Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Disconnection-aware Attack Detection in Networked Control Systems ⋆

Hampei Sasahara * **Takayuki Ishizaki** ** **Jun-ichi Imura** **
Henrik Sandberg *

* *KTH Royal Institute of Technology, Stockholm, SE-100 44 Sweden (e-mail: {hampei,hsan}@kth.se).*
** *Tokyo Institute of Technology, Tokyo, 1528552 Japan (e-mail: {ishizaki,imura}@sc.e.titech.ac.jp)*

Abstract: This study deals with security issues in dynamical networked control systems. The goal is to establish a unified framework of the attack detection stage, which includes the four processes of monitoring the system state, making a decision based on the monitored signal, disconnecting the corrupted subsystem, and operating the remaining system during restoration. This paper, in particular, considers a disconnection-aware attack detector design problem. Traditionally, observer-based attack detectors are designed based on the system model with a fixed network topology and cannot cope with a change of the topology caused by disconnection. The disconnection-aware design problem is mathematically formulated and a solution is proposed in this paper. A numerical example demonstrates the effectiveness of the proposed detector through an inverter-based voltage control system in a benchmark model.

Keywords: Detection algorithms, distributed detection, large-scale systems, networks, on-line security analysis, system security.

1. INTRODUCTION

This study deals with security issues for physical networked systems, which face challenges in ensuring a safe and secure operation as in Ding et al. (2018). In modern and future cyber-physical systems, components are densely interconnected in both information and physical layers and as a result once an adversarial attack irrupts its effects propagate over a broad range in the entire networked system as reported in Kushner (2013); Lee et al. (2016) For secure operation of physical systems, novel security techniques in the physical layer are required aside from the existing information security techniques because of difference between the requirements of information systems and physical systems. For instance, patching and frequent updates are not well suited for control systems as mentioned in Cárdenas et al. (2011). Furthermore, implementing security technologies in physical layers in addition to information layers fits the notion of "defense in depth" advocated in Kulpers and Fabro (2006), which argues the importance of multiplication of protections.

When considering security in the physical layer we have to handle dynamics of the system because in most cases inertia cannot be ignored. In Dibaji et al. (2019) the defense mechanism of physical systems is classified into three stages: prevention, resilience, and detection stages. The process of the detection stage is further divided into the four steps: monitoring, decision, disconnection, and restoration steps. In the monitoring step, the state of the operation is observed. Based on the monitored signal, decision on attack injection is made in the decision step. The attacked components are disconnected for suppressing propagation of attack effects. Finally, after removing the cause of the attack the isolated part is reconnected. Traditionally, those steps have separately been studied based on the assumption that the steps are able to operate independently.

However, those steps should be discussed in a unified manner especially for dynamical networked systems under control, because unplugging some input ports would lead to loss of the original function owing to disconnection of feedback in controlled systems. For example, when an attack is injected into one of the reference signals, simply disconnecting the input port of the reference signal results in deviation of all output signals from the reference. In the worst case, stability of the system would be lost if the signal under attack configures a closed-loop system. Therefore, it is required to design the network system with a protection system to be compatible with disconnection and to discuss the entire security flow in the detection stage in an integrated fashion.

The goal of this study is to establish a unified framework of the detection stage. First, the system must be resilient to disconnection of a part of the networked system. Since existing researches (e.g., Ishizaki et al. (2018); Anderson et al. (2019); Qu and Simaan (2014); Akhavan et al. (2020)) enable the system continue to work under disconnection, in this paper we suppose that the entire system can be constructed to have the resilience property to disconnection. Based on this premise, this paper treats

⋆ This work was supported in part by the Swedish Research Council (grant 2016-00861), the Swedish Foundation for Strategic Research (project CLAS), the Swedish Civil Contingencies Agency (project CERCES), and JST CREST Grant Number JPMJCR15K1, Japan.

the design problem of a distributed attack detector compatible with disconnection. Conventionally, observer-based residual generators, which monitor the state of the system operation by comparing the ideal signal under the normal operation and the actual measured signal, are designed based on the system model with a fixed network topology. The design policy leads to lack of the capability to handle topology changes caused by disconnection. We formulate the disconnection-aware dynamical attack detector design problem and propose a solution based on *retrofit control* developed in Ishizaki et al. (2018, 2019). The effectiveness of the proposed approach is illustrated through a numerical example of power distribution network systems.

Several related works that propose design methods of physical systems being secure throughout the entire detection stage can be found. In Sasaki et al. (2015), the authors have proposed fallback control. In the framework, fallback operation is switched to as an incident response of the control system where facilities are preferentially protected and minimal system that can continue to operate during cyber attacks. A system design method that can perform plug-and-play fault detection and control-reconfiguration in a unified manner is proposed in Riverso et al. (2016), the method of which can possibly be applied to attack detection as well.

This paper is organized as follows. In Sec. 2, we first review security operation for dynamical networked systems. Then we provide a motivating example of power distribution network systems. Moreover, the disconnection-aware distributed attack detector design problem is formulated. We review retrofit control, which is the key to solve the formulated problem in Sec. 3, and subsequently, we provide a solution based on retrofit control. Sec. 4 provides a numerical example that illustrates the effectiveness of the proposed method and Sec. 5 draws conclusion.

2. UNIFIED FRAMEWORK FOR ATTACK DETECTION AND PROBLEM FORMULATION

In this section, we review the security flow for dynamical networked systems under attack. Then we provide a motivating example highlighting the necessity of a unified framework of the detection stage. We formulate the problem of distributed attack detector design in which the distributed attack detector can cope with topology variation caused by disconnection.

2.1 Security Operation for Dynamical Networked Systems

Consider a linear time-invariant interconnected system Σ composed of N subsystems

$$\Sigma_i : \begin{cases} \dot{x}_i &= A_i x_i + L_i v_i + B_i r_i + a_{i,x} \\ w_i &= W_i x_i + Z_i v_i + U_i r_i + a_{i,w} \\ y_i &= C_i x_i + E_i v_i + D_i r_i + a_{i,y} \end{cases} \tag{1}$$

for $i = 1, \ldots, N$ with an interconnection $v_i = \sum_{j \in \mathcal{N}_i} M_{ij} w_j$, where x_i, v_i, w_i, y_i, r_i denote the state, the interconnection input, the interconnection output, the measurement output, and the reference signal, respectively. The index set associated with the neighborhood of Σ_i is denoted by $\mathcal{N}_i$. The exogenous inputs $a_{i,x}, a_{i,w}, a_{i,y}$ represent the effects to the system behavior caused by the attack. For simplicity, the entire interconnected system is assumed to be well-posed throughout this paper.

Defense mechanisms of physical systems can be categorized into three stages: prevention, resilience, and detection as in Dibaji et al. (2019). In this study, we pay attention to the detection stage, the process of which includes monitoring the state of the system operation for detection of attacks, making a decision of attack injection, disconnecting the part under the attack for suppressing propagation of the influence of the attack, and restoring the normal operation. The process of each step, which exploits techniques in fault detection and isolation (Ding (2013)) is reviewed in the following.

As the first step of the detection stage, the residual $\epsilon(t)$ is generated by comparing the ideal output signal $\hat{y}(t)$ under the normal operation and the actual measurement output $y(t)$ by $\epsilon(t) := y(t) - \hat{y}(t)$, which is around zero when no attack is injected. The ideal output signal is generated through an observer with the information on the input signals and the dynamical system model. For instance, for the interconnected system composed of (1), a distributed observer, each of whose subobservers interacts an estimated interconnection signal $\hat{w}_i$ to each other, can be built as

$$\hat{\Sigma}_i : \begin{cases} \hat{x}_i &= A_i \hat{x}_i + L_i \hat{v}_i + B_i r_i + H_i (y_i - \hat{y}_i) \\ \hat{w}_i &= W_i \hat{x}_i + Z_i \hat{v}_i + U_i r_i \\ \hat{y}_i &= C_i \hat{x}_i + E_i \hat{v}_i + D_i r_i \\ \hat{v}_i &= \sum_{j \in \mathcal{N}_i} M_{ij} \hat{w}_j \end{cases} \tag{2}$$

with appropriate observer gains H_i. At the next step, based on the residual $\epsilon(t)$, decisions are made by a testing method.Subsequently isolation, namely, identifying the location of the attack is performed as well. Then for eliminating influence of the attack the input port is isolated by disconnecting the facilities under attack, e.g., turning off the switch of the interaction between the plant and the controller. Finally, the system is restored to regain its original function by removing the source of the attack and reconnecting the islanded components.

2.2 Motivating Example

This subsection provides a motivating example that demonstrates operation failures owing to disconnection. We consider low-voltage power distribution networks as shown in Fig. 1. In the distribution network system, each customer has a distributed generation (DG) with which an inverter is equipped as a controller. The purpose of the control is to regulate the voltage magnitudes v_k to given reference values $\overline{\mathrm{v}}_k$, which is identically set to be 230 [V] (for the detail of the modeling, see Sec. 4). As a residual generator, we design a distributed observer (2) with certain observer gains H_i designed to guarantee stability of the error dynamics for the original system Σ.

Let us observe behaviors of the system under an attack. The distribution network has five customers with an attack at the ith DGs for $i = 3, 4, 5$. As a reaction against the attack, the DGs under the attack are disconnected at the time $t_0 = 5$. The top of Fig. 2 illustrates the voltage magnitudes under the disconnection. It can be confirmed that the voltage magnitudes in the remaining parts are kept to be around the setpoint even after the disconnection.

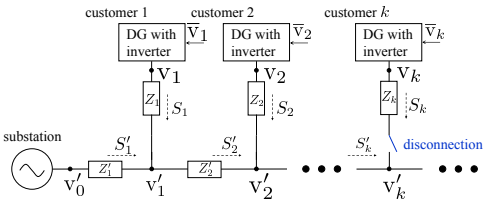


Fig. 1. Infrastructure of a low-voltage distribution network system.

This result implies that the distribution network system is resilient to disconnection caused by attacks. On the other hand, the bottom of Fig. 2 illustrates the residual with a distributed observer (2) under the disconnection. It can be observed that, although the estimation error converges to zero before the disconnection, the error diverges after the disconnection. The instability is arisen from the inadequate choice of the observer gains H_i, which are designed only for the original system without disconnection. The resulting distributed observer is not compatible with network topology change caused by the disconnection.

As shown in the example, the conventional attack detector would lead to loss of monitoring function owing to instability of the distributed observer when disconnection happens. Therefore, it is required to develop a novel attack detector design method that can cope with disconnection. Moreover, since the system dynamics is changed by the disconnection, all processes that utilize system model information have to be designed taking the system variation into account explicitly. With the observation, we consider developing a unified framework for the detection stage.

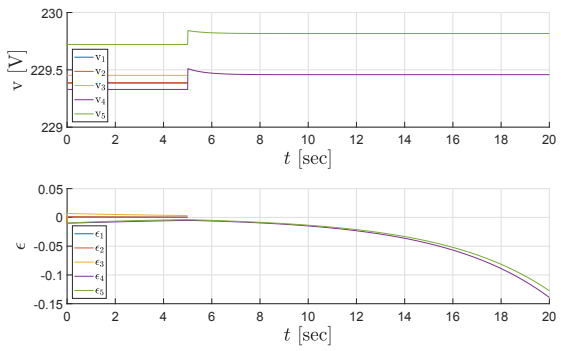


Fig. 2. Top: The voltage magnitudes under disconnection at $t_0 = 5$. Bottom: The residual with a distributed observer (2) under disconnection at $t_0 = 5$.

2.3 Disconnection-aware Attack Detector Design Problem

The motivating example above has shown the necessity of a novel framework for the security flow:

(0) design process: designing the interconnected system Σ to be resilient to disconnection caused by attacks,
(1) monitoring process: monitoring the local time series of the residual $\epsilon_i(t)$ generated by comparing the actual measurement signal $y_i(t)$ with the ideal signals $\hat{y}_i(t)$ created through a distributed observer,
(2) decision process: making a decision of attack and the attacked place at each detector based on the local monitored residual $\epsilon_i(t)$,
(3) disconnection process: disconnecting the ith subsystem where an attack is detected, namely, w_i is set to be zero,
(4) operation and restoration process: operating the remaining interconnected system during restoration.

Clearly, at least the first three processes from the top of the list are required to be compatible with disconnection. Our goal is to establish a unified framework for the entire security flow.

In this study, we regard stability under disconnection as resiliency of the networked system. We treat the resiliency of the original system as a premise. The following assumption is made.

Assumption 1. For any $\mathcal{I} \subset \{1, \dots, N\}$, the remaining interconnected system $\Sigma_{\mathcal{I}}$ is internally stable where $\Sigma_{\mathcal{I}}$ is given by the interconnected system composed of the subsystems Σ_i for $i \in \mathcal{I}$ with $w_j = 0$ for $j \notin \mathcal{I}$ in (1).

Assumption 1 states that the stability of the interconnected system itself is guaranteed for any disconnection with respect to the interconnection between the subsystems. Although the detail of designing the interconnected system satisfying Assumption 1 is not focused on in this paper, we provide brief description of two possible approaches for the design method in the following. The first approach is given by simply applying existing methods in the design process for designing a dynamical interconnected system to be resilient to disconnection, see Ishizaki et al. (2018); Anderson et al. (2019); Qu and Simaan (2014); Akhavan et al. (2020). In this approach the points to be disconnected are naturally determined and subsystems Σ_i are given as the designed components. In the other approach, for a given system Σ, we choose the interconnection points to be disconnected such that Assumption 1 is satisfied. The pick of connecting points leads to the system description as an interconnected system of Σ composed of the subsystems Σ_i. The policy of deciding interconnection points can specifically be given when the entire system is built by a negative feedback interconnection of passive systems (Brogliato et al. (2006)). In this case, Assumption 1 is satisfied simply choosing each element as a subsystem. If the original system does not have such a property, connecting points can be determined through off-line simulation in advance.

In this paper, we focus on the problem of designing a distributed observer to generate $\hat{y}$ in the monitoring process as a preliminary step. The difficulty here is that, even if the dynamics of the error signal $e := x - \hat{x}$ with the distributed observer in the form of (2) with fixed observer gains H_i is stable, the remaining error signal dynamics under some disconnection is possibly destabilized with the original observer gains, which are compatible only with the original fully interconnected system Σ.

We formulate the dynamical attack detector design problem as follows.

Problem 1. Under Assumption 1, design a distributed observer $(\hat{x}_i, \hat{y}_i, \hat{w}_i) = \mathcal{O}_i(y_i, \{\hat{w}_j\}_{j \in \mathcal{N}_i}, r_i)$ such that $\hat{x}_{\mathcal{I}} \to x_{\mathcal{I}}$ and $\hat{y}_{\mathcal{I}} \to y_{\mathcal{I}}$ as $t \to +\infty$ for any $\mathcal{I} \in \{1, \dots, N\}$ under any initial conditions when the attack signals are zero, where $\hat{x}_{\mathcal{I}}, \hat{y}_{\mathcal{I}}$ are the states and outputs of $\mathcal{O}_{\mathcal{I}}$, which

is given by the distributed observers composed of $\mathcal{O}_i$ for $i \in \mathcal{I}$ with $\hat{w}_j = 0$ for $j \notin \mathcal{I}$.

Note that a naive approach to guarantee stability under network topology variation caused by disconnection is to not utilize feedback of the error signal, namely, setting $H_i = 0$ for any i in (2). Then the stability of the error signal dynamics is preserved owing to Assumption 1 because it is equivalent to the stability of the original system. However, the poles of the error signal dynamics cannot be moved at all without feedback and hence early attack detection cannot be performed with the naive approach. Note also that, if we represent the stability condition as linear matrix inequalities (LMIs), the number of the constraints reaches 2^N and hence it is practically impossible to solve the problem via the naive LMI approach.

3. PROPOSED DISCONNECTION-AWARE ATTACK DETECTOR DESIGN METHOD

In this section, we propose a design method of a distributed observer that meets the requirement in the formulated problem. First, we briefly review retrofit control, developed in Ishizaki et al. (2018, 2019), which plays the key role to solve the problem. Subsequently, we propose a solution based on the retrofit control.

3.1 Brief Review of Retrofit Control

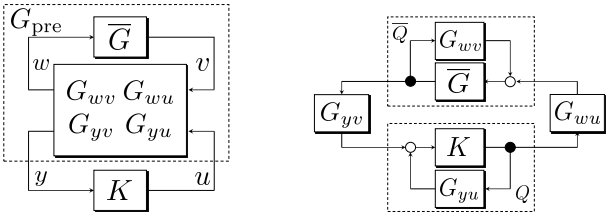
Retrofit control is a newly developed controller design method the main purpose of which is to achieve modular design of large-scale interconnected control systems. In the framework, each subcontroller is supposed to be designed only using the model information of the subsystem to which the subcontroller is attached. For the controller design it is naturally required that the function held by the entire interconnected system must be preserved even with introduction of subcontrollers in a sequential manner. Retrofit control provides a specific controller design method that fulfills this requirement.

The schematic diagram of the interconnected system to be controlled in the retrofit control framework is depicted in Fig. 3 (a). It is assumed that the model information of $G = (G_{wv}, G_{wu}, G_{yv}, G_{yu})$, the subsystem associated with the subcontroller to be designed, is available while the model information of $\overline{G}$, the other part of the interconnected system, is unavailable. We suppose that the interconnected system without the controller K, denoted by G_{pre}, stably operates and consider attaching K to improve the control performance while preserving the stability. To mathematically handle the problem, we first consider the following set $\overline{\mathcal{G}} := \{\overline{G} : G_{\mathrm{pre}} \text{ is internally stable.}\}$ and provide the definition of retrofit controllers as follows.

Definition 1. The controller K in Fig. 3 (a) is said to be a retrofit controller if the entire feedback system in Fig. 3 (a) is internally stable for any $\overline{G} \in \overline{\mathcal{G}}$.

A necessary and sufficient condition for retrofit controllers is given as follows:

Proposition 1. Assume that G in Fig. 3 (a) is stable. Then K is a retrofit controller if and only if $G_{wu}QG_{yv} = 0$ and Q is stable where $Q := K(I - G_{yu}K)^{-1}$ denotes the Youla parameter with respect to K for G_{yu}.



(a) Entire interconnected system to be controlled in the retrofit control framework.

(b) Equivalent system explicitly representing the Youla parameters Q and $\overline{Q}$.

Fig. 3. Equivalent block diagrams of the closed-loop system.

The interpretation of the conditions is given as follows. Fig. 3 (b) is obtained by transforming the system Fig. 3 (a) with the Youla parameters Q and $\overline{Q}$, which correspond to K and $\overline{G}$, respectively. Since $\overline{Q}$ is an arbitrary stable transfer matrix, the condition is necessary and sufficient for internal stability. For the complete proof, see Ishizaki et al. (2019).

3.2 Proposed Design Method

We propose a disconnection-aware distributed attack detector design method as a solution to Problem 1 based on retrofit control. Let us consider a distributed observer in the following form:

$$\mathcal{O}_i : \begin{cases} \dot{\hat{x}}_i &= A_i\hat{x}_i + L_i\sum_{j\in\mathcal{N}_i} M_{ij}\hat{w}_j + B_i r_i + \mu_i \\ \hat{w}_i &= W_i\hat{x}_i + Z_i\sum_{j\in\mathcal{N}_i} M_{ij}\hat{w}_j + U_i r_i + \nu_i \\ \hat{y}_i &= C_i\hat{x}_i + E_i\sum_{j\in\mathcal{N}_i} M_{ij}\hat{w}_j + D_i r_i \end{cases} \tag{3}$$

where μ_i and ν_i are artificial signals injected into the sub-observer. It should be noted that the distributed observer has the form of (2) if we choose $\mu_i = H_i(y_i - \hat{y}_i)$ and $\nu_i = 0$.

Let $e_i := \hat{x}_i - x_i$ be the local state estimation error. Then the dynamics of e_i is represented as

$$\mathcal{E}_i : \begin{cases} \dot{e}_i &= A_i e_i + L_i\sum_{j\in\mathcal{N}_i} M_{ij}\omega_j + \mu_i \\ \omega_i &= W_i e_i + Z_i\sum_{j\in\mathcal{N}_i} M_{ij}\omega_j + \nu_i \\ \psi_i &= C_i e_i + E_i\sum_{j\in\mathcal{N}_i} M_{ij}\omega_j \end{cases}$$

where $\omega_i := \hat{w}_i - w_i$ and $\psi_i := \hat{y}_i - y_i$ denote the local estimation errors in terms of the interconnection signal and the measurement output, respectively. Obviously, Problem 1 is equivalent to the stabilization problem of the interconnected systems composed of $\mathcal{E}_i$ under disconnection as long as the distributed observer is assumed to have the form in (3). Therefore, we consider designing a stabilizing controller $K_i : \psi_i \mapsto (\mu_i, \nu_i)$ in the following discussion.

Now we take a particular $\mathcal{I} \subset \{1, \ldots, N\}$ and let $\mathcal{E}_\mathcal{I}$ be the interconnected system composed of $\mathcal{E}_i$ for $i \in \mathcal{I}$. Then the error dynamics can be interpreted as an interconnection system of $\mathcal{E}_\mathcal{I}$ and the others with a controller K_i. From Assumption 1 and Proposition 1, if K_i satisfies the condition for retrofit controllers for any $i \in \{1, \ldots, N\}$, then the resulting distributed observer with the controllers K_i meets the requirement of the problem.

The difficulty here is that the procedure to design a retrofit controller for K_i is not obvious. The following theorem, the main result of this paper from a theoretical perspective, provides a retrofit controller design method for K_i.

Theorem 1. Design K_i as

$$K_i : \begin{cases} \dot{\chi}_i &= A_i\chi_i + \mu_i \\ \mu_i &= H_i\psi_i,\ \nu_i = -W_i\chi_i \end{cases}$$

with an observer gain H_i such that $A_i + H_iC_i$ is Hurwitz. Then K_i satisfies the condition for retrofit control.

Proof. See Sasahara et al. (2020). □

Solution: In summary, the structure of the proposed disconnection-aware distributed dynamical attack detector is given by

$$\mathcal{O}_i : \begin{cases} \dot{\hat{x}}_i = A_i\hat{x}_i + L_i\sum_{j\in\mathcal{N}_i} M_{ij}\hat{w}_j + B_ir_i + \mu_i \\ \hat{w}_i = W_i\hat{x}_i + Z_i\sum_{j\in\mathcal{N}_i} M_{ij}\hat{w}_j + U_ir_i + \nu_i \\ \hat{y}_i = C_i\hat{x}_i + E_i\sum_{j\in\mathcal{N}_i} M_{ij}\hat{w}_j + D_ir_i \\ K_i : \begin{cases} \dot{\chi}_i &= A_i\chi_i + \mu_i \\ \mu_i &= H_i(\hat{y}_i - y_i),\ \nu_i = -W_i\chi_i \end{cases} \end{cases} \tag{4}$$

with observer gains H_i such that $A_i + H_iC_i$ is Hurwitz for any $i \in \{1, \dots, N\}$, which gives a solution to Problem 1.

4. NUMERICAL EXAMPLE

In this section, we examine how the proposed attack detector works through a numerical example of a power distribution network system. For security issues in power distribution network systems, see Teixeira et al. (2015); Isozaki et al. (2016).

The infrastructure of the distribution network is shown in Fig. 1 where v'_0 represents the substation bus voltage magnitude. Lines are represented by a series impedance $Z_k = R_k + jX_k$ or $Z'_k = R'_k + jX'_k$ and each customer has a DG with an inverter. For each customer, the received voltage level is v_k and the voltage level at the point of connection with the distribution line is v'_k. At the kth connecting point, $S_k = P_k + j(q_{\mathrm{g},k} - q_{\mathrm{c},k})$ and $S'_k = P'_k + jQ'_k$ denote the power supplies from the DG and the distribution line, respectively. We employ the LinDistFlow model developed in Baran and Wu (1989) for power flow and voltage drop equations represented by $S'_k = S'_{k+1} - S_k, \mathrm{v}'^2_{k+1} = \mathrm{v}'^2_k - 2\beta'_{k+1}(S'_{k+1}), \mathrm{v}^2_k = \mathrm{v}'^2_k + 2\beta_k(S_k)$ where $\beta_k(S_k) := R_kP_k + X_kQ_k$ and $\beta'_k(S'_k) := R'_kP'_k + X'_kQ'_k$. The inverter dynamics is given by $\dot{q}_{\mathrm{g},k} = -q_{\mathrm{g},k}/\tau_k + \kappa_k(\overline{\mathrm{v}}^2_k - \mathrm{v}^2_k)/\tau_k$ where τ_k is the time-constant and κ_k is the droop gain as in Chong et al. (2019). As a specific network, we consider a benchmark European low-voltage distribution network in Strunz et al. (2014) as in Chong et al. (2019). It can numerically be confirmed that the interconnected system satisfies Assumption 1.

We now suppose that an attack is injected into the power distribution network system. The k_0th customer's DG is assumed to be attacked and the squared voltage reference signal $\overline{\mathrm{v}}^2_{k_0}$ is fabricated. The attack signal is modeled as a simple step function beginning at t_{a}. Accordingly, the fabricated squared reference voltage signal $\overline{\mathrm{v}}^2_{\mathrm{f},k_0}$ is represented by $\overline{\mathrm{v}}^2_{\mathrm{f},k_0}(t) = \overline{\mathrm{v}}^2_{k_0}(t) + \overline{a}\iota_{\{t:t\geq t_{\mathrm{a}}\}}(t)$ where $\overline{a}$ is a scaler value that represents the amplitude of the attack and ι_Ω denotes the indicator function of the set Ω.

For monitoring process, we design the proposed distributed attack detector. Let Σ_i be the subsystem composed of the ith customer and branch for $i \in \{1, \dots, 5\}$. The internal state and the interconnection signal of Σ_i are $q_{\mathrm{g},i}$ and (v'^2_i, S'_i), respectively. We suppose that each $q_{\mathrm{g},i}$ is measurable at each subobserver $\mathcal{O}_i$, namely, $y_i = q_{\mathrm{g},i}$, for attack detection. Finally, the distributed observer is designed based on the structure (4) where observer gains H_i are determined based on the state-feedback linear quadratic regulator design method under certain weights.

Let us compare the time responses of residuals generated through the naive approach, namely, (2) with $H_i = 0$, and the proposed detector. Fig. 4 shows the time response of the normalized residuals for the attack with $\overline{a} = 1$ and $t_{\mathrm{a}} = 1$ for $k_0 = 4$, where the normalized factor γ is determined by the direct-current (DC) gain of the transfer function from the attack to the residual. Suppose that 0.95 is taken as the baseline for the decision making, which is depicted by the broken line in Fig. 4. When the value of the normalized residual exceeds the baseline, the decision of attack injection is made. Then the times of attack detection are $t = 3.5$ and $t = 1.5$, which are represented by dash-dot lines, with the no-feedback detector and the proposed detector, respectively. It can be observed that the detection time is much earlier owing to the estimation error feedback.

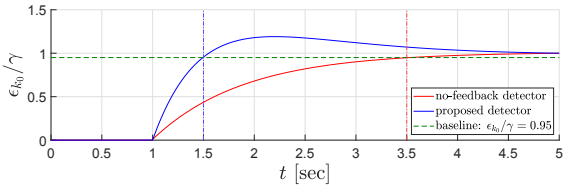


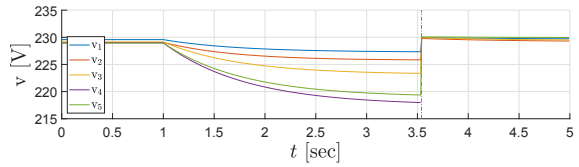
Fig. 4. Time series of normalized residuals ϵ_{k_0}/γ with the distributed attack detector without error feedback and the proposed distributed attack detector where $\overline{a} = 1$, $t_{\mathrm{a}} = 1$, and γ is set to be the DC gain of the transfer function from the attack to the residual.

The time series of each customer's voltage are shown in Fig. 5.The attack is made with the parameters $\overline{a} = -5 \times 10^4$ and $t_{\mathrm{a}} = 1$ with $k_0 = 4$. The decision on disconnection in terms of Σ_i is made when $|\overline{a}|\epsilon_i(t)/\gamma > 0.95$. As shown in Fig. 5, owing to the early detection achieved by the proposed detector the voltage deviations with respect to non-attacked customers are suppressed compared with the naive detector.

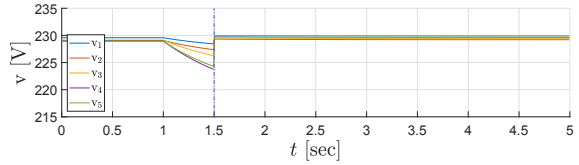
Finally, we confirm that the stability of the error dynamics is guaranteed under disconnection in the same setting of Fig. 2. Fig. 6 illustrates the time series of the residual with the proposed attack detector where the observer gains are the same as the ones used in Fig. 2. It can be observed that the stability is preserved even after the disconnection owing to the proposed structure.

5. CONCLUSION

In this paper, we have considered detection process of defense mechanism in physical layers towards attack-resilient cyber-physical systems. In particular, we have paid attention to the disconnection-aware attack detector design problem and proposed a novel design method based on retrofit control. Through the numerical example, the effectiveness has been verified.



(a) No-feedback distributed detector



(b) Proposed distributed detector

Fig. 5. Time series of each customer's voltage under the attack injected into the k_0th subsystem from $t_a = 1$ where Σ_i is disconnected when the attack is detected at $\mathcal{O}_i$.

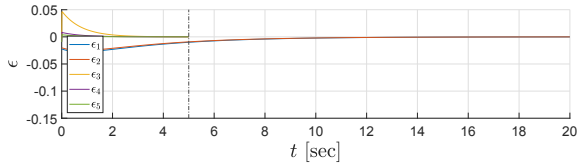


Fig. 6. Time series of the residual with the proposed distributed observer under the disconnection.

The important direction of future works is to develop an entire framework for the detection stage. For instance, while in the numerical example we employ a simple criterion for making a decision with a fixed baseline, the criterion is desirable to be dependent on the time instant, the system model, the supposed attack, and also the variable network topology. Moreover, worst-case analysis against white-box attacks is considered to be another direction.

REFERENCES

Akhavan, A., Mohammadi, H.R., Vasquez, J.C., and Guerrero, J.M. (2020). Passivity-based design of plug-and-play current-controlled grid-connected inverters. *IEEE Trans. Power Electron.*, 35(2), 2135–2150.

Anderson, J., Doyle, J., Low, S., and Matni, N. (2019). System level synthesis. *Annual Reviews in Control*, 47, 364–393.

Baran, M. and Wu, F.F. (1989). Optimal sizing of capacitors placed on a radial distribution system. *IEEE Trans. Power Del.*, 4(1), 735–743.

Brogliato, B., Lozano, R., Maschke, B., and Egeland, O. (2006). *Dissipatie Systems Analysis and Control: Theory and Applications*. Communications and Control Engineering. Springer, 2nd edition.

Cárdenas, A.A., Amin, S., Lin, Z.S., Huang, Y.L., Huang, C.Y., and Sastry, S. (2011). Attacks against process control systems: Risk assessment, detection, and response. In *Proc. the 6th ACM ASIA Conference on Computer and Communications Security*.

Chong, M.S., Umsonst, D., and Sandberg, H. (2019). Local voltage control of an inverter-based power distribution network with a class of slope-restricted droop controllers. In *Proc. the 8th IFAC Workshop on Distributed Estimation and Control in Networked Systems*.

Dibaji, S.M., Pirani, M., Flamholz, D.B., Annaswamy, A.M., Johansson, K.H., and Chakrabortty, A. (2019). A systems and control perspective of CPS security. *Annual Reviews in Control*, 47, 394–411.

Ding, D., Han, Q.L., Xiang, Y., Ge, X., and Zhang, X.M. (2018). A survey on security control and attack detection for industrial cyber-physical systems. *Neurocomputing*, 275, 1674–1683.

Ding, S.X. (2013). *Model-Based Fault Diagnosis Techniques Design Schemes, Algorithms and Tools*. Advances in Industrial Control. Springer, 2nd edition.

Ishizaki, T., Sadamoto, T., Imura, J., Sandberg, H., and Johansson, K.H. (2018). Retrofit control: Localization of controller design and implementation. *Automatica*, 95, 336–346.

Ishizaki, T., Sasahara, H., Inoue, M., Kawaguchi, T., and Imura, J. (2019). Modularity-in-design of dynamical network systems: Retrofit control approach. [Online]. Available: `https://arxiv.org/abs/1902.01625`.

Isozaki, Y., Yoshizawa, S., Fujimoto, Y., Ishii, H., Ono, I., Onoda, T., and Hayashi, Y. (2016). Detection of cyber attacks against voltage control in distribution power grids with PVs. *IEEE Trans. Smart Grid*, 7(4), 1824–1835.

Kulpers, D. and Fabro, M. (2006). Control systems cyber security: Defense in depth strategies. Technical report, Idaho National Laboratory. [Online]. Available: `https://inldigitallibrary.inl.gov/sites/sti/sti/3375141.pdf`.

Kushner, D. (2013). The real story of Stuxnet. *IEEE Spectrum*, 50(3), 48–53.

Lee, R.M., Assante, M.J., and Conway, T. (2016). Analysis of the cyber attack on the Ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*. [ONLINE] Available: `https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf`.

Qu, Z. and Simaan, M.A. (2014). Modularized design for cooperative control and plug-and-play operation of networked heterogeneous systems. *Automatica*, 50(9), 2405–2414.

Riverso, S., Boem, F., Ferrari-Trecate, G., and Parisini, T. (2016). Plug-and-play fault detection and control-reconfiguration for a class of nonlinear large-scale constrained systems. *IEEE Trans. Autom. Control*, 61(12), 3963–3978.

Sasahara, H., Ishizaki, T., Imura, J., and Sandberg, H. (2020). Disconnection-aware attack detection in networked control systems. [Online]. Available: `https://arxiv.org/abs/1911.05084`.

Sasaki, T., Sawada, K., Shin, S., and Hosokawa, S. (2015). Model based fallback control for networked control system via switched Lyapunov function. In *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, 2000–2005.

Strunz, K., Abbasi, E., and Abbey, C., et al. (2014). Benchmark systems for network integration of renewable and distributed energy resources. *CIGRE Task Force*.

Teixeira, A., Paridari, K., Sandberg, H., and Johansson, K.H. (2015). Voltage control for interconnected microgrids under adversarial actions. In *2015 IEEE 20th Conference on Emerging Technologies Factory Automation (ETFA)*, 1–8.