Optimal Partial Feedback Attacks in Cyber-Physical Power Systems

Guangyu Wu, Gang Wang, *Member, IEEE*, Jian Sun, *Member, IEEE*, and Jie Chen, *Fellow, IEEE*

***Abstract*—This paper considers false data injection attacks constructed based on partial feedback of generator frequencies in a cyber-physical power system. The goal of the attacker is to destabilize the system, by compromising a subset of frequency control signals with false data injection, without consuming much energy. In this context, two attack design problems are studied, considering both location-fixed attacks and location-switching attacks based on measurable generator bus frequencies. They are further modeled as switched control problems, for which closed-form solutions can be attained. Leveraging the Maximum Principle, the diagonal partial feedback matrix is optimized by solving a convex optimization problem. The convexified switching variables describing the switching behaviour are solved in a quadratic optimization problem and a fractional optimization problem respectively. As a result, optimal switching conditions to select the best attack locations are obtained, along with optimal partial feedback attack matrices. Case studies on the IEEE 9-bus test system validate the practical merits of theory and numerical effectiveness of the proposed attack schemes.**

Index terms— Partial feedback, location-switching attacks, switching condition, mixed integer, convex relaxation.

I. Introduction

THE cyber-physical power system composed of a massive amount of highly coupled heterogeneous network components, becomes more interconnected and more interdependent than conventional power systems, both physically and informatively. Due to the deep fusion and close interaction of physical and information processes, contemporary cyber-physical power systems are facing both cyber vulnerabilities and physical threats [1].

A. Motivation

Due to the widespread use of computerized elements, cyber-security of the power grid has become a critical and growing challenge [2]. A new class of attacks emerged rapidly over the last decades–attacks launched in the cyber domain compromise digital units in the physical domain [3]. Such attacks can be launched at various electronic devices distributed in different regions of large-scale cyber-physical power systems, such as PMUs, circuit breakers, transformers, inductive loads, instruments, etc. Attackers can launch intermittent attacks in a cooperative manner at multiple points, increasing the difficulty and burden of preventing attacks. Investigating the worst case impact of resource-constrained attackers on cyber-phsical power systems can be used for vulnerability analysis and risk assessment.

B. Related Work

Among a multitude of contributions on exploiting attack strategies, there is a growing interest in attacks with switching behaviors where attack locations are usually governed by a switching mechanisms [4]. In the related literature, the adversary can switch jammed channels [5] to break data availability, switch compromised sensors [6] to break data integrity, and target circuit breakers to break network connectivity [7]. Accordingly, sensor observations, timestamps of transmitted data [8], or network topology can be manipulated to maximize a malicious objective. Unfortunately, even if fault detection is employed in an estimator or controller, convergence of the estimation error [5] or stability of the attacked states [9] cannot be guaranteed in general, incurring abnormal operations and disruptions to the system.

Using an attack matrix (stacking attack vectors) to describe switching location attacks, initial state recovery from compromised measurements was studied in [10]. A multiple-model state filtering algorithm was developed in [6], despite false data injection attacks with unknown magnitude and locations, as well as attacks that change the system's mode of operation. Similar to [6], a switching signal is introduced to describe the switching among different network topologies in [11]. Cyberattacks implementing stochastic switching laws among multiple network typologies were investigated for linear multiagent systems. Binary variables are used to describe the switching behaviour between multiple channels [12]. The attacker was restricted to jam only one channel or, not to take action, where an optimal schedule to decide when and which channel to attack was obtained. A state-dependent switching signal was constructed to determine whether to switch from one subsystem to another [13]. The switching attack was modeled as a constant-switching signal [14], generating a modulating sequence of possible injections. An iterated game

The work of G. Wu and J. Sun was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61621063, 61522303, by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under U1509215, by the Project of Major International (Regional) Joint Research Program NSFC under Grant 61720106011, and by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1208).

G. Wu is with the Clean Energy Automotive Engineering Center, School of Automotive Studies, Tongji University, Shanghai 201804, China. J. Sun is with the State Key Lab of Intelligent Control and Decision of Complex Systems, and the School of Automation, Beijing Institute of Technology, Beijing 100081, China. G. Wang is with the Digital Technology Center and the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA. J. Chen is with the State Key Lab of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, China, and also the Tongji University, Shanghai 200092, China. Emails: mebest21@163.com; gangwang@umn.edu; sunjian@bit.edu.cn; chenjie@bit.edu.cn.

between attacks and distributed control was formulated to stabilize the power system.

C. Contributions

The switching attacks reviewed so far are conceived beforehand to guide resilient system design, which pursued only constant-frequency switching [10], and deterministic or random switching laws [6], [7]. The optimal switching law with respect to a given objective, even for linear time-invariant systems, has rarely been studied. Our precursor [15] developed the optimal switching policy, yet assuming availability of full state information of the attacked system. In practice, acquiring full state information of a continent-scale power grid is almost impossible for an attacker. This motivates well our present work on the optimal attack design based only on partial feedback information. Two key challenges arise: Q1) Whether and how one can design an optimal feedback attack law to maximize the damage to a system when only partial state information can be acquired? and Q2) How can one design an optimal switching law for selecting most favorable locations to attack as well as devise a partial feedback attack law simultaneously? In this paper, we answer affirmatively the two questions, under suitable assumptions on the form of attacks.

The switching partial feedback attack design problem is formulated as a nonconvex optimization involving both continuous and discrete variables, which is generally hard to solve. Leveraging advances in convex relaxation, seeking the globally optimal solution is possible. The optimal partial feedback matrices can be found by efficiently solving a convex program, and each entry of the optimal attack signal is constructed based on a measurable state. The main contributions of this work are summarized as follows:

c1) For partial feedback attacks at fixed locations, we develop an optimal partial state feedback law instead of obtaining a partial state observer to yield a full state feedback law, to maximize a quadratic function. By appropriately designing feedback matrices, a closed-form solution can be obtained by solving a convex program.

c2) To deal with the discrete location selection variables, the nonconvex optimal partial feedback attack design problem is convexified. We show that the optimal solution of the latter recovers an optimal solution of the original nonconvex problem. That is, optimal solutions of both the discrete switch input and the continuous attack signal can be found efficiently by solving a convex program, including an algebraic switching condition as well as a feedback attack law based on partial states.

The rest of this paper is structured as follows: In Sec. II, system modeling is outlined. In Sec. III, optimal partial feedback attacks with location fixed or switching is studied. Numerical tests on the IEEE 9-bus system are presented in Sec. IV, and the present paper is concluded in Sec. V.

II. System Model

Consider a power system with $\mathcal{G} := \{1, \ldots, m\}$ and $\mathcal{L} := \{1, \ldots, n\}$ being the set of generator buses and load buses, respectively. The linearized power flow equations at each bus can be written as

$$P_i^G = \sum_{j\in\mathcal{G}} H_{ij}(\delta_i - \delta_j) + \sum_{j\in\mathcal{L}} H_{ij}(\delta_i - \theta_j), \quad \forall i \in \mathcal{G}$$
$$-P_i^L = \sum_{j\in\mathcal{G}} H_{ij}(\theta_i - \delta_j) + \sum_{j\in\mathcal{L}} H_{ij}(\theta_i - \theta_j), \quad \forall i \in \mathcal{L}$$

where P_i^G is the power injection of the generator at bus i, P_i^L is the power consumption of the load at bus i, δ_i and θ_i are the voltage phase angle and the voltage phase angle at generator bus i, and H_{ij} is the admittance of the transmission line between buses i and j. The generator dynamics at each generator bus $i \in \mathcal{G}$ can be written as

$$\dot{\delta}_i = \omega_i$$
$$M_i^g \dot{\omega}_i = P_i^M - D_i^g \omega_i - P_i^G$$

where ω_i is the rotor frequency deviation at the generator bus i, M_i^g is the inertia of the rotor, D_i^g is the damping coefficient, and P_i^M is the mechanical power input.

The overall dynamics of a power system can be commonly modeled using the following state-space equations (see e.g., [9], [16])

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}^g & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{\delta}} \\ \dot{\boldsymbol{\omega}} \\ \dot{\boldsymbol{\theta}} \end{bmatrix} = - \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{H}^{gg} & \boldsymbol{D}^g & \boldsymbol{H}^{gl} \\ \boldsymbol{H}^{lg} & \boldsymbol{0} & \boldsymbol{H}^{ll} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\omega} \\ \boldsymbol{\theta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{P}^M \\ -\boldsymbol{P}^L \end{bmatrix} \tag{1}$$

where $\boldsymbol{\delta}, \boldsymbol{\omega} \in \mathbb{R}^m$ collect the voltage phase angles, the rotor frequency deviations at all generator buses, respectively, and $\boldsymbol{\theta} \in \mathbb{R}^n$ the voltage phase angles at all load buses; vector $\boldsymbol{P}^M \in \mathbb{R}^m$ concatenates the mechanical power input at all generator buses, and $\boldsymbol{P}^L \in \mathbb{R}^n$ the sum of controllable but frequency-insensitive loads and uncontrollable loads [9].

Diagonal matrices $\boldsymbol{M}^g$ and $\boldsymbol{D}^g$ hold entries of $\{M_i^g > 0\}$ and $\{D_i^g > 0\}$ on their main diagonals, respectively, along with the imaginary part of the admittance matrix as follows

$$\boldsymbol{H}_{bus} = \begin{bmatrix} \boldsymbol{H}^{gg} & \boldsymbol{H}^{gl} \\ \boldsymbol{H}^{lg} & \boldsymbol{H}^{ll} \end{bmatrix}.$$

The goal of this paper is to design feedback data injection attacks from the viewpoint of the adversary to maximally corrupt the closed-loop system performance. Towards this objective, we consider two controllers that affect the mechanical power input, i.e., the governor and load frequency controller for generators with automatic generation control (AGC). AGC is a fundamental control system used in all power grids to maintain the grid frequency at its nominal value, by adjusting the output power of generators based on measurements collected from sensors distributed in the grid [17]. In this paper, we consider the integral and proportional controller given by (see e.g., [9])

$$P_i^M = -\left(K_i^P \omega_i + K_i^I \int_0^t \omega_i \, dt\right) \tag{2}$$

where $K_i^P \geq 0$ and $K_i^I \geq 0$ are the proportional and integral controller coefficients, respectively. Controller parameters are set so as to keep the system stable in absence of an attack. The healthy controller (2) can be abbreviated by

$$\boldsymbol{u} := -\boldsymbol{K}^P \boldsymbol{\omega} - \boldsymbol{K}^I \boldsymbol{\delta}. \tag{3}$$

where $\boldsymbol{u} := [u_1 \ \cdots \ u_m]^\top$ is a m dimensional vector, $\boldsymbol{K}^P$ and $\boldsymbol{K}^I$ are diagonal matrices holding entries of $\{K_j^P\}_{j=1}^m$ with $\{K_j^I\}_{j=1}^m$ on their main diagonals, respectively. The control signals u_i are transmitted over communication lines (cyber space) between the local controller and the generator buses [14].

We assume the controllable loads can be actively controlled and the uncontrollable loads change over time but is pre-specified [18]. Since the problem of interest does not depend on $\boldsymbol{P}^L$, we take without loss of generality that $\boldsymbol{P}^L = \boldsymbol{0}$. Eliminating the last row of the linear descriptor system (1), we obtain a general linear system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \tag{4}$$

where $\boldsymbol{x} = \left[\boldsymbol{\delta}^\top \ \boldsymbol{\omega}^\top\right]^\top$ is the state vector and matrices $\boldsymbol{A}$, $\boldsymbol{B}$ are given by

$$\begin{aligned}\boldsymbol{A} = &\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}^g \end{bmatrix}^{-1} \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{H}_{gl} \end{bmatrix} \boldsymbol{H}_{ll}^{-1} \begin{bmatrix} \boldsymbol{H}_{lg} & \boldsymbol{0} \end{bmatrix} \right. \\ &\left. - \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{I} \\ \boldsymbol{H}_{gg} & \boldsymbol{D}_g \end{bmatrix} \right) \\ \boldsymbol{B} = &\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}^g \end{bmatrix}^{-1}.\end{aligned}$$

Before presenting the main results, several standard assumptions on the adversary are made.

Assumption 1. *To investigate the worst case attack consequences, the adversary is assumed to know the system dynamics [19] in* (1) *and the control law in* (2)*.*

Assumption 2. *It is possible for a malicious third party to access the communication channel between the controller and generator buses secretly [20]. Similar to spoofing-based man-in-the-middle (MIM) attack described in [21] and [22], the adversary can intercept transmitted packets $\boldsymbol{u}$ and send false control signals $\boldsymbol{u}^c$ into target generator buses. Thereby, the actual control signal is completely replaced by the constructed false control signal. The adversary can switch the compromised channels (attack locations) frequently, but the number of channels that can be attacked has an upper bound r [23].*

When an attack occurs, the adversary adds false data $\hat{\boldsymbol{f}}^a$ to the healthy controller.

$$\boldsymbol{u}^c := \boldsymbol{u} + \hat{\boldsymbol{f}}^a \tag{5}$$

and

$$\hat{\boldsymbol{f}}^a := \boldsymbol{D}^a \boldsymbol{f}^a \tag{6}$$

where $\boldsymbol{D}^a := \left[\boldsymbol{d}_1^{a\top} \ \cdots \ \boldsymbol{d}_m^{a\top}\right]^\top$ is a $m \times m$ matrix, which is designed as a priori. Let $\boldsymbol{f}^a := [f_1^a \ \cdots \ f_m^a]^\top$ be an m-dimensional vector representing the partial feedback attack signal to be optimized. Let $\mathcal{S} := \{1, \ldots, r\} \subsetneqq \mathcal{G}$ collect the indices of all attackable channels. If the attack is realized at the channel transmitting signal u_i ($i \in \mathcal{S}$), then $\boldsymbol{d}_i^a$ is a nonzero vector. Thus, $\boldsymbol{D}^a$ can be viewed as some 'indicator' matrix, which reveals the locations of attacks.

Assumption 3. *The adversary injects datum $\boldsymbol{d}_i^{a\top}\boldsymbol{f}^a$ into u_i, where $\boldsymbol{f}^a$ is constructed based on measurable generator frequencies. Scalar $\boldsymbol{d}_i^{a\top}\boldsymbol{f}^a$ is a combination of measurable frequencies, and $\boldsymbol{d}_i^a$ collects the combination coefficients. For simplicity, we consider $\boldsymbol{d}_i^a$ takes $\boldsymbol{0}$ or $\boldsymbol{1}$ (a vector with all elements takes 0 or 1) here.*

The attacker can deploy frequency disturbance recorders (FDRs) [24] on generator buses or intrude into the supervisory control and data acquisition (SCADA) system [25] to acquire local synchronized frequency measurements. In practice, the adversary does not have sufficient instrumentation or capability to acquire frequencies of all generator buses. On the other hand, there is only limited phasor measurement unit (PMU) installation in real-world power grids [26], so phase angles at most buses are typically not available. These two facts prompt us to consider partial frequency feedback attacks rather than the less practical case where full frequencies are assumed accessible.

Suppose the adversary can only measure ℓ out of the m generator bus frequencies. Let $\mathcal{M} := \{z_1, \ldots, z_\ell\} \subsetneqq \mathcal{G}$ collect the indices of all generator buses whose frequencies are measured. If G_j represents the feedback gain of f_j^a and ω_j the frequency of generator bus $j \in \mathcal{M}$, one can write

$$f_a^j := \begin{cases} 0, & \forall j \notin \mathcal{M} \\ G_j\omega_j, & \forall j \in \mathcal{M} \end{cases}.$$

The attack signal can be compactly expressed as

$$\boldsymbol{f}^a = \boldsymbol{G}^a\boldsymbol{\omega} \tag{7}$$

in which $\boldsymbol{G}^a \in \mathbb{R}^{m\times m}$ is a diagonal matrix holding entries $\{G_j\}_{j=1}^m$ on its main diagonal. The diagonal entry corresponding to the non-measured frequency is set to 0. Compared with other structures of partial feedback matrices, the diagonal structure requires a minimum number of entries to be optimized while ensuring completeness of feedback information. Evidently, one can express $\boldsymbol{G}^a$ as a linear combination of unit diagonal matrices, which are obtained by zeroing out $m-1$ diagonal entries of the $m \times m$ identity matrix. To see this, consider the following example.

Example 1. *For $m = 3$ and $\ell = 2$, suppose that the metered frequencies are ω_2 and ω_3. It holds that*

$$\boldsymbol{G}^a = \begin{bmatrix} 0 & 0 & 0 \\ 0 & G_1 & 0 \\ 0 & 0 & G_2 \end{bmatrix} = G_1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + G_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where G_1 and G_2 are nonzero controller gains.

Substituting (7) into (6) and merging (6) into (3) yields

$$\boldsymbol{u}^c := -\boldsymbol{K}^P\boldsymbol{\omega} - \boldsymbol{K}^I\boldsymbol{\delta} + \boldsymbol{D}^a\boldsymbol{G}^a\boldsymbol{\omega} \tag{8}$$

The system stability can be destroyed by carefully designing feedback matrix $\boldsymbol{G}^a$. Failure to stabilize the frequency may cause damages to equipment and reduction or interruption to electricity supply [9]. Following conventions, we use symbol $\boldsymbol{x}^c$ to denote the state vector of the attacked system. Precisely, plugging (8) into (4), the attacked system model can be written as

$$\begin{aligned}\dot{\boldsymbol{x}}^c &= \boldsymbol{A}\boldsymbol{x}^c + \boldsymbol{B}^a\boldsymbol{u}^a \\ \boldsymbol{u}^a &= \boldsymbol{G}\boldsymbol{x}^c\end{aligned} \tag{9}$$

where

$$\boldsymbol{B}^a = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{D}^a \end{bmatrix} \tag{10}$$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}^a \end{bmatrix}, \quad \boldsymbol{u}^a = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{f}^a \end{bmatrix}. \tag{11}$$

For convenience, the measurable frequencies are placed at the bottom of $\boldsymbol{x}^c$. The partial feedback matrix can be decomposed into

$$\boldsymbol{G}(t) = \sum_{i=1}^{\ell} G_i(t) \boldsymbol{E}_i \tag{12}$$

where

$$\boldsymbol{E}_i = \mathrm{diag}([\underbrace{0 \ \cdots \ 0}_{m+z_i-1} \ 1 \ 0 \ \cdots \ 0]), \ z_i \in \mathcal{M}. \tag{13}$$

The coefficients collected into $\boldsymbol{g}(t) := [G_1(t) \ \cdots \ G_\ell(t)]^\top$ are time-varying variables to be sought, each now corresponding to a measurable frequency.

III. Optimal Partial Feedback Attacks

This section derives the optimal data injection attacks when only partial feedback information is available. A quadratic objective function is utilized for the attacker to reflect the state deviations from the origin as well as the energy cost of the attack signal over a finite time interval $[t_0, \ t_f]$ [15], [27]. In words, the adversary designs $\boldsymbol{u}^a$ by solving

$$\begin{aligned} \max \ \ J = & \frac{1}{2} \boldsymbol{x}^{c\top}(t_f) \boldsymbol{W} \boldsymbol{x}^c(t_f) \\ & + \frac{1}{2} \int_{t_0}^{t_f} \Big[\boldsymbol{x}^{c\top}(t) \boldsymbol{Q} \boldsymbol{x}^c(t) - \boldsymbol{u}^{a\top}(t) \boldsymbol{R} \boldsymbol{u}^a(t) \Big] dt \end{aligned} \tag{14}$$

where $\boldsymbol{W}$, $\boldsymbol{Q}$, and $\boldsymbol{R}$ are preselected $2m \times 2m$-dimensional symmetric coefficient matrices. In addition, matrices $\boldsymbol{W}$ and $\boldsymbol{Q}$ are positive semi-definite, and $\boldsymbol{R}$ is positive definite. The traditional linear quadratic regulator (LQR) aims to minimize the state deviation as well as the control cost [28]. Contrarily, the worst-case attack is designed to maximize the state deviation while minimizing an attacking cost [29] or setting an upper bound on the attacking cost [30].

On the other hand, the attacker tries to bypass the detection mechanism in order not to trigger an alarm and subsequent countermeasures [31]. The stealth requirements imposed by detection mechanisms can be typically converted into constraints related to attack magnitude and frequency [32], [33]. Therefore, this work minimizes the attacking cost to restrict the attack magnitude both for energy reduction and stealthiness. The oscillation magnitudes of generator frequencies and phase angles can be controlled by adjusting $\boldsymbol{Q}$ and $\boldsymbol{R}$ to avoid being detected. Meanwhile, the location-switching mechanism is utilized to form intermittent attacks launched at each location and the attack frequency can also be regulated by adjusting weighting matrices.

In the following, the case where attack locations are fixed by the adversary a priori is investigated first, while location-switching attacks with a fixed number of compromised control signals are subsequently studied.

A. Optimal Partial Feedback Attacks with Fixed Locations

If the attack location is fixed over the entire control period, then $\boldsymbol{D}^a$ is time-invariant. The maximum principle can be applied to find the optimal partial feedback matrix $\boldsymbol{G}^*(t)$, which is tantamount to computing the optimal diagonal entries $\boldsymbol{g}^*(t) := [G_1^*(t) \ \cdots \ G_\ell^*(t)]^\top$.

Theorem 1. *The optimal coefficients of the partial feedback matrix are given by*

$$\boldsymbol{g}^*(t) = \boldsymbol{M}^{-1}(t) \boldsymbol{c}(t). \tag{15}$$

where entries of $\boldsymbol{M} \in \mathbb{R}^{\ell \times \ell}$ and $\boldsymbol{c} \in \mathbb{R}^\ell$ are given by

$$M(i,j) = \boldsymbol{x}^{c\top}(t) \boldsymbol{E}_i^\top \boldsymbol{R} \boldsymbol{E}_j \boldsymbol{x}^c(t), \quad \forall i,j = 1, \ldots, \ell \tag{16}$$

$$c(i) = \boldsymbol{\lambda}^{*\top}(t) \boldsymbol{B}^a \boldsymbol{E}_i \boldsymbol{x}^c(t). \quad \forall i = 1, \ldots, \ell \tag{17}$$

where $\boldsymbol{\lambda}^*(t) := [\lambda_1^*(t) \ \cdots \ \lambda_n^*(t)]^\top$ is the solution of

$$\dot{\boldsymbol{\lambda}}(t) = -\boldsymbol{Q} \boldsymbol{x}^c(t) - \boldsymbol{A}^\top \boldsymbol{\lambda}(t). \tag{18}$$

with the boundary condition $\boldsymbol{\lambda}(t_f) = \boldsymbol{W} \boldsymbol{x}(t_f)$.

Proof. Our proof is based on Pontryagin's maximum principle. We start by writing the Hamilton function

$$\begin{aligned} H(\boldsymbol{x}^c, \boldsymbol{u}^a; \boldsymbol{\lambda}) = & \frac{1}{2} \Big[\boldsymbol{x}^{c\top}(t) \boldsymbol{Q} \boldsymbol{x}^c(t) - \boldsymbol{u}^{a\top}(t) \boldsymbol{R} \boldsymbol{u}^a(t) \Big] \\ & + \boldsymbol{\lambda}^\top(t) \Big[\boldsymbol{A} \boldsymbol{x}^c(t) + \boldsymbol{B}^a \boldsymbol{u}^a(t) \Big]. \end{aligned} \tag{19}$$

Plugging (12) into (19) yields

$$\begin{aligned} H = & \frac{1}{2} \boldsymbol{x}^{c\top}(t) \boldsymbol{Q} \boldsymbol{x}^c(t) \\ & - \frac{1}{2} \sum_{i,j} G_i(t) G_j(t) \boldsymbol{x}^{c\top}(t) \boldsymbol{E}_i^\top \boldsymbol{R} \boldsymbol{E}_j \boldsymbol{x}^c(t) \\ & + \boldsymbol{\lambda}^\top(t) \Big[\boldsymbol{A} \boldsymbol{x}^c(t) + \sum_i G_i(t) \boldsymbol{B}^a \boldsymbol{E}_i \boldsymbol{x}^c(t) \Big]. \end{aligned}$$

By the co-state equation, we deduce (18). For brevity, the dependence on t will be dropped. After removing terms in H that do not depend on $\boldsymbol{g}$, maximizing H with respect to $\boldsymbol{g}$ is equivalent to maximizing the following reduced Hamilton

$$\bar{H}(\boldsymbol{g}) := -\frac{1}{2} \boldsymbol{g}^\top \boldsymbol{M} \boldsymbol{g} + \boldsymbol{c}^\top \boldsymbol{g}$$

where $\boldsymbol{M}$ and $\boldsymbol{c}$ are defined in (16) and (17), respectively.

The Hessian matrix of $\bar{H}$ with respect to $\boldsymbol{g}$ is

$$\frac{\partial^2 \bar{H}}{\partial \boldsymbol{g}^2} = -\boldsymbol{M} = -\boldsymbol{X}^\top \widetilde{\boldsymbol{R}} \boldsymbol{X} \tag{20}$$

where $\boldsymbol{X} \in \mathbb{R}^{\ell \times \ell}$ is a diagonal matrix that holds the metered frequencies on its main diagonal, and $\widetilde{\boldsymbol{R}} \in \mathbb{R}^{\ell \times \ell}$ is the corresponding block submatrix of $\boldsymbol{R}$. Consider Example 1 again, for which we have the following

$$\boldsymbol{X} = \begin{bmatrix} \omega_2 & 0 \\ 0 & \omega_3 \end{bmatrix}$$

and

$$\widetilde{\boldsymbol{R}} = \begin{bmatrix} R(5,5) & R(5,6) \\ R(6,5) & R(6,6) \end{bmatrix}.$$

Since every principal submatrix of a positive definite matrix is positive definite, we deduce that the submatrix $\tilde{\boldsymbol{R}}$ of $\boldsymbol{R} \succ \boldsymbol{0}$

is positive definite. Therefore, the Hessian $-\boldsymbol{X}^\top \tilde{\boldsymbol{R}} \boldsymbol{X}$ in (20) is negative definite, implying that $\bar{H}$ is strictly concave. According to standard convex optimization, the unique maximum of $\bar{\boldsymbol{H}}$ (and thus $\boldsymbol{H}$) is attained at the stationary point $\boldsymbol{g}^*$ dictated by setting

$$\left.\frac{\partial \bar{H}}{\partial \boldsymbol{g}}\right|_{\boldsymbol{g}=\boldsymbol{g}^*} = -\boldsymbol{M}\boldsymbol{g}^* + \boldsymbol{c} = \boldsymbol{0}$$

yielding the optimal coefficients of partial feedback matrices

$$\boldsymbol{g}^* = \boldsymbol{M}^{-1}\boldsymbol{c}$$

and

$$\max\ \bar{H} = \frac{1}{2}\boldsymbol{c}^\top \boldsymbol{M}^{-1}\boldsymbol{c}$$

which completes the proof. □

B. Optimal Partial Feedback Attacks with Switching Locations

Control of large-scale systems (power networks, in particular) is often implemented in a distributed manner, involving multiple controllers distributed over a large geographical area. The adversary, in a real-world setting, has only limited resources, and is thus capable of compromising only a part of the vulnerable local controllers at a time, due to lack of resources. This paper considers an attacker who constantly changes attack locations expecting to acquire more severe consequences. As the fixed-location attack is a special case of switching-location attacks, it has been shown that the latter performs better than the former [15].

We consider the attack scenario in which a fixed number r of m control signals are attacked during the attack, and the total number of candidate attack locations (i.e., size-r location sets) is $N := \binom{m}{r}$. These N location sets can be represented by the 'indicator' matrices $\{\boldsymbol{D}_s^a\}_{s=1}^N$. The attacker's goal is to determine an optimal switching sequence from all candidate location sets to attack with an optimal partial feedback attack law. Next, we introduce N binary variables to describe the switching behavior. The attacked system dynamics can be given by

$$\dot{\boldsymbol{x}}^c(t) = \boldsymbol{A}\boldsymbol{x}^c(t) + \sum_{s=1}^{N} w_s(t)\boldsymbol{B}_s^a \boldsymbol{u}^a(t) \tag{21}$$

where matrices $\{\boldsymbol{B}_s^a\}_{s=1}^N$ can be obtained by substituting $\boldsymbol{D}_s^a$ into (10). The switch inputs $\{w_s(t)\}_{s=1}^N$ belong to

$$\mathcal{W}_0 := \left\{ \boldsymbol{w}(t) \,\middle|\, \sum_{s=1}^{N} w_s(t) = 1, \text{ and } w_s(t) \in \{0,1\},\ \forall s \right\}. \tag{22}$$

For all $t \in [t_0, t_f]$, since only one location set (namely, $\boldsymbol{D}_s^a$ for some s) is to be chosen, its corresponding switch input $w_s(t)$ is set 1, while the others are set 0. Finding the optimal selection criterion of the location sets boils down to find optimal values of vector $\boldsymbol{w}(t) := [w_1(t)\ \cdots\ w_N(t)]^\top$. The location sets at all switching instants and their corresponding partial feedback attack law define the so-called switching sequence

$$\boldsymbol{\zeta} := \Big\{ \big(\boldsymbol{w}(t_0), \boldsymbol{G}(t_0)\big),\ \ldots,\ \big(\boldsymbol{w}(t_F), \boldsymbol{G}(t_F)\big) \Big\} \tag{23}$$

where $t_0 \le t_1 \le \ldots \le t_F \le t_f$, the set $\{t_1, \ldots, t_F\}$ collects all switching instants (the time when a switching operation is performed), and F is the total number of switching operations.

The optimal partial feedback attack design problem with location switching is to find $\boldsymbol{w}(t)$ and $\boldsymbol{G}(t)$ that

$$\max \quad J \tag{24a}$$

$$\text{s. to} \quad \dot{\boldsymbol{x}}^c(t) = \boldsymbol{A}\boldsymbol{x}^c(t) + \sum_{s=1}^{N} w_s(t)\boldsymbol{B}_s^a u^a(t) \tag{24b}$$

$$\boldsymbol{u}^a(t) = \sum_{i=1}^{\ell} G_i(t)\boldsymbol{E}_i \boldsymbol{x}^c(t) \tag{24c}$$

$$\boldsymbol{w}(t) \in \mathcal{W}_0,\ \forall t. \tag{24d}$$

In fact, constraint (24d) involving integer variables, rendering (24) nonconvex and NP-hard in general [34]. Seeking an optimal solution becomes challenging, due to additionally the coupling between the continuous coefficients $\{G_i(t)\}$ and the discrete switch inputs $\{w_s(t)\}$. Most previous results assume a fixed switching sequence, which is known *a priori* and thus significantly decreases the difficulty (e.g., [34]); only optimal switching instants are to be searched by traditional nonlinear optimization approaches.

To tackle this challenge, we view the attacked system (24b) as a linear switched system (see e.g., [35] for related definitions). Interestingly, problem (24) can be treated as an optimal partial feedback control problem of a linear switched system. The closed-form solution of linear quadratic regulator of switched systems (SLQR) with full state information is obtained leveraging convex relaxation in our previous work [36]. The SLQR problem using partial state information has not yet been studied. Most advances have been focused on designing full state observers based on partial state information so as to construct the full state feedback control law [37],[38]. In contrast, this paper addresses how to directly solve the SLQR problem in the context of cyber-physical power systems, when some frequency states are unobservable and only partial state feedback can be constructed.

According to the optimal coefficients $G_i^*(t)$ of partial feedback attack matrices in Theorem 1, we hopefully tackle (24) by means of convex relaxation to acquire both the optimal coefficients and the switching inputs. The idea of convex relaxation is to relax each discrete variable $w_s(t) \in \{0, 1\}$ to a continuous one $w_s(t) \in [0, 1]$. Rather than dealing with constraint (24d), we deal with the switch input vector $\boldsymbol{w}(t)$ belonging to the following set

$$\mathcal{W}_1 := \left\{ \boldsymbol{w}(t) \,\middle|\, \sum_{s=1}^{N} w_s(t) = 1, \text{ and } 0 \le w_s(t) \le 1, \forall s \right\}. \tag{25}$$

After replacing the constraint $\boldsymbol{w}(t) \in \mathcal{W}_0$ in (24) with $\boldsymbol{w}(t) \in \mathcal{W}_1$, we arrive at the following relaxed partial feedback attack design problem

$$\max \quad \text{(24a)} \tag{26a}$$

$$\text{s. to} \quad \text{(24b), (24c), and } \boldsymbol{w}(t) \in \mathcal{W}_1. \tag{26b}$$

Interestingly, this relaxed problem can be cast as an optimal control problem, which can be solved leveraging Pontryagin's

maximum principle. If luckily, the optimal solution of $\boldsymbol{w}(t)$ in (26) is achieved at one of the vertices of the polytope of $\mathcal{W}_1$ (i.e., $\boldsymbol{w}(t) \in \mathcal{W}_0$) for all t, it is safe to conclude that this solution is also the optimal solution of the original problem (24) [39]. To proceed, we discuss separately the following two cases depending on whether a single objective function or the sum of multiple objective functions is maximized.

1) Single objective function: If the objective function in (14) is adopted, we have the following result.

Theorem 2. *If J is maximized, the optimal switching condition for the original design problem* (24) *is given by*

$$s^*(t) := \arg \max_{s\in\{1,\dots,N\}} \ \boldsymbol{c}_s{}^\top(t)\boldsymbol{M}^{-1}(t)\boldsymbol{c}_s(t) \tag{27}$$

where entries of $\boldsymbol{c}_s$ are

$$c_s(i) = \boldsymbol{\lambda}^\top(t)\boldsymbol{B}_s^a\boldsymbol{E}_i\boldsymbol{x}^c(t), \quad \forall i,j = 1,\dots,\ell. \tag{28}$$

The optimal coefficients of partial feedback matrices are

$$\boldsymbol{g}^*(t) = \boldsymbol{M}^{-1}(t)\boldsymbol{c}_{s^*}(t) \tag{29}$$

where $\boldsymbol{\lambda}(t)$ is the solution of (18) *with the boundary condition $\boldsymbol{\lambda}(t_f) = \boldsymbol{W}\boldsymbol{x}(t_f)$.*

Proof. We start with the Hamilton function

$$\begin{aligned} H = &\frac{1}{2}\boldsymbol{x}^c(t)^\top\boldsymbol{Q}\boldsymbol{x}^c(t) - \frac{1}{2}\boldsymbol{u}^a(t)^\top\boldsymbol{R}\boldsymbol{u}^a(t) \\ &+ \boldsymbol{\lambda}^\top(t)\Big[\boldsymbol{A}\boldsymbol{x}^c(t) + \sum_s w_s\boldsymbol{B}_s^a\boldsymbol{u}^a(t)\Big]. \end{aligned} \tag{30}$$

By the co-state equation, we have (18). Recalling entries of $\boldsymbol{M}$ in (16), the reduced Hamilton function becomes

$$\bar{H} = -\frac{1}{2}\boldsymbol{g}^\top(t)\boldsymbol{M}(t)\boldsymbol{g}(t) + \boldsymbol{d}^\top(t)\boldsymbol{g}(t)$$

where entries of $\boldsymbol{d}(t)$ are

$$d(i) = \sum_s w_s\boldsymbol{\lambda}^\top(t)\boldsymbol{B}_s^a\boldsymbol{E}_i\boldsymbol{x}^c(t) \tag{31}$$

Obviously, invoking Theorem 1, the optimal coefficients are

$$\boldsymbol{g}^*(t) = \boldsymbol{M}^{-1}(t)\boldsymbol{d}(t)$$

and

$$\begin{aligned} \max_{\boldsymbol{g},\boldsymbol{w}} \ \bar{H} &= \max_{\boldsymbol{w}} \ \frac{1}{2}\boldsymbol{d}^\top(t)\boldsymbol{M}^{-1}(t)\boldsymbol{d}(t) \\ &= \frac{1}{2}\sum_{s=1}^N\sum_{k=1}^N w_s w_k \boldsymbol{c}_s^\top(t)\boldsymbol{M}^{-1}(t)\boldsymbol{c}_k(t) \end{aligned}$$

where entries of $\boldsymbol{c}_i$ are defined in (28). Maximizing $\bar{H}$ with respect to the switching input $\boldsymbol{w}(t)$ is a quadratic optimization problem. The Hessian matrix of $\bar{H}$ is

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{c}_1^\top\boldsymbol{M}^{-1}\boldsymbol{c}_1 & \cdots & \boldsymbol{c}_1^\top\boldsymbol{M}^{-1}\boldsymbol{c}_N \\ \vdots & \ddots & \vdots \\ \boldsymbol{c}_N^\top\boldsymbol{M}^{-1}\boldsymbol{c}_1 & \cdots & \boldsymbol{c}_N^\top\boldsymbol{M}^{-1}\boldsymbol{c}_N \end{bmatrix} \\ = &\begin{bmatrix} \boldsymbol{c}_1^\top \\ \vdots \\ \boldsymbol{c}_N^\top \end{bmatrix} \boldsymbol{M}^{-1} \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_N \end{bmatrix} \succeq \boldsymbol{0} \end{aligned}$$

We have showed that the Hessian matrix of $\bar{H}$ is positive semidefinite; that is, function $\bar{H}$ is convex. The minimum of maximizing a convex function over a convex set $\mathcal{W}_1$ is always attained at a vertex of the convex polytope determined by the N box constraints in $\mathcal{W}_1$. In a nutshell, the switch input $\boldsymbol{w}(t)$ obtains its optimal solution in $\mathcal{W}_0$. Therefore,

$$\max \ \bar{H} = \max_{s\in\{1,\dots,N\}} \boldsymbol{c}_s^\top\boldsymbol{M}^{-1}\boldsymbol{c}_s$$

completing the proof. □

Algorithm 1: Optimal Partial Feedback Switching Attack Algorithm.

```
Determine r, ℓ, N, {D_a^s}_{s=1}^N, and {E_i}_{i=1}^ℓ as a prior.
Set: W, Q, and R according to the attacker's preference.
Initialize: attack horizon [t_0, t_f], initial state x_c(t_0),
  terminal state x_c(t_f), and initial co-state λ(t_0).
for k = 0, ..., f do
    for i, j = 1, ..., ℓ do
        Compute M(i, j) in (16);
    end
    for s = 1, ..., N do
        Compute c_s in (28);
        Evaluate (27) to determin s*(t);
    end
    Compute g*(t_k) in (29) and G^a(t_k) in (12);
    Compute u_a(t_k) and x^c(t_{k+1}) in (9);
    Compute λ(t_{k+1}) in (18);
end
```

Remark 1. The initial co-state $\boldsymbol{\lambda}(t_0)$ can be found by solving a two-point boundary value problem.

2) Sum of multiple objective functions: In diverse practical setups, the adversary is likely to set a tradeoff between the quadratic function of $\boldsymbol{x}^c$ and that of $\boldsymbol{u}^a$ based on their degree of importance or the adversary's preferences. Furthermore, different objective functions can be employed for different attack locations. This prompts us to choose a more meaningful objective function J_o, constructed by the summation of the excited local objective functions J_s; that is

$$J_o = \sum_{s=1}^N w_s(t)J_s \tag{32}$$

where $\boldsymbol{w}(t) \in \mathcal{W}_0$ and

$$\begin{aligned} J_s = &\frac{1}{2}\boldsymbol{x}^{c\top}(t_f)\boldsymbol{W}_s\boldsymbol{x}^c(t_f) \\ &+ \frac{1}{2}\int_{t_0}^{t_f}\Big[\boldsymbol{x}^{c\top}(t)\boldsymbol{Q}_s\boldsymbol{x}^c(t) - \boldsymbol{u}^{a\top}(t)\boldsymbol{R}_s\boldsymbol{u}^a(t)\Big]dt. \end{aligned}$$

Similar to (14), $\boldsymbol{W}_s$ and $\boldsymbol{Q}_s$ are positive semi-definite matrices. In order to obtain an analytical solution, we consider all $\boldsymbol{R}_s$'s are positive diagonal matrices holding entries $\{\gamma_s^j > 0\}_{j=1}^{2m}$ on their main diagonals.

Theorem 3. *The optimal switching condition to maximize J_o is given by*

$$s^*(t) := \arg \max_{s\in\{1,\dots,N\}} \ \boldsymbol{c}_s^\top(t)\widetilde{\boldsymbol{M}}^{-1}(t)\boldsymbol{c}_s(t). \tag{33}$$

The optimal coefficients of partial feedback matrices are

$$\boldsymbol{g}^*(t) = \widetilde{\boldsymbol{M}}^{-1}(t)\boldsymbol{d}(t) \tag{34}$$

where entries of $\boldsymbol{c}$ *and* $\boldsymbol{d}$ *are given in* (28) *and* (31)*, respectively, and entries of* $\widetilde{\boldsymbol{M}}$ *are*

$$\widetilde{M}(i,j) = \boldsymbol{x}^{c\top}(t)\boldsymbol{E}_i^\top \boldsymbol{R}_{s^*}\boldsymbol{E}_j\boldsymbol{x}^c(t) \quad \forall i,j = 1,\ldots,\ell \tag{35}$$

where $\boldsymbol{\lambda}(t)$ *is the solution of*

$$\dot{\boldsymbol{\lambda}}(t) = -\boldsymbol{Q}_{s^*}\boldsymbol{x}^c(t) - \boldsymbol{A}^\top\boldsymbol{\lambda}(t) \tag{36}$$

with the boundary condition $\boldsymbol{\lambda}(t_f) = \boldsymbol{W}_{s^*}\boldsymbol{x}(t_f)$.

Proof. Appealing again to the Pontryagin maximum principle and convex relaxation, the Hamilton function is given by

$$H = \frac{1}{2}\sum_s w_s[\boldsymbol{x}^{c\top}(t)\boldsymbol{Q}_s\boldsymbol{x}^c(t) - \boldsymbol{u}^{a\top}(t)\boldsymbol{R}_s\boldsymbol{u}^a(t)] + \boldsymbol{\lambda}^\top(t)\Big[\boldsymbol{A}\boldsymbol{x}^c(t) + \sum_s w_s\boldsymbol{B}_s^a\boldsymbol{u}^a(t)\Big]. \tag{37}$$

The co-state equation confirms that

$$\dot{\boldsymbol{\lambda}} = -\sum_s w_s\boldsymbol{Q}_s\boldsymbol{x}^c - \boldsymbol{A}^\top\boldsymbol{\lambda}.$$

The reduced Hamilton function can be written as

$$\bar{H} = -\frac{1}{2}\sum_i\sum_j G_iG_j\sum_s w_s\boldsymbol{x}^{c\top}\boldsymbol{E}_i^\top\boldsymbol{R}_s\boldsymbol{E}_j\boldsymbol{x}^c + \sum_s w_s\Big(\frac{1}{2}\boldsymbol{x}^{c\top}\boldsymbol{Q}_s\boldsymbol{x}^c + \boldsymbol{\lambda}^\top\boldsymbol{B}_s^a\sum_i G_i\boldsymbol{E}_i\boldsymbol{x}^c\Big) \tag{38}$$

which can be simplified as

$$\bar{H}(\boldsymbol{g}) = -\frac{1}{2}\boldsymbol{g}^\top\widetilde{\boldsymbol{M}}\boldsymbol{g} + \boldsymbol{d}^\top\boldsymbol{g} \tag{39}$$

where entries of $\widetilde{\boldsymbol{M}}$ are given by

$$\widetilde{M}(i,j) := \sum_s w_s\boldsymbol{x}^{c\top}\widetilde{\boldsymbol{R}}_s\boldsymbol{x}^c = \sum_s w_s\boldsymbol{x}^{c\top}\boldsymbol{E}_i^\top\boldsymbol{R}_s\boldsymbol{E}_j\boldsymbol{x}^c$$

where $\widetilde{\boldsymbol{R}}_s := \boldsymbol{E}_i^\top\boldsymbol{R}_s\boldsymbol{E}_j$. For Example 1, we have that $\widetilde{\boldsymbol{R}}_s := \mathrm{diag}([\gamma_s^5\ \gamma_s^6])$. Recalling (31) and leveraging Theorem 2, we have that

$$\boldsymbol{g}^* = \widetilde{\boldsymbol{M}}^{-1}\boldsymbol{d}. \tag{40}$$

Substituting (40) into (39) yields

$$\bar{H}(\boldsymbol{g}^*) = \frac{1}{2}\boldsymbol{d}^\top\widetilde{\boldsymbol{M}}^{-1}\boldsymbol{d} = \sum_s\sum_k w_sw_k\boldsymbol{c}_s^\top\widetilde{\boldsymbol{M}}^{-1}\boldsymbol{c}_k = \frac{1}{\boldsymbol{x}^{c\top}\boldsymbol{x}^c}\sum_s\frac{\psi_s^2(\boldsymbol{w})}{\phi_s(\boldsymbol{w})}.$$

where $\psi_s(\boldsymbol{w}) := \sum_k w_kc_k(s)$ and $\phi_s(\boldsymbol{w}) := \sum_k w_k\gamma_k^s$. Clearly, it holds that $\phi_s(\boldsymbol{w}) > 0$. Maximizing $\bar{H}$ with respect to $\boldsymbol{w}(t)$ is a fractional optimization problem. The second derivative of $\bar{H}(\boldsymbol{g}^*)$ with respect to w_k is

$$\frac{\partial^2\bar{H}}{\partial w_k^2} = 2\sum_s\frac{\big(c_k(s)\phi_s - \gamma_k^s\psi_s\big)^2}{\phi_s^3} \geq 0. \tag{41}$$

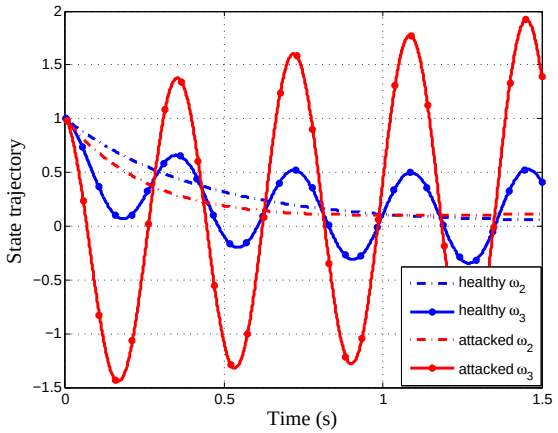

Fig. 1: State trajectories under partial feedback attacks at u_5.

Likewise, the second partial derivative of $\bar{H}$ with respect to w_s and w_k can be found as

$$\frac{\partial\bar{H}}{\partial w_s\partial w_k} = 2\sum_s\frac{\big(c_k(s)\phi - \gamma_k^s\psi_s\big)\big(c_k(s)\phi_s - \gamma_k^s\psi_s\big)}{\phi_s^3}. \tag{42}$$

For brevity, define $\boldsymbol{v}_s := [v_1^s\ \cdots\ v_N^s]^\top$ with entries being $v_k^s = c_k(s)\phi_s - \gamma_k^s\psi_s$. Then using (41) and (42), along with some algebraic manipulations, the Hessian matrix of $\bar{H}(\boldsymbol{g}^*)$ with respect to $\boldsymbol{w}$ can be compactly written as

$$\frac{\partial^2\bar{H}}{\partial\boldsymbol{w}^2} = \frac{2}{\boldsymbol{x}^{c\top}\boldsymbol{x}^c\phi_s^3}\sum_s\boldsymbol{v}_s\boldsymbol{v}_s^\top \succeq \boldsymbol{0}$$

which confirms that function $\bar{H}$ is convex over $\mathcal{W}_1$. Similar to Theorem 2, the solution of maximizing $\bar{H}(\boldsymbol{g}^*)$ is attained at least at a vertex of $\mathcal{W}_1$, i.e., the switch input $\boldsymbol{w}(t)$ obtains its optimal solution in $\mathcal{W}_0$. Therefore,

$$\max_{\boldsymbol{g}}\ \bar{H}(\boldsymbol{g}) = \max_{\boldsymbol{w}}\ \bar{H}(\boldsymbol{g}^*) = \max_{s\in\{1,\ldots,N\}}\ \boldsymbol{c}_s^\top\widetilde{\boldsymbol{M}}^{-1}\boldsymbol{c}_s \tag{43}$$

completing the proof. □

IV. Illustrative Example

The attack design approaches discussed in Theorems 1-3 were numerically examined using the IEEE 9-bus benchmark system [40], which consists of 3 generator buses and 6 load buses. We consider 2 frequency control signals that can be strategically modified by a knowledgeable attacker. As a result, the system frequencies will deviate from their nominal values. Specifically, we assume that the attacker can measure the frequencies of generators 2 and 3, construct the attack signal $\boldsymbol{f}^a = [0\ G_1\omega_2\ G_2\omega_3]^\top$, and alter the frequency control signals. Values of the system parameters simulated in this paper were taken from [40].

Set $\boldsymbol{K}^P = \mathrm{diag}([0.01\ 0.01\ 0.01])$, $\boldsymbol{K}^I = \boldsymbol{0}$, $\boldsymbol{Q} = \mathrm{diag}([0\ 0\ 0\ 2\ 2\ 2])$, $\boldsymbol{R} = \mathrm{diag}([0\ 0\ 0\ 10\ 10\ 10])$, and $\boldsymbol{d}_5^a = [1\ 1\ 1]^\top, \boldsymbol{d}_k^a = \boldsymbol{0}, \forall 1 \leq k \leq 6, k \neq 5$ (only u_5 is attacked, that is fixed-location attacks). Using Theorem 1, Fig. 1 shows the frequency response of the attacked system.

Considering the single objective case that the attack signal switches between u_5 and u_6, let $\boldsymbol{d}_5^a$ be $\boldsymbol{d}_5^{a,1}$, and set

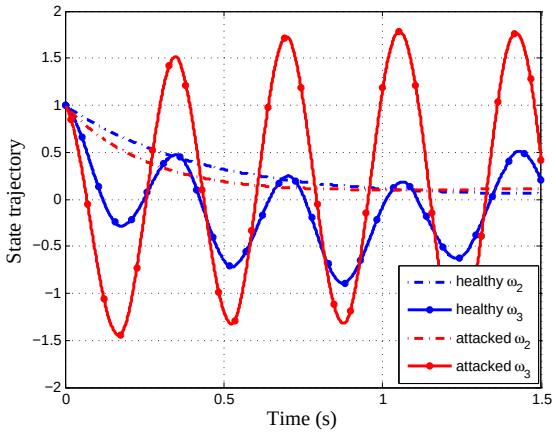

Fig. 2: State trajectories under switching attacks.

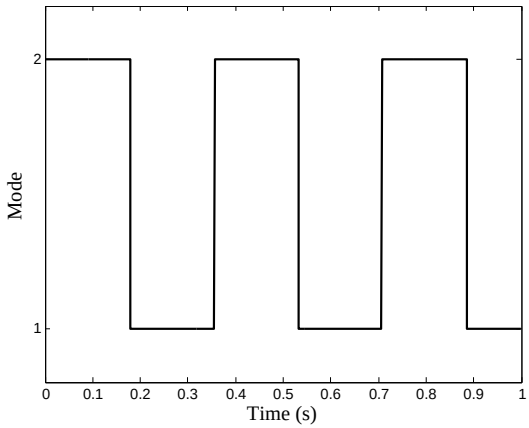

Fig. 3: Optimal switching times.

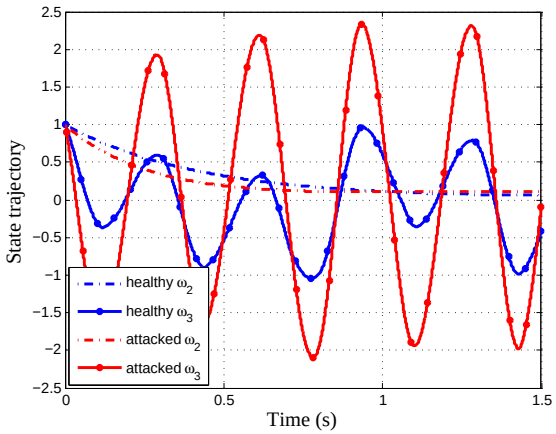

Fig. 4: State trajectories under switching attacks.

$\boldsymbol{d}_6^{a,2} = [1\ 1\ 1]^\top$, $\boldsymbol{d}_k^{a,2} = \mathbf{0}, \forall 1 \leq k \leq 5$ (only u_6 was attacked). Leveraging Theorem 2, Fig. 3 depicts the frequency response under switching attacks, along with the optimal attack switching sequence presented in Fig. 4. Observing from Fig. 2, at the switching instants, the curves corresponding to attacked

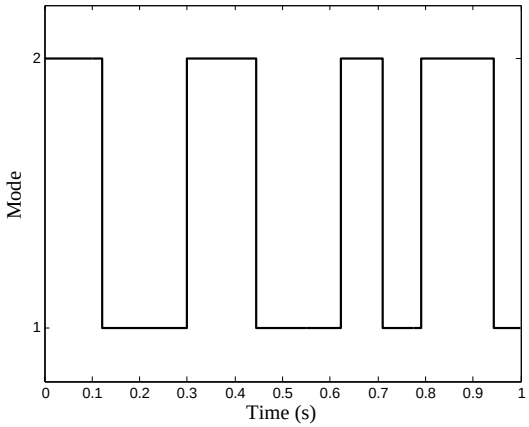

Fig. 5: Optimal switching times.

states change the variation law, which is the so-called vibration phenomenon [36]. The switching time and its corresponding mode are given by

$$\Big\{(0s,2),(0.18s,1),(0.356s,2),(0.533s,1) \\ (0.707s,2),(0.886s,1)\Big\} \tag{44}$$

Considering the multiple objective case, let $\boldsymbol{Q}$ be $\boldsymbol{Q}_1$, and $\boldsymbol{R}$ be $\boldsymbol{R}_1$; choose $\boldsymbol{Q}_2 = \mathrm{diag}([0\ 0\ 0\ 3\ 3\ 3])$, and $\boldsymbol{R}_2 = \boldsymbol{R}_1$. The objective for attacking g_3 puts more emphasis on the attacked state, so the oscillation frequencies of attacked states are increased (observe Figs. 2 and 4).

V. Conclusions

This work addressed the design of optimal partial feedback based switching data injection attacks for cyber-physical power systems The goal is to manipulate a subset of control signals, and alter the attack locations persistently to degrade system performance. Explicit forms for optimal coefficients of partial feedback attack matrices were provided when attacked signals satisfy certain canonical forms. We showed that the nonconvex Hamilton function of the optimal switching attack design problem can be reduced into simpler forms involving switch inputs. Leveraging convex relaxation and Pontryagin maximum principle, we further proved that all optimal switch inputs are attained at least at a vertex of the convexified counterpart, and derived the switching condition to select the optimal attack locations. Case studies were presented to assess the power system vulnerabilities, as well as practical merits of the theory on the IEEE 9-bus benchmark system.

References

[1] M. D. Ilic, L. Xie, U. A. Khan, and J. M. F. Moura, "Modeling of future cyber-physical energy systems for distributed sensing and control," *IEEE Trans. Syst., Man, Cybern. A*, vol. 40, no. 4, pp. 825–838, July 2010.

[2] F. Pasqualetti, F. Dörfler, and F. Bullo, "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design," in *Proc. of Decision and Control and European Control Conf.* Orlando, FL: IEEE, 2011, pp. 2195–2201.

[3] M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, "Multi-agent approach for enhancing security of protection schemes in cyber-physical energy systems," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 436–447, Apr. 2017.

[4] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proc. IEEE*, vol. 105, no. 7, pp. 1389–1407, Jul. 2017.

[5] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 3023–3028, Nov. 2015.

[6] S. Z. Yong, M. Zhu, and E. Frazzoli, "Switching and data injection attacks on stochastic cyber-physical systems: Modeling, resilient estimation, and attack mitigation," *ACM Trans. Cyber-Physical Syst.*, vol. 2, no. 2, pp. 9:1–9:2, Jun. 2018.

[7] Z. Feng, G. Hu, and G. Wen, "Distributed consensus tracking for multi-agent systems under two types of attacks," *Int. J. Robust. Nonlin. Control*, vol. 26, no. 5, pp. 896–918, Apr. 2016.

[8] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Preventing time-delay switch attack on load frequency control in distributed power systems," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 1176–1185, Mar. 2016.

[9] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic load altering attacks against power system stability: Attack models and protection schemes," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2862–2872, Jul. 2018.

[10] C. Liu, J. Wu, C. Long, and Y. Wang, "Dynamic state recovery for cyber-physical systems under switching location attacks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 14–22, Mar. 2017.

[11] Z. Feng, G. Wen, and G. Hu, "Distributed secure coordinated control for multiagent systems under strategic attacks," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1273–1284, May 2017.

[12] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal DoS attack scheduling in wireless networked control system," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 3, pp. 843–852, May 2016.

[13] S. Liu, S. Mashayekh, D. Kundur, T. Zourntos, and K. Butler-Purry, "A framework for modeling cyber-physical switching attacks in smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 2, pp. 273–285, Dec. 2013.

[14] A. Farraj, E. Hammad, A. Al Daoud, and D. Kundur, "A game-theoretic analysis of cyber switching attacks and mitigation in smart grid systems," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1846–1855, July 2016.

[15] G. Wu, J. Sun, and J. Chen, "Optimal data injection attacks in cyber-physical systems," *IEEE Trans. Cybern.*, no. 99, pp. 1–11, Jun. 2018.

[16] F. Pasqualetti, A. Bicchi, and F. Bullo, "A graph-theoretical characterization of power network vulnerabilities," in *Proc. of American Control Conf.*, San Francisco, CA, USA, 2011, pp. 3918–3923.

[17] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*. McGraw-hill New York, 1994, vol. 7.

[18] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Trans. on Autom. Control*, vol. 59, no. 5, pp. 1177–1189, May 2014.

[19] S. Amin, X. Litrico, S. Sastry, and A. M. Bayen, "Cyber security of water scada systemspart i: Analysis and experimentation of stealthy deception attacks," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1963–1970, Sep. 2013.

[20] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 2027–2051, Mar. 2016.

[21] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1294–1305, July 2013.

[22] R. Liu, C. Vellaithurai, S. S. Biswas, T. T. Gamage, and A. K. Srivastava, "Analyzing the cyber-physical impact of cyber events on the power grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2444–2453, Sep. 2015.

[23] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *Proc. of IEEE Conf. on Decision and Control*, Osaka, Japan, Dec. 15-18, 2015, pp. 5162–5169.

[24] Zhian Zhong, Chunchun Xu, B. J. Billian, Li Zhang, S. . S. Tsai, R. W. Conners, V. A. Centeno, A. G. Phadke, and Yilu Liu, "Power system frequency monitoring network (fnet) implementation," *IEEE Trans. on Power Syst.*, vol. 20, no. 4, pp. 1914–1921, Nov. 2005.

[25] Y. Zhang, L. Wang, Y. Xiang, and C. Ten, "Power system reliability evaluation with scada cybersecurity considerations," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1707–1721, July 2015.

[26] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1521–1530, Aug. 2012.

[27] M. Zhu and S. Martnez, "Attack-resilient distributed formation control via online adaptation," in *Proc. of Decision and Control and European Control Conf.* Orlando, FL: IEEE, 2011, pp. 6624–6629.

[28] L. Shi, Y. Yuan, and J. Chen, "Finite horizon lqr control with limited controller-system communication," *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1835–1841, July 2013.

[29] M. M. Kogan, "Solution to the inverse problem of minimax control and worst case disturbance for linear continuous-time systems," *IEEE Trans. Autom. Control*, vol. 43, no. 5, pp. 670–674, May 1998.

[30] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska, and J. Dong, "Finite energy and bounded attacks on control system sensor signals," in *Proc. of American Control Conf.*, Portland, OR, 2014, pp. 1716–1722.

[31] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *2013 American Control Conference*, Washington, DC, USA, 2013, pp. 3344–3349.

[32] M. N. Kurt, Y. Yilmaz, and X. Wang, "Real-time detection of hybrid and stealthy cyber-attacks in smart grid," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 2, pp. 498–513, Feb. 2019.

[33] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 106–117, Mar. 2017.

[34] X. Xu and P. J. Antsaklis, "Optimal control of switched systems based on parameterization of the switching instants," *IEEE Trans. Autom. Control*, vol. 49, no. 1, pp. 2–16, Jan. 2004.

[35] F. Zhu and P. J. Antsaklis, "Optimal control of hybrid switched systems: A brief survey," *Discrete Event Dyn. Syst.*, vol. 25, no. 3, pp. 345–364, Sept. 2015.

[36] G. Wu, J. Sun, and J. Chen, "Optimal linear quadratic regulator of switched systems," *IEEE Trans. Autom. Control*, to appear 2018.

[37] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Partial state feedback control of induction motors with magnetic saturation: Elimination of flux measurements," *Automatica*, vol. 38, no. 2, pp. 191–203, Feb. 2002.

[38] G. Song and G. Tao, "A model reference adaptive control scheme with partial-state feedback for output tracking," in *Proc. of American Control Conf.*, Seattle, WA, USA, 2017, pp. 2465–2470.

[39] S. C. Bengea and R. A. DeCarlo, "Optimal control of switching systems," *Automatica*, vol. 41, no. 1, pp. 11–27, Jan. 2005.

[40] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.