the risk management process, the next sections illustrate and describe in detail methods that can be used for risk analysis and risk treatment in networked control systems.

**RISK ANALYSIS FOR STEALTHY DECEPTION ATTACKS**

Quantitative approaches to risk analysis of stealthy deception attacks are discussed in the remainder of this article. First, a simplified static case is analyzed in detail and illustrated with a power systems example. Then, the general dynamic case is presented and illustrated on a wireless quadruple-tank test bed.

Recall that the adversary aims to drive the system to an unsafe state while remaining stealthy. Additionally, the adversary also has resource constraints, in the sense that only a small number of attack points to the system are available. This section describes a framework for performing risk analysis of data deception attacks on networked control systems, where an attack is deemed less likely the more resources it requires. Particularly, the plant (1), feedback controller (2), and anomaly detector (3) are considered to be linear time-invariant systems. Defining $\eta_k = [x_k^T \ z_k^T \ s_k^T]^T$ and $a_k = [\Delta u_k^T \ \Delta y_k^T]^T$, the closed-loop dynamics of the networked control system driven by deception attacks are [13]

$$
\eta_{k+1} = A \eta_k + B a_k,
$$

$$
\tilde{y}_k = C_x \eta_k + D_x a_k,
$$

$$
r_k = C_s \eta_k + D_s a_k.
$$

### Risk Analysis for Static Models

The risk assessment in this subsection focuses on analyzing the threat’s likelihood, indicated by the minimum number of sensors that need to be compromised by the adversary for a given attack scenario. The minimum number of compromised sensors is a relevant indicator of
the threat’s likelihood because the sensors are often geographically distributed in networked control systems. As a result, coordinated attacks compromising multiple sensors need to be carried out simultaneously in different locations and are therefore difficult to implement.

The model in Figure 1 is simplified in two regards. First, the plant is in steady state. That is, in (5) the state vector \( \eta_k \) is constant for all \( k \), so the subscript \( k \) is omitted. The second simplification is that there is no feedback control. The simplifications are made because they can lead to a more streamlined illustration of the main concept of risk assessment. In addition, the simplified structure is relevant in its own right in analyzing the cyberphysical security of power systems. The risk assessment for general dynamic models will be deferred to a later section.

The model for risk assessment is the relationship between the static plant states \( x \) and the measurements \( \tilde{y} \) received by the anomaly detector. This is described by the expression

\[
\tilde{y} = C_yx + \Delta y,
\]

where \( C_y \) is the measurement matrix, and \( \Delta y \) is the measurement data attack. In a typical static state estimation problem, such as the power network case, there are more measurements than states, and hence \( C_y \) is assumed to have full column rank [42], [43]. Based on the risk assessment model, the least-squares estimate of the states is \( (C_y^TC_y)^{-1}C_y^T\tilde{y} \), and the estimate of measurements can be expressed as \( C_y(C_y^TC_y)^{-1}C_y^T\tilde{y} \). Thus, the anomaly detector, which is based on measurement residue, can be described by

\[
\begin{align*}
    r & \triangleq S\tilde{y} = (I - C_y(C_y^TC_y)^{-1}C_y^T)\tilde{y}. \\
\end{align*}
\]

Such an anomaly detector is, in general, sufficient to detect \( \Delta y \) in the form of a single error involving only one faulty measurement [42], [43]. However, in the face of a coordinated malicious attack on multiple measurements, the anomaly detector can fail. In particular, in [44] it was reported that an attack of the form

\[
\Delta y = C_y\Delta x
\]

for an arbitrary \( \Delta x \) would not result in any additional residue in (6), apart from the residue caused by other factors, such as measurement noise. In fact, the set of stealthy deception attacks with respect to the anomaly detector (6) and a zero detection threshold is characterized by (7), and these attacks were also experimentally verified in a realistic test bed [8]. Although stealthy attacks may be obtained from (7), distinct choices of \( \Delta x \) may yield attack vectors \( \Delta y \) requiring significantly different amounts of adversary resources, in terms of the number of nonzero entries of the attack vector \( \Delta y \). This number is also an indicator of the likelihood of the success of stealthy attack, as discussed earlier in this subsection. The rest of this subsection focuses on the characterization of the stealthy attack vectors with the minimum number of nonzero entries, as a concrete example of the quantitative method for risk assessment.

Minimum-Resource Attacks

There is a significant amount of literature studying the stealthy attack (7) and its consequences to state-estimation data integrity (for example, [15] and [44]–[49]). It was shown numerically that stealthy attacks \( \Delta y = C_y\Delta x \) are often sparse [44]. To analyze the stealthy attacks with the minimum number of nonzero entries, in [15] the notion of security index \( a_j \) for a measurement \( j \) was introduced as the optimal objective value of the following cardinality minimization problem

\[
\begin{align*}
    a_j & \triangleq \min_{\Delta x \in \mathbb{R}^n} \|C_y\Delta x\|_1 \\
    \text{subject to } & C_y(j,:)\Delta x \neq 0,
\end{align*}
\]

where \( \|C_y\Delta x\|_1 \) denotes the cardinality (that is, the number of nonzero entries) of the vector \( C_y\Delta x \). \( j \) is the label of the measurement for which the security index \( a_j \) is computed, and \( C_y(j,:) \) denotes the \( j \)th row of \( C_y \). The security index \( a_j \) is the minimum number of measurements an attacker needs to compromise to attack measurement \( j \) without being detected by the anomaly detector. In particular, a small \( a_j \) implies that measurement \( j \) is relatively easy to compromise in a stealthy attack, therefore indicating the higher likelihood of such a threat. As a result, the knowledge of the security indices for all measurements allows the network operator to pinpoint the security vulnerabilities of the network and to better protect the network with limited resources. For example, [45] proposed a method to optimally assign limited encryption protection resources to improve the security of the network based on its security indices.

The security index (8) is a quantitative tool for risk assessment that can provide a security assessment the standard detection procedure [42], [43] might not be able to
For each attack scenario, the attack policy is designed according to the adversary’s intent, namely the attack goals and constraints.

provide. As a concrete example [15], consider the measurement matrix

$$
C_y = \begin{pmatrix}
-1 & -1 & 0 \\
-1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & -1 \\
0 & -1 & 0
\end{pmatrix}.
$$

The “hat matrix” [42], [43], denoted $K$, captures how the received measurements $\hat{y}$ are weighted together to form a measurement estimate $\hat{x}$ and is defined according to

$$
\hat{y} = C_y \hat{x} = C_y (C_y^T C_y)^{-1} C_y^T \hat{y} \triangleq K \hat{y}.
$$

Corresponding to the $C_y$ in (9), is the hat matrix

$$
K = \begin{pmatrix}
0.6 & 0.2 & -0.2 & 0 & 0.4 \\
0.2 & 0.4 & -0.4 & 0 & -0.2 \\
-0.2 & -0.4 & 0.4 & 0 & 0.2 \\
0 & 0 & 0 & 1 & 0 \\
0.4 & -0.2 & 0.2 & 0 & 0.6
\end{pmatrix}.
$$

The rows of the hat matrix can be used to study the measurement redundancy [42], [43]. Typically a large degree of redundancy (many nonzero entries in each row) is desirable to compensate for noisy or missing measurements. In (10), all measurements are redundant in this example except the fourth. Such nonredundant measurement is called a critical measurement. Without the critical measurement, observability is lost, meaning that it becomes impossible to uniquely determine the states based on the available measurement information. The hat matrix indicates that the critical measurement is sensitive to attacks. This is indeed the case, but some other measurements can also be vulnerable to attacks. The security indices $\alpha_j$, $j = 1, \ldots, 5$, respectively, are 2, 3, 3, 1, 2. Therefore, the fourth (critical) measurement has security index one, indicating that it is vulnerable to stealthy attacks. However, the first and the last measurements also have relatively small security indices. This is not obvious from $K$ in (10). Hence, the information of the security indices can enhance the vulnerability analysis compared to the hat matrix.

Because of the cardinality minimization, computing the security indices can sometimes be hard. In fact, it can be established that problem (8) is NP-hard using techniques from [50] and [51]. As a result, known exact solution algorithms for (8) are enumerative by nature. Three different typical exact algorithms include a) enumeration on the support of $C_y \Delta x$, b) finding the maximum feasible subsystem for an appropriately constructed system of infeasible inequalities [52], and c) the big $M$ method (for example, [53]). This article focuses on the big $M$ method because the resulting optimization problem can be modeled and solved using available software such as CPLEX. The big $M$ method sets up and solves the following optimization problem.

$$
\begin{align*}
\text{minimize} & \quad \sum_i w(i) \\
\text{subject to} & \quad C_y \Delta x \leq Mw, \\
& \quad -C_y \Delta x \leq Mw, \\
& \quad C_y(j,:) \Delta x = 1, \\
& \quad w(i) \in [0,1] \text{ for all } i.
\end{align*}
$$

In (11), the inequalities are interpreted entry-wise, and $0 < M < \infty$ is a user-defined constant scalar. If $M$ is greater than the maximum entry of $C_y \Delta x$ in absolute value, for some optimal solution $\Delta x^*$ of (8), then the optimal solution to (11) is exactly an optimal solution to (8). Otherwise, solving (11) yields a suboptimal solution, optimal among all solutions $\Delta x$ such that the maximum entry of $C_y \Delta x$ is less than or equal to $M$ in absolute value. The procedure described in [54] can always find a sufficiently large $M$ to ensure that the big $M$ method indeed provides the optimal solution to (8). In addition, the physics and insights of the underlying application problem can also lead to a suitable $M$. The optimization problem in (11) is a mixed integer linear programming (MILP) problem; see “Mixed Integer Linear Programming” for additional details.

For large-scale system analysis, it might be deemed impractical to obtain the exact solution to the security index problem in (8). In this case, it might be necessary to settle for an approximate solution instead. A particular method to obtain an approximate solution is $\ell_1$ relaxation. For general information about $\ell_1$ relaxation; see, for example, [55]–[57]. Here the properties that are most relevant to this article are described. Instead of solving (8), the $\ell_1$ relaxation method sets up and solves the following optimization problem:

$$
\begin{align*}
\text{minimize} & \quad \|C_y \Delta x\|_1 \\
\text{subject to} & \quad C_y(j,:) \Delta x = 1, \\
& \quad \Delta x \in \mathbb{R}^{n_x}.
\end{align*}
$$

where $\|C_y \Delta x\|_1$ denotes the vector $\ell_1$-norm (sum of absolute values of the entries) of $C_y \Delta x$. In addition, the right-hand side of the constraint in (12) needs to be normalized to ensure that the problem is well posed. Problem (12) can be
Mixed Integer Linear Programming

A mixed integer linear programming (MILP) problem is an optimization problem over both real and integer decision variables with a linear objective function and linear constraints. It is basically an LP problem except that some of the decision variables are integer valued. In general, an MILP problem can be written as

\[
\begin{align*}
\text{minimize} & \quad c'x + d'y \\
\text{subject to} & \quad Ax + By = b, \\
& \quad x, y \geq 0, \\
& \quad x \text{ integer},
\end{align*}
\]

where \(A, B, b\) are matrices or a vector with commensurate dimensions. If the integer constraint is relaxed, an MILP problem becomes an LP problem. The MILP problem has many applications (for example, [53]). For instance, \(0 \rightarrow 1\) binary decision variables can be used to model logical “on–off” decisions. If \(x_1\) and \(x_2\) are both \(0 \rightarrow 1\) binary decision variables, then the constraint \(x_1 + x_2 = 1\) means that either \(x_1 = 1\) or \(x_2 = 1\) but not both. This modeling capability is not available with an LP, because LP decision variables can take fractional values. Another well-known example of MILP modeling is the traveling salesman problem, where a map of cities and pairwise distances between cities are given and the salesman has to make a shortest-distance tour visiting each city exactly once. The traveling salesman problem has many important applications including circuit-board drilling and DNA sequencing. The MILP model of the traveling salesman problem cannot be relaxed to an LP model since the decision of whether or not a road is traversed is a binary one. The MILP problem is NP-hard, as it includes as a special case the \(0 \rightarrow 1\) integer program. As a result, unless \(P = NP\), it is impossible to find a polynomial-time algorithm to solve the MILP problem. This implies that the computational effort for solving MILP problems in general increases very rapidly as the size of the problem increases. For example, suppose that a basic computation requires \(10^{-9}\) s to perform on a computer. On a graph with \(|V| = 30\) nodes and \(|E| = |V|(|V| - 1)/2\) edges, to solve the traveling salesman problem by enumeration requires \(O(2^{|V|})\) basic computations, or about \(2 \times 10^{122}\) years. On the other hand, for the same graph, if instead the minimum cut problem is solved with a polynomial-time algorithm that requires \(O(|V||E| + |V|^2 \log(|V|))\) basic computations, then the solution time is only about 17 ms. Nevertheless, solution algorithms for the MILP problem are well studied and well developed. They include, for instance, branch-and-bound methods and cutting-plane methods. Software implementations of MILP problem solution algorithms include, for example, CPLEX [69] and Gurobi [70].

written as a linear program. Hence, it can be solved efficiently to obtain an exact optimal solution. The optimal solution to (12) is feasible to problem (8) because \(C_y(j,:) Ax = 1\) implies \(C_y(j,:), Ax \neq 0\). Therefore, the optimal solution to (12) is an approximate solution to (8), with the former leading to an objective value that is greater than or equal to the true minimum of (8). Therefore, the \(\epsilon_1\)-relaxation approach provides an overestimate of the security index.

An alternative approach to handle the large-scale system computation difficulty is to develop specialized algorithms for particular instances of (8). For example, when the underlying application is power network state estimation and when the measurement system satisfies certain assumptions, such as the full measurement assumption to be described, problem (8) can be solved exactly in a time-efficient manner. The details of this result and an illustration with large-scale numerical examples when applied to electric power systems will be given in the following section.

Risk Analysis and Treatment for Electric Power Network

Power transmission networks are complex and spatially distributed systems. They are operated through SCADA systems, which represent the backbone IT and control infrastructure, as illustrated in Figure 7. SCADA systems collect data from remote terminal units (RTUs) installed in substations and relay aggregated measurements to the central master station located at the control center. The technological limitations of legacy measurement equipment limits the sampling periods to the order of tens of seconds, thus the system is mainly observed at a quasi-static state.

SCADA systems for power networks are complemented by a set of application-specific software, usually called energy management systems (EMSs). EMSs enable state and measurement estimation and optimal operation under safety and reliability constraints by providing human operators with state-awareness and recommended control actions. In the past, the malfunction of EMS components, in particular the state estimator, has led to a large-scale blackout with severe economic consequences [58]. Furthermore, as discussed in [2], there are several vulnerabilities in the SCADA system architecture, including the direct tampering of RTUs, communication links between the RTUs and the control center, and the IT software and databases in the control center. Thus cybersecurity of SCADA and EMS in power networks is of major importance.

Given the relevance of power networks, in this part of the article the risk assessment method described in the previous section on static models will be specialized to the case where the plant is an electric power network. Focusing on the power network case enables the risk assessment to be performed in a computationally efficient manner. In addition, some of the risk treatment tools for power
network applications will be highlighted. At the end of this section, a numerical case study with IEEE benchmark systems will illustrate the effectiveness of the risk assessment tools described in this section.

DC Power Flow Measurement Model
Assume that the electric power network has \(n+1\) buses and \(L\) transmission lines. The state of the network is determined by the complex voltages at the buses, whose magnitudes and phase angles are, respectively, denoted by \(V_i\) and \(\delta_i\) for \(i = 0, 1, \ldots, n\). In power networks, commonly considered measurements include line power flows, bus power injections, bus voltage magnitudes, and line current flow magnitudes. This section focuses only on active power flows on transmission lines and active power injections at buses, which are functions of the bus voltage magnitudes and phase angles. However, for the analysis of cyberphysical security, bad-data detection, and network observability, it is customary to describe the dependency of active power flows and injections through an approximate model called the dc-power flow model. By assuming that the voltage magnitudes \(V_i\) are all fixed to 1 p.u. (that is, unity in the per unit system [42]), the dc power flow model depends only on the voltage phase angles. In this model, the transmission-line active power flow from bus \(i\) to bus \(j\) is

\[
P_{ij} = \frac{x_{ij} V_i V_j}{x_0} \tag{13}
\]

where \(x_{ij} = x_i - x_j\) and \(x_0 > 0\) is the reactance of the line between bus \(i\) and bus \(j\). On the other hand, the active power injection at bus \(i\) is

\[
P_i = \sum_{j \in N_i} P_{ij} \tag{14}
\]

where \(N_i\) is the set of all indices of the neighboring buses of bus \(i\), excluding \(i\).

Equations (13) and (14) give rise to a linear measurement model in matrix-vector form. Let \(x\) denote the \(n\)-vector of voltage phase angles on all buses except the reference bus. The reference bus is arbitrarily defined, with voltage phase angle fixed at zero. In addition, let \(y\) denote the vector of active power flow and active power injection measurements. Then, \(y\) and \(x\) are related by the equation

\[
y = \begin{bmatrix} T_x & D \end{bmatrix} x = C y x. \tag{15}
\]

In (15), the term \(T_x D A_i^T x\) corresponds to transmission line power flow measurements. On the other hand, the term \(T_x A_i D A_i^T x\) corresponds to power injection measurements. The symbols in (15) are as follows: \(A_i \in \mathbb{R}^{(n+1) \times L}\) is the incidence matrix of the network defined as

\[
A_i(i,l) = \begin{cases} 1 & \text{line } l \text{ starts at bus } i, \\ -1 & \text{line } l \text{ ends at bus } i, \\ 0 & \text{otherwise}, \end{cases} \text{ for each transmission line } i.
\]

The directions of the lines in \(A_i\) are irrelevant to the application in this article. They can be fixed arbitrarily. Matrix \(A_i\) is the truncated incidence matrix with the row of \(A_i\) corresponding to the reference bus removed. Matrix \(D\) is a diagonal matrix with the diagonal entries being the reciprocals of \(X_{ij}\) for all lines. Matrices \(T_x\) and \(T_y\) are stacked by the rows of identity matrices, and they indicate which line power flows and bus power injections are actually measured. The total number of rows of \(T_x\) and \(T_y\) is the total number of measurements, which is denoted by \(m\). The matrix \(C_y\) is again referred to as the measurement matrix.

The measurement model in (15) has a network potential flow interpretation. A particular \(x\) corresponds to an
assignment of the bus voltage phase angles. The phase angle differences between two neighboring buses induce power flows along the connecting lines. The vector of induced line power flows is described by $DA^T x$. At each bus, any difference between total incoming and total outgoing line power flows has to be balanced by an external power injection or extraction. The vector of external power injections at the buses is described by $A_bDA^T x$. The matrices $T_1$ and $T_2$ allows the flexibility that all line power flows and bus power injections are not measured. The network potential flow interpretation is illustrated in Figure 8.

Security Index Problem Under dc Power Flow Model

In this section, unless otherwise noted, the security index problem in (8) should be interpreted with the measurement matrix $C_y$ restricted to the form in (15). With the restriction of $C_y$, problem (8) has a network potential flow interpretation. The vector $Ax$ can be considered as an assignment of fictitious voltage phase angles. The constraint in (8) states that either a particular line has a nonzero flow or a particular bus has a nonzero injection, depending on the meaning of measurement $j$. The objective is to minimize $\|C_y Ax\| = \|T_1 DA^T Ax\| + \|T_2 A_b DA^T Ax\|$, being the sum of the number of lines with nonzero measured flows and the number of buses with nonzero measured injections.

Even though $C_y$ is specialized, (8) is still NP-hard [59]. However, (8) becomes solvable in polynomial time under certain additional assumptions. One such example is when $T_1$ is an identity matrix and $T_2$ is a zero matrix. In this case, minimizing the number of lines with nonzero phase angle differences is the only objective. It can be seen that an assignment of $Ax$ using only two distinct values (say zero and one) is optimal. In fact, the above assumption can be generalized: the $0-1$ assignment remains optimal if both $T_1$ and $T_2$ are identity matrices of appropriate dimensions [59]. In this case, all line power flows and bus injections are measured and the condition is referred to as the full measurement assumption. In summary, under the full measurement assumption, (8) is equivalent to

$$\begin{align*}
\text{minimize} & \quad \|C_y Ax\|_0 \\
\text{subject to} & \quad C_y(j,:)Ax \neq 0.
\end{align*}$$

(16)

The only difference between (8) and (16) is that the decision vector of real numbers in (8) is replaced by the decision vector of 0–1 binary values in (16).

The $0-1$ assignment of the entries of $Ax$ in (16) leads to yet another graph interpretation. The binary choice of entries of $Ax$ specifies a partitioning of the buses into two disjoint sets, a set with buses with fictitious voltage phase angles being zero and the complementary set. A line connecting two buses in two different sets is cut. The objective is to minimize the sum of the number of cut lines with line power flow meters and the number of buses that have injection meters and are incident to at least one cut line. The constraint can also be described as a particular line being cut [59]. As a result, (16) can be interpreted as a generalization of the standard minimum cut problem; see “Standard Minimum Cut Problem.” The only difference is that the standard minimum cut problem does not consider the cost associated with the number of buses incident to cut lines. The generalized minimum cut problem is illustrated in Figure 9.

The optimal solution to the generalized minimum cut problem can be interpreted in an alternative way: if all transmission lines in the cut are removed, then the remaining network contains at least two isolated subnetworks. This, in fact, implies that the measurement system with the compromised measurements removed is unobservable because it is impossible to deduce the phase angle information of one subnetwork from the information of another subnetwork [60]. In fact, in [26] it was shown that a set of compromised measurements is the optimal solution to the security index problem (8) if and only if the set is a sparsest critical tuple containing the target measurement $j$. A critical tuple is a set of measurements whose removal renders the remaining measurement system unobservable, but the removal of any strictly proper subset of the critical tuple would not lead to loss of observability. The interpretation of the security index problem optimal solution set as a critical tuple is illustrated in Figure 10. While this article focuses on interpreting the security index problem as a power network observability problem, the converse interpretation can be utilized to solve an observability analysis problem as a security index problem; see [61] for details.
A standard minimum cut problem (on an undirected graph) is an optimization problem whose instance is defined by an undirected graph with nonnegative edge weights and two distinct nodes called source and sink; see Figure S3 for an illustration.

The problem seeks a partition of the set of all nodes into two parts, with the source in one part and the sink in the other part, so that the sum of the weights of cut edges is minimized. An edge is cut if and only if it has one end node in the partition part including the source and the other end node in the partition part including the sink; see Figure S4 for an illustration. As an application, consider the network in Figure S3 as a power network $s$ is the supply (generator) and $t$ is the demand (load), respectively. If all edge weights are set to unity, then the solution to the minimum cut problem specifies the minimum number of transmission lines to break to cause the disruption of power supply.

Efficient solution algorithms are available to solve the standard minimum cut problem with computation effort proportional to a polynomial function of the size of the problem. The algorithms can be direct [62], or based on solving the dual maximum flow problem enabled by the max-flow, min-cut theorem [71].

The generalized minimum cut problem can be solved in polynomial time, due to its connection to the standard minimum cut problem, which is polynomial-time solvable [62]. The main result is that an instance of the generalized minimum cut problem on an auxiliary graph that is roughly three times as large as the original graph [59]. Additionally, the auxiliary graph can be constructed efficiently as discussed in “Mixed Integer Linear Programming,” the minimum cut problem has much lower computational complexity than enumeration methods. Therefore, the generalized minimum cut formulation can potentially be applied to large problems, as illustrated in the case studies reported at the end of this section.

In summary, given an instance of the security index problem in (8) under the full measurement assumption or similar conditions found in [59], it can first be specialized by restricting the entries of $\Delta x$ to either zero or one, as in problem (16). The specialized problem can be viewed as an instance of the generalized minimum cut problem. Then, an instance of the standard minimum cut problem on an auxiliary graph can be set up and solved in polynomial time. The solution to the standard minimum cut problem can be used to construct the solution to the security index problem due to the equivalence mentioned above. Notice that, even without the full measurement assumption, the solution to problem (16) still serves as an overestimate of the true security index. In particular, the security index of a measurement must be small if the overestimate is small. This indicates the vulnerability of the measurement. The accuracy of this approach was demonstrated in [59].

Risk Treatment Approaches

One possible approach to decrease the risk of stealthy deception attacks is to encrypt the data and communication channels. Since a large part of today’s power grid equipment is...
old, data encryption can be costly to implement because of the corresponding update of the equipment. Therefore, the following question is of great importance to measurement data integrity. Given limited protection resources (the number of devices for data encryption), which measurements should be encrypted to maximize the benefits of the protection resources? The risk analysis outcome from computing the measurements’ security indices may be used to sort the measurements in terms of their vulnerability and identify those that should be protected. In fact, a variant of the security index problem can help provide an answer to the previous question, namely

\[
\begin{align*}
\minimize_{\Delta x \in \mathbb{R}^n} & \quad \| C_x \Delta x \| \\
\text{subject to} & \quad C_y(j,:)\Delta x \neq 0, \\
& \quad C_y(P,:)\Delta x = 0, \\
& \quad (17)
\end{align*}
\]

where \( P \) is the index set of the encrypted measurements, which cannot be attacked. By comparing the security indices for different index sets \( P \), it is possible to evaluate the effect of different protection strategies and determine the best one to implement. For example, [28] considers a lexicographic optimization of some security metrics that are based on the security index computation related to (17).

In the case where it is impractical to encrypt all measurements, it becomes critical to detect and isolate the measurements that are under attack. Effective attack isolation enables the damage control (for example, removing attacked measurements for state estimation) to be performed in a timely fashion, that is, before the attack can lead to an incident with significant consequences. A distributed procedure for isolating the data attacks on power system transmission line power flow measurements is presented in [63], based on secure bus voltage magnitude measurements. The work in [18] develops a generalized likelihood ratio test to detect the presence of data attack, based on the assumption that the normal measurements follow a known Gaussian distribution. Mechanisms to detect data attacks based on known-secure phasor measurement unit PMU measurements and a known pattern of system states are presented in [29].

Security Index Problem Case Studies on the Benchmark Systems

As a numerical demonstration, the security index problems for all measurements in two benchmark systems are considered. The two benchmarks are the IEEE 14-bus and IEEE 118-bus systems [64]. In this case study, all line power flows and bus injections of the benchmark systems are measured. In other words, the full measurement assumption holds and

![Figures 9 and 10](image-so.png)
the security index can be computed by solving (16). Figure 11 shows the IEEE 14-bus system with measurements color coded to indicate which ones are more vulnerable to stealthy data attack. The vulnerable measurements have small security indices (< 7) and are color coded red. The resilient measurements have large security indices (> 7) and are color coded green. The other measurements are color coded blue and their resilience lies somewhere in between. The efficient computation of security indices enables the rapid determination of the security weak points in the measurement system.

**Figure 12** The IEEE 118-bus benchmark system [64].

**Risk Analysis for Dynamic Models**

Having discussed methods for threat likelihood estimation in the static case, the dynamic case and methods for more general quantitative risk assessment are considered. The proposed risk assessment methods are not executed based on real-time data of the system. Instead, these methods are used offline, for a given configuration and corresponding model of the system, to assess the risk of different hypothetical attack scenarios. Consider the time interval [0, N] and define the vectors \( \mathbf{n} = [\eta_0 \ldots \eta_N]^T \) and \( \mathbf{a} = [a_0 \ldots a_N]^T \), which capture the system states and attack signals over the time interval of interest. The state and residue trajectories are described by the following static mappings obtained (5)

\[
\begin{align*}
\mathbf{n} &= O \eta_0 + T \mathbf{a}, \\
\mathbf{x} &= C \mathbf{n}, \\
\mathbf{r} &= C \mathbf{n} + D \mathbf{a},
\end{align*}
\]

where \( O \) describes the effect of the initial condition \( \eta_0 \) on the system’s state trajectory, and \( T \) is a lower triangular
block-Toeplitz matrix mapping the attack signals to the state trajectory; see [27] for details. For simplicity, \( \eta_1 \) is assumed to be zero. These mappings can be used for risk assessment of deception attacks on networked control systems, as described in the remainder of this section.

**Maximum-Impact, Minimum-Resource Attacks**

For illustration purposes, the quantitative risk assessment methods in the previous section considered only the adversary resources. In this subsection, full risk assessment is performed on dynamical systems by simultaneously considering impact and resources. Recall that the adversary aims to perturb the networked control system operation and drive the system to an unsafe set. To better clarify the proposed approach, suppose that the safe set is defined as

\[
S_\delta = \{ x : \| x \|_p < \delta \},
\]

where \( \| x \|_p \) denotes the \( \ell_p \)-norm of \( x \) for \( p \geq 1 \), and \( x \) depends on the attack signal \( a \) as described in (18). Therefore, the attack impact during the time interval \([0, N]\) can be characterized as the perturbation of the state trajectory \( x \) due to the attack quantified by \( \| x \|_p \). Notice that the proposed framework can be straightforwardly applied to safe sets that consider linear transformations of the state trajectory.

To illustrate the resources required for a given attack signal, suppose that \( \Delta u_k \) and \( \Delta y_k \) are scalars and recall the attack vector at time \( k \), \( a_k = [\Delta u_k \, \Delta y_k]^T \). Consider an attack having \( \Delta u_k \) equal to zero for all \( k \), while \( \Delta y_k \) is equal to zero at all times except for \( k = 0 \). Since \( \Delta y_0 \) is nonzero, the adversary must have access to one communication channel to inject the false data, namely the measurement channel. The attack with \( \Delta y_k \neq 0 \) only at \( k = 0 \) requires as many resources as the one having \( \Delta y_k \neq 0 \) for \( k > 2 \), thus corrupting the measurement signal at other times does not require additional adversary resources. On the other hand, corrupting also the control signal such that \( \Delta u_k \neq 0 \) at a given time \( k \) requires an additional resource, namely the access to the control signal. More generally, the attack vector can be rewritten as \( a_k = [a_{0(1, k)}, \ldots, a_{0(n, k)}]^T \), where \( a_{0,k} \in \mathbb{R} \) denotes the corrupted data introduced in the \( i \)th adversary resource at time \( k \). Denoting the attack signal at the \( i \)th resource by \( a_{i0} = [a_{0(1, k)}, \ldots, a_{0(n, k)}]^T \), the \( i \)th resource is used during the attack if \( \| a_{i0} \|_p \) is nonzero. Defining the vector

\[
h_p(a) = [\| a_{0(1, k)} \|_p \ldots \| a_{0(n, k)} \|_p]^T,
\]

the number of resources used in a given attack corresponds to the number of nonzero elements of \( h_p(a) \), which is denoted as \( \| h_p(a) \|_e \).

The attack impact and resources are jointly considered in the multiobjective optimization problem [65]

\[
\begin{align*}
\text{maximize} & \quad (\| x \|_p, -\| h_p(a) \|_e) \\
\text{subject to} & \quad \| r \|_e \leq \delta,
\end{align*}
\]

where \( x \) and \( r \) are determined by \( a \) according to (18). The multiobjective optimization problem (19) can be interpreted as computing the attack signal \( a \) that simultaneously maximizes the impact \( \| x \|_p \) and minimizes the resources \( \| h_p(a) \|_e \), while remaining stealthy by ensuring \( \| r \|_e \leq \delta \).

The tradeoff analysis in the multiobjective optimization problem can be performed by studying the Pareto solutions [65]. These solutions can be obtained through several techniques, for instance, the bounded objective function method in which all but one of the objectives are posed as constraints, thus obtaining a scalar-valued objective function. Applying this method to (19) and constraining \( \| h_p(a) \|_e \) yields

\[
\begin{align*}
\text{maximize} & \quad |x| \\
\text{subject to} & \quad |r| \leq \delta, \\
& \quad |h_p(a)|_e \leq \epsilon, \\
& \quad n = Ot_0 + Ta, \\
& \quad x = C_n, \\
& \quad r = C_n + D_a,
\end{align*}
\]

where \( \epsilon \) is a fixed \( \in [1, \ldots, n_r] \). For a fixed \( \epsilon \) and \( p = q = \infty \), the optimization problem (20) can be formulated as an MILP, while the problem with parameters \( \epsilon = \infty \) and \( p = q = 2 \) reduces to a generalized eigenvalue problem; see [27] for a detailed discussion.
Case Study: Quadruple-Tank Process

Risk analysis and risk treatment approaches for dynamic control systems are illustrated for a particular networked control system in this section. The physical plant consists of the quadruple-tank process (QTP) [66], depicted in Figure 14, while the networked control system architecture is depicted in Figure 15.

The plant model is

\[
\frac{dh_1}{dt} = -\frac{a_1}{A_1}\sqrt{2gh_1} + \frac{a_3}{A_3}\sqrt{2gh_3} + \gamma_1 k_1 u_1,
\]

\[
\frac{dh_2}{dt} = -\frac{a_2}{A_2}\sqrt{2gh_2} + \frac{a_4}{A_4}\sqrt{2gh_4} + \gamma_2 k_2 u_2,
\]

\[
\frac{dh_3}{dt} = -\frac{a_3}{A_3}\sqrt{2gh_3} + \frac{(1-\gamma_2)k_2}{A_2} u_2,
\]

\[
\frac{dh_4}{dt} = -\frac{a_4}{A_4}\sqrt{2gh_4} + \frac{(1-\gamma_1)k_1}{A_4} u_1,
\]

where \( h_i \in [0, 30] \) are the heights of water in each tank, \( A_i \) is the cross-sectional area of each tank, \( a_i \) is the cross-sectional area of each tank’s outlet, \( k_i \) is the pump constant, \( \gamma_i \) the flow ratios, and \( g \) is gravitational acceleration. The nonlinear plant model is linearized for a given operating point and the state of the linearized plant model \( x \) corresponds to the water level deviations from the operating point. Moreover, given the range of the water levels and the operating point, the safe set is considered to be \( \mathcal{S}_x = \{ x \in \mathbb{R}^n : \| x \| \leq 5 \} \).

The process is controlled using a centralized linear-quadratic-Gaussian (LQG) controller with integral action running on a remote computer and a wireless network is used for the communications. A Kalman filter-based anomaly detector is also running on the remote computer and alarms are triggered according to (4). The measurements of the lower tanks’ water levels, \( y_1 \) and \( y_2 \), are sent to the controller and anomaly detector through the communication network. Given the received measurements \( \tilde{y}_1 \) and \( \tilde{y}_2 \), the LQG controller computes the pump control signals, \( u_1 \) and \( u_2 \), which are sent to the water pumps through the wireless network. In the present attack scenario, the adversary can corrupt the data transmitted through the wireless network, namely \( u_1 \), \( u_2 \), \( y_1 \), and \( y_2 \).

Risk Analysis for Stealthy Deception Attacks

For the time interval \([0, 50]\), the maximum-impact, minimum-resource attacks were computed for the process in minimum and nonminimum phase settings by choosing \( p = q = 2 \) and iteratively solving (20) with respect to \( \epsilon \). The respective impacts correspond to the energy of the state signal \( x \) for value of \( \epsilon \), and are presented in Table 1, while...
the risk is depicted by the risk matrix plot in Figure 16(a) (recall Figure 5).

As expected, the nonminimum phase system is less resilient than the minimum-phase one. In both settings, the attack impact can be made arbitrarily large by corrupting three or more channels, and thus the adversary can drive the state out of the safe set while remaining stealthy. The results indicate that the threats compromising three or more channels have high risk and should therefore be analyzed in more detail. The risk of such threats can be mitigated by protecting the data channels, which is shown in the next subsection.

For illustrative purposes, the maximum-impact attack signal for the nonminimum phase system with \( \epsilon = 2, \delta = 0.15, \) and \( p = q = 2 \) is presented in Figure 17(a). For the parameters \( \epsilon = 2, \delta = 0.025, \) and \( p = q = \infty, \) the maximum-impact attack signal shown in Figure 17(b) was computed by solving an MILP; see [27]. In both cases the optimal attack corrupts both actuator channels and ensures \( \| r \|_\infty \leq \delta, \) while maximizing the attack impact. Although the impact results in Table 1 do not quantify the impact according to the safe set \( S = \{ x \in \mathbb{R}^n : \| x \|_p \leq 5 \}, \) the state trajectory does leave the safe set in both cases.

The attack signals illustrated in Figure 17 are related to the zero dynamics of the QTP system. The zero-dynamics attack signal and other scenarios were analyzed and performed in an experimental test bed of the QTP by [13]. Videos of the experiments are available [67].

### Risk Treatment Approaches

The risk analysis identifies the data channels that, when corrupted, may lead to a large impact on the system. The subsequent step in the risk management framework is the risk

<table>
<thead>
<tr>
<th>Number of Compromised Channels</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum phase</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>140.39</td>
<td>1.15</td>
</tr>
<tr>
<td>Nonminimum phase</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>689.43</td>
<td>2.8</td>
</tr>
</tbody>
</table>

The risk analysis results for the quadruple-tank process. Each entry corresponds to the maximum impact \( \| x \|_p \) for a given number of corrupted channels, computed through (20), with \( p = q = 2 \) and \( \delta = 0.15. \)

The figure in (b) indicates that, when pairs of resources can be protected in the nonminimum phase process, the most effective choice for risk treatment is to protect both actuator channels, \( \{ u_1, u_2 \}. \)
treatment stage, in which actions reducing the risk are chosen and implemented. A common approach to decrease the risk of threats is to deploy protective resources such as encryption, thus preventing the attacks from occurring. To assess the effectiveness of protecting a given set of data channels $P$ the optimization problem (20) may be modified to

$$
\text{maximize} \quad \| \mathbf{x} \|_p \\
\text{subject to} \quad \| \mathbf{r} \|_p \leq \delta, \\
\| h_{ij}(\mathbf{a}) \|_p < \epsilon, \\
(18), \\
\mathbf{a}_i = 0, \text{ for all } i \in P.
$$

The QTP example is now considered to illustrate the risk treatment step using channel encryption. The preventive action under study is the encryption of one pair of data channels so that the risk is minimized. The optimization problem (21) is solved for each pair of data channels, and the corresponding risk matrices plots are depicted in Figure 16(b).

From the results in Figure 16(b), it is evident that the pair of actuators $\{u_1, u_2\}$ should be protected to minimize the risk. Moreover, recalling the original risk matrix plot in Figure 16(a), observe that the impact when two channels are corrupted is substantially decreased by protecting $\{u_1, u_2\}$. This protection choice is expected since the adversary can no longer inject an attack exciting the unstable zero dynamics of the system when both actuators are protected. Furthermore, since the resources accessible to the adversary are $y_1$ and $y_2$, the adversary cannot have a direct impact on the physical system but instead needs to affect the system through the feedback controller by corrupting the measurement signals.

Methods other than encryption have been proposed in the literature to reduce the risk of threats. Concerning replay attacks, [36] proposes the use of a hypothesis test as the anomaly detector and the injection of random, zero-mean Gaussian noise with an optimally designed covariance in the control input channels. The injected noise increases the performance of the hypothesis test since the noise statistics are assumed to be unknown to the adversary. Similarly, [30] proposes the insertion of uncertainty in the adversary’s model knowledge by modifying the system dynamics and control and output channels. The effects of such actions on zero-dynamics attacks are also characterized in detail.
CONCLUSIONS AND FUTURE WORK
The pervasive use of IT infrastructures supporting the operation of networked control systems has introduced vulnerabilities into these systems, raising numerous challenges regarding the cyberphysical security of networked control systems. The specific nature of these threats and the coupling between the cyber and physical realms of the system requires the development of new paradigms and frameworks to study and tackle security-related problems.

This article described a cybersecurity problem in networked control systems, covering some of the key aspects such as the networked control-system architecture, the adversary model, and the defense methodology. The networked control-system architecture consists of the physical plant, the feedback controller colocated with the anomaly detector, and the communication network, through which the measurement and control data are sent. The adversary is modeled by a resource-constrained policy with limited model knowledge, disruption resources, and disclosure resources. Moreover, the attack policy is shaped according to the adversary’s intent: to drive the state of the physical plant to an unsafe region while remaining undetected by the anomaly detector. The defense methodology is based on the risk management framework, where the concept of risk is defined as a function of the threat’s likelihood and the threat’s impact to the system. The risk management cycle continuously minimizes the risk of threats by performing risk analysis, risk treatment, and risk monitoring.

Recent quantitative methods developed for risk analysis are also presented, which are of major importance to enhance the cyberphysical security of networked control systems. The problem of quantifying the likelihood of threats for static systems is discussed, and computationally efficient methods are described and illustrated for large-scale electric power networks. Possible risk treatment approaches proposed in the literature are also mentioned.

Finally, the full risk analysis problem for dynamic systems simultaneously considering the attack impact and likelihood is described. The quantitative risk analysis methods are formulated and illustrated on the wireless QTP test bed. Additionally, the effectiveness of protection-based risk treatment schemes is evaluated in terms of their effect on the risk of threats, and alternative approaches proposed in the literature are summarized.

Possible extensions to the work described in this article include the study of efficient tools for risk assessment of dynamic systems and the analysis of attack and defense policies under noise and uncertain communication channels, among several other interesting directions.

ACKNOWLEDGMENTS
The research leading to these results received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 608224, the Swedish Research Council under grants 2009-4565 and 2013-5523, and the Knut and Alice Wallenberg Foundation.

AUTHOR INFORMATION
André Teixeira (andretei@kth.se) is a Ph.D. candidate in automatic control at KTH Royal Institute of Technology, Stockholm, Sweden. He received the M.Sc. degree in electrical and computer engineering in 2009 from the Faculdade de Engenharia da Universidade do Porto, Portugal. He was a finalist for the NecSys 2012 Best Student Paper Award and one of his publications is listed in ACM Computing Review’s Notable Computing Books and Articles of 2012. His main research interests include secure control, distributed fault detection and isolation, distributed optimization, power systems, and multiagent systems. He can be contacted at the Department of Automatic Control, KTH School of Electrical Engineering, Osquldas väg 10, plan 6, SE-100 44 Stockholm, Sweden.

Kin Cheong Sou received a Ph.D. degree in electrical engineering and computer science at the Massachusetts Institute of Technology in 2008. From 2008 to 2010 he was a postdoctoral researcher at Lund University, Sweden. From 2010 to 2013 he was a postdoctoral researcher at KTH Royal Institute of Technology, Stockholm, Sweden. Since 2013 he has been an assistant professor with the Department of Mathematical Sciences at Chalmers University of Technology and the University of Gothenburg, Sweden. His research interests include power-system cybersecurity analysis, environment-aware buildings and communities, convex and nonconvex optimization, and model reduction for dynamical systems.

Henrik Sandberg received the M.Sc. degree in engineering physics and the Ph.D. degree in automatic control from Lund University, Sweden, in 1999 and 2004, respectively. He is an associate professor with the Automatic Control Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. From 2005 to 2007, he was a post-doctoral scholar with the California Institute of Technology, Pasadena. In 2013, he was a visiting scholar at the Laboratory for
Information and Decision Systems at the Massachusetts Institute of Technology, Cambridge. He has also held visiting appointments with the Australian National University and the University of Melbourne, Australia. His current research interests include secure networked control, power systems, model reduction, and fundamental limitations in control. He was a recipient of the Best Student Paper Award from the IEEE Conference on Decision and Control in 2004 and an Ingvar Carlsson Award from the Swedish Foundation for Strategic Research in 2007. He is currently an associate editor of *Automatica*.

**Karl Henrik Johansson** is director of the KTH ACCESS Linnaeus Centre and professor at the School of Electrical Engineering, Royal Institute of Technology, Sweden. He is a Wallenberg Scholar and has held a six-year senior researcher position with the Swedish Research Council. He is director of the Stockholm Strategic Research Area ICT The Next Generation. He received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University. He has held visiting positions at the University of California, Berkeley (1998–2000) and the California Institute of Technology (2006–2007). His research interests are in networked control systems; hybrid and embedded systems; and applications in transportation, energy, and automation systems. He has been a member of the IEEE Control Systems Society Board of Governors and chair of the IFAC Technical Committee on Networked Systems. He has been on the editorial boards of several journals, including *Automatica*, *IEEE Transactions on Automatic Control*, and *IET Control Theory and Applications*. He is currently on the editorial board of *IEEE Transactions on Control of Network Systems* and the *European Journal of Control*. He has been guest editor for special issues, including the 2011 issue on wireless sensor and actuator networks in *IEEE Transactions on Automatic Control*. He was the general chair of the ACM/IEEE Cyberphysical Systems Week 2010 in Stockholm and IPC chair of many conferences. He has served on the executive committees of several European research projects in the area of networked embedded systems. In 2009, he received the Best Paper Award of the IEEE International Conference on Mobile Ad-hoc and Sensor Systems. In 2009, he was also appointed as a Wallenberg Scholar as one of the first ten scholars from all sciences, by the Knut and Alice Wallenberg Foundation.

He was awarded an Individual Grant for the Advancement of Research Leaders from the Swedish Foundation for Strategic Research in 2005. He received the triennial Young Author Prize from IFAC in 1996 and the Pecccei Award from the International Institute of System Analysis, Austria, in 1993. He received Young Researcher Awards from Scania in 1996 and from Ericsson in 1998 and 1999. He is a Fellow of the IEEE.

**REFERENCES**


