A secure state estimation algorithm for nonlinear systems under sensor attacks

Michelle S. Chong, Henrik Sandberg, João P. Hespanha

***Abstract*— The state estimation of continuous-time nonlinear systems in which a subset of sensor outputs can be maliciously controlled through injecting a potentially unbounded additive signal is considered in this paper. Analogous to our earlier work for continuous-time linear systems in [1], we term the convergence of the estimates to the true states in the presence of sensor attacks as 'observability under M attacks', where M refers to the number of sensors which the attacker has access to. Unlike the linear case, we only provide a sufficient condition such that a nonlinear system is observable under M attacks. The condition requires the existence of asymptotic observers which are robust with respect to the attack signals in an input-to-state stable sense. We show that an algorithm to choose a compatible state estimate from the state estimates generated by a bank of observers achieves asymptotic state reconstruction. We also provide a constructive method for a class of nonlinear systems to design state observers which have the desirable robustness property. The relevance of this study is illustrated on monitoring the safe operation of a power distribution network.**

I. Introduction

The *cyber* security of dynamical systems have garnered the attention of our community in the past decade, see [2] and [3] for a tutorial overview. This is indeed a timely concern as the increasing (cyber) connectivity between physical systems creates vulnerabilities where malicious cyber attacks can lead to disastrous consequences.

The focus of this work is on the state estimation of nonlinear dynamical systems where the sensors have been compromised. This context has been studied in detail for linear systems in both discrete [4], [5], [6], [7], [8], [9], [10] and continuous-time [1], [11]. The main thread that underlies achieving state reconstruction is characterizing the number of sensors which are allowed to be attacked, and the resulting algorithm is an optimization problem which is combinatorial in nature. The computational complexity of these algorithms are addressed in various ways including transforming an l_0 minimization problem into a convex one [4], using gradient descent algorithms [5], employing Satisfiability Modulo Theory (SMT) solvers to reduce search time [6] and to reduce the number of candidates [11], [10], [9], [8], to name a few.

M. Chong is with the Control Systems Technology section at the Department of Mechanical Engineering, Eindhoven University of Technology. m.s.t.chong@tue.nl

H. Sandberg is with the Division of Decision and Control Systems at KTH Royal Institute of Techonology. hsan@kth.se

J. Hesphana is with the Electrical and Computer Engineering Department at the University of California Santa Barbara. hespanha@ece.ucsb.edu

This material is based upon work supported by the U.S. Office of Naval Research under the MURI grant No. N00014-16-1-2710.

Relatively little work has been done for nonlinear systems, where algorithms were proposed for classes of nonlinear systems in discrete-time [12], [13], [14] and in continuous-time [15], [16]. Feedback linearizable systems are considered in [13] and differentially flat systems in [12], which then enables state estimation using linear techniques. The authors of [14] consider Lur'e systems and employs the same framework as in [1]. An adaptive observer is designed to estimate both the states and the attack signals for asymptotically stable nonlinear systems in [16]. In [15], a uniformly observable nonlinear system is considered and a high gain observer is designed for each measured output. An algorithm which exploits redundancy then collects all the state estimates and provides a state estimate.

In this paper, we consider a continuous-time nonlinear system with N outputs where each is measured by a potentially compromised sensor. Under the scenario where M out of the N sensors have been maliciously manipulated, we aim to reconstruct the states given that we do not know which M of the N sensors have been compromised. If this objective is met, we call such a system *observable under M attacks*, a term coined in our earlier work for linear systems [1].

We first provide a sufficient condition in Section IV for observability under M attacks. The condition calls for the total number of sensors N to be larger than twice the number of attacked sensors M, i.e. $N > 2M$. Moreover, it also requires an observer to be constructed for every combination of $N - 2M$ sensor measurements received by the observer, with the crucial property that the observer is robust with respect to the attack signals. In other words, each observer must have an estimation error system which is input-to-state stable (ISS) [17] with respect to the attack signals. These conditions are consistent with the key results in the literature for linear systems [4], [1] and a class of nonlinear systems [15].

This gives rise to an algorithm in Section V, which employs the same framework proposed in an earlier work for linear systems in [1]. The algorithm uses a bank of observers designed to satisfy the aforementioned properties and picks the state estimate which satisfies a consistency measure involving a subset of the other state estimates. The chosen state estimate is shown to converge asymptotically to the true state, in the presence of sensor attacks, provided that the system is M attack observable.

In Section VI, we consider a class of nonlinear systems and provide a systematic method for designing observers which have the desired ISS property with respect to the attack signal. This work is highly relevant in the remote

monitoring of the local voltage regulation of each customer who is connected to a power distribution network, which we present in Section VII. Due to page limitations, all proofs are provided in the full version of this paper [18].

II. Preliminaries

- Let $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_{\geq 0} = [0, \infty)$, $\mathbb{R}_{>0} = (0, \infty)$.
- Let the set of complex numbers be denoted by $\mathbb{C}$.
- We denote the set of integers $\{i, i+1, i+2, \ldots, i+k\}$ as $\mathbb{N}_{[i,i+k]}$.
- The number of k-element subsets of an n-element set is denoted $\binom{n}{k}$.
- Let (u, v) where $u \in \mathbb{R}^{n_u}$ and $v \in \mathbb{R}^{n_v}$ denote the column vector $(u^T, v^T)^T$.
- The cardinality of a set $\mathcal{J}$ is denoted as $\#(\mathcal{J})$.
- The identity matrix of dimension n is denoted by $\mathbb{I}_n$ and a matrix of dimension m by n with all elements 1 is denoted by $\mathbf{1}_{m\times n}$.
- A diagonal matrix with elements d_i, $i \in \mathbb{N}_{[1,n]}$ is denoted by $\text{diag}(d_1, d_2, \ldots, d_n)$.
- Given a symmetric matrix P, its maximum (minimum) eigenvalue is denoted by $\lambda_{\max}(P)$ ($\lambda_{\min}(P)$).
- The infinity norm of a vector $x \in \mathbb{R}^n$, is denoted $|x| := \max_{i\in\mathbb{N}_{[1,n]}} |x_i|$ and for a matrix $A \in \mathbb{R}^{n\times n}$, $|A| := \max_{i\in\mathbb{N}_{[1,n]}} \sum_{j\in\mathbb{N}_{[1,n]}} |a_{ij}|$, where a_{ij} is the row i-th and column j-th element of matrix A.
- A continuous function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a class $\mathcal{K}$ function, if it is strictly increasing and $\alpha(0) = 0$; additionally, if $\alpha(r) \to \infty$ as $r \to \infty$, then α is a class $\mathcal{K}_\infty$ function. A continuous function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a class $\mathcal{KL}$ function, if: (i) $\beta(., s)$ is a class $\mathcal{K}$ function for each $s \geq 0$; (ii) $\beta(r, .)$ is non-increasing and (iii) $\beta(r, s) \to 0$ as $s \to \infty$ for each $r \geq 0$.

III. Problem statement

We consider the problem of state observation for a class of nonlinear systems under sensor attacks of the following form

$$\begin{aligned} \dot{x} &= f(x, z, w),\ z = (z_1, z_2, \ldots, z_N), \\ z_i &= h_i(x, w), \\ y_i &= z_i + a_i, \qquad i \in \mathbb{N}_{[1,N]}, \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^{n_x}$ is the state, $y_i \in \mathbb{R}^{n_i}$ is the measured output at sensor i, $w \in \mathbb{R}^{n_u}$ is a measured input, f and h_i are locally Lipschitz functions and $a_i \in \mathbb{R}^{n_i}$ is a possibly unbounded attack signal that cannot be measured.

Assumption 1: Further assumptions about the attack signals a_i are

(i) Sensors $i \in \mathbb{N}_{[1,N]}$ which are not under attack satisfy $a_i(t) = 0$, for all $t \geq 0$.

(ii) Given an index set $\mathcal{I} \subseteq \mathbb{N}_{[1,N]}$, the set of non-attacked sensors remain constant, i.e. the attack vector $a = (a_1, a_2, \ldots, a_N) \in \mathcal{N}_\mathcal{I}$, where $\mathcal{N}_\mathcal{I} := \{(a_1, a_2, \ldots, a_N) : a_i(t) = 0, \forall t \geq 0, \forall i \notin \mathcal{I}\}$.

□

In this paper, we derive conditions such that the state x of system (1) with N outputs can be estimated when M of the sensors have been attacked, which we term *observable under M attacks* and formally define below.

Definition 1: System (1) is **observable under M attacks** if for any

- initial conditions $x(0)$, $\bar{x}(0) \in \mathbb{R}^{n_x}$,
- measured input $w \in \mathbb{R}^{n_u}$,
- index sets $\mathcal{I}_a$, $\mathcal{I}_b \subset \mathbb{N}_{[1,N]}$ with not more than M elements,
- attack vectors $a = (a_1, a_2, \ldots, a_N) \in \mathcal{N}_{\mathcal{I}_a}$, $\bar{a} = (\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_N) \in \mathcal{N}_{\mathcal{I}_b}$,

there exists an index set $\mathcal{J} \subset \mathbb{N}_{[1,N]}$ with at least $N - 2M$ elements, such that the output trajectories of system (1) satisfy

$$\begin{aligned} &\underline{\alpha}_y\left(|x(0) - \bar{x}(0)|, t\right) \\ &\quad \leq |y_i(t; x(0), w, a_i) - y_i(t; \bar{x}(0), w, \bar{a}_i)| \\ &\quad \leq \bar{\alpha}_y\left(|x(0) - \bar{x}(0)|\right), \end{aligned} \tag{2}$$

for all $i \in \mathcal{J}$, $t \geq 0$, and $\underline{\alpha}_y$ is a class $\mathcal{KL}$ function and $\bar{\alpha}_y$ is a class $\mathcal{K}$ function. □

We have denoted the output trajectories of system (1) initialized at $x(0)$ for the input w and attack a_i as $y_i(t; x(0), w, a_i)$, for all $i \in \mathbb{N}_{[1,N]}$.

Definition 1 means that when a system (1) is observable under M attacks, there is at most one initial condition in which system (1) generates a compatible measured output y_i for any given input signal w, for at least $N - 2M$ of the measured outputs. This has to be achieved regardless of which of the M sensors have been compromised and the attack signal a_i that has been chosen by the attacker.

IV. A sufficient condition for observabilty under M attacks

We provide a sufficient condition for system (1) to be observable under M attacks.

Theorem 1: For any integer $M \geq 0$, (ii) implies (i):

(i) System (1) is observable under M attacks.

(ii) $N > 2M$ and, for every set $\mathcal{J} \subset \mathbb{N}_{[1,N]}$ with $\#(\mathcal{J}) \geq N - 2M$, there exists a function $\hat{f} : \mathbb{R}^{n_x} \times \mathbb{R}^{\#(\mathcal{J})} \times \mathbb{R}^{\#(\mathcal{J})} \to \mathbb{R}^{n_x}$ such that the solution to

$$\dot{\hat{x}}_\mathcal{J} = \hat{f}\left(\hat{x}_\mathcal{J}, y_\mathcal{J}, w\right), \tag{3}$$

and the solution to system (1), respectively satisfy

$$\begin{aligned} |x(t) - \hat{x}_\mathcal{J}(t)| \leq\ &\hat{\beta}\left(|x(0) - \hat{x}_\mathcal{J}(0)|, t\right) \\ &+ \hat{\gamma}\left(\sup_{s\in[0,t)} |a_\mathcal{J}(s)|\right), \end{aligned} \tag{4}$$

for all $t \geq 0$ and initial conditions $x(0)$, $\hat{x}_\mathcal{J}(0) \in \mathbb{R}^{n_x}$, where $\hat{\beta}$ is a $\mathcal{KL}$ function, $\hat{\gamma}$ is a $\mathcal{K}_\infty$ function, and $a_\mathcal{J}$ denotes a stacked vector of a_i indexed by $i \in \mathcal{J}$.

□

Theorem 1 specifies that the number of available sensors N has to be strictly larger than twice the number of compromised sensors M. This is consistent with the results for

linear systems in [1] for continuous-time systems and [4] for discrete-time systems, as well as in [15] and [14] for classes of nonlinear systems in continuous and discrete-time, respectively.

Further, condition (ii) means that the estimation error $e_{\mathcal{J}} := x - \hat{x}_{\mathcal{J}}$ system constructed out of system (1) and (3) is input-to-state stable (ISS) [17] with respect to the attack vector $a_{\mathcal{J}}$. This property can be fulfilled with Luenberger observers in the case of linear systems (see [1, Section III.B]), and with high gain observers [19] or circle criterion observers [20] for classes of nonlinear systems. We will provide a constructive example of our case study in Section VI.

V. Algorithm

Using Theorem 1, we formulate the following estimation algorithm to estimate the states of system (1) when M out of N of its sensors have been compromised. Our algorithm follows the idea presented in [1], where the results were derived for linear dynamical systems.

The crux of the algorithm lies in the fact that for each combination of $N - M$ outputs (note that this is greater than $N - 2M$ outputs, which satisfies condition (ii) in Theorem 1), there is one observer which receives attack-free sensor outputs and hence provides state estimates that converges to the true state. Further, for each set of these $N - M$ outputs, there is at least one subset consisting of $N - 2M$ outputs which is attack-free. Thus, the observer which receives the attack-free subset of $N - 2M$ outputs will provide a state estimate which converge to the true one. Therefore, in the algorithm presented in this section, we employ two banks of observers: one bank of observers employing $N-M$ outputs, and the other employing $N-2M$ outputs.

Suppose that at most M out of N of system (1)'s outputs can be compromised and condition (ii) of Theorem 1 holds. Then, for every set $\mathcal{S} \subset \mathbb{N}_{[1,N]}$ of $N - M$ elements, an observer which employs $N - M$ outputs from system (1) is constructed as follows

$$\dot{\hat{x}}_{\mathcal{S}} = \hat{f}(\hat{x}_{\mathcal{S}}, y_{\mathcal{S}}, w), \tag{5}$$

which has an estimation error system that is ISS with respect to the attack vector $a_{\mathcal{S}}$ as stated in (4) of Theorem 1. This forms the first bank of $\binom{N}{N-M}$ observers. We define the consistency measure $\pi_{\mathcal{S}}$ to be the worst case deviation between the estimate $\hat{x}_{\mathcal{S}}$ given by (5) and the estimate $\hat{x}_{\mathcal{P}}$ generated in the same manner as (5) for $\mathcal{P} \subset \mathcal{S}$ with $N-2M$ elements, which is

$$\pi_{\mathcal{S}}(t) = \max_{\mathcal{P} \subset \mathcal{S} : \#(\mathcal{P}) = N-2M} \left| \hat{x}_{\mathcal{S}}(t) - \hat{x}_{\mathcal{P}}(t) \right|. \tag{6}$$

For the set $\mathcal{S}$ in which all the attack vectors are zero, i.e. $a_i(t) = 0$, for all $i \in \mathcal{S}$ and $t \geq 0$, all the state estimates $\hat{x}_{\mathcal{S}}$ and $\hat{x}_{\mathcal{P}}$ will be consistent and this motivates the choice of

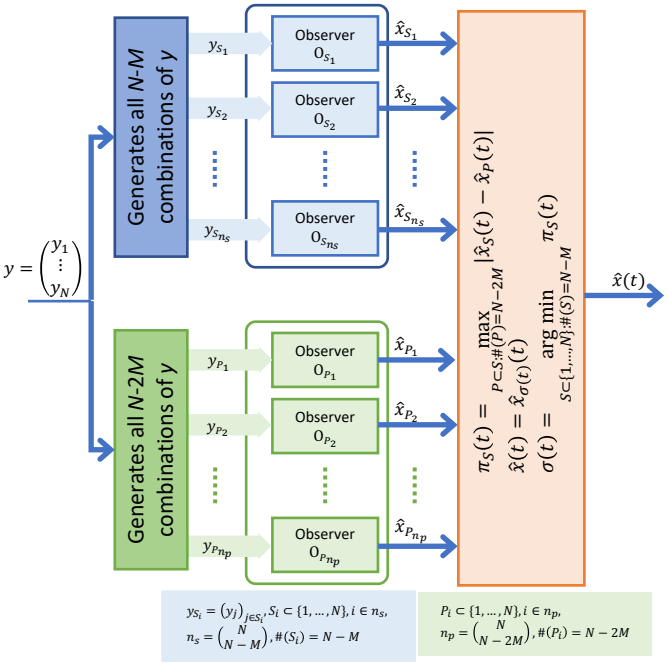

Fig. 1. Infrastructure of a secure state estimation algorithm for nonlinear systems (1).

the state estimate $\hat{x}$ produced by the algorithm as follows

$$\begin{aligned} \hat{x}(t) &= \hat{x}_{\sigma(t)}(t), \\ \sigma(t) &= \underset{\mathcal{S} \subset \mathbb{N}_{[1,N]} : \#(\mathcal{S}) = N-M}{\arg\min} \pi_{\mathcal{S}}(t). \end{aligned} \tag{7}$$

We summarize the algorithm (5), (6), (7) in Figure 1 and provide the following state estimation convergence guarantees.

Theorem 2: Consider system (1) with N-outputs of which at most M is compromised, i.e. the attack vector a belongs to $\mathcal{N}_{\mathcal{I}}$, for some set $\mathcal{I} \subset \mathbb{N}_{[1,N]}$ where $\#(\mathcal{I}) \leq M$. Assuming that (ii) of Theorem 1 holds, then there exists a class $\mathcal{KL}$ function $\tilde{\beta}$ such that the solution to system (1) and the secure state estimation algorithm (5), (6), (7) satisfy

$$|x(t) - \hat{x}(t)| \leq \tilde{\beta}\left(|x(0) - \hat{x}(0)|, t\right), \ \forall t \geq 0, \tag{8}$$

for any initial conditions $x(0)$, $\hat{x}_{\mathcal{S}}(0)$, $\hat{x}_{\mathcal{P}}(0) \in \mathbb{R}^{n_x}$. □

VI. Case Study: A Class of Nonlinear Systems

We consider a specific form of system (1) as follows:

$$\begin{aligned} \dot{x} &= Ax + \phi(z), \ \phi(z) = (\phi_1(z_1), \phi_2(z_2), \dots, \phi_N(z_N)), \\ z_i &= H_i x + w_i, \ i \in \mathbb{N}_{[1,N]}, \\ y_i &= z_i + a_i, \end{aligned} \tag{9}$$

where the nonlinearities $\phi_i : \mathbb{R}^{n_i} \to \mathbb{R}$ are slope-restricted, i.e.

Assumption 2: For $i \in \mathbb{N}_{[1,N]}$, the nonlinearity ϕ_i satisfies

$$\underline{d}_i \leq \frac{\phi_i(\xi) - \phi_i(\psi)}{\xi - \psi} \leq \bar{d}_i, \ \forall \xi, \psi \in \mathbb{R}, \ \xi \neq \psi. \tag{10}$$

□

For system (9) with N outputs, of which M can be compromised, we show that (ii) of Theorem 1 is satisfied by designing each observer (3) in the following manner for every set $\mathcal{J} \subset \mathbb{N}_{[1,N]}$.

$$\begin{aligned} \dot{\hat{x}}_{\mathcal{J}} &= A\hat{x}_{\mathcal{J}} + \phi(\xi_{\mathcal{J}}) + \bar{o}(L_{\mathcal{J}}, y_{\mathcal{J}}, \hat{x}_{\mathcal{J}}, w_{\mathcal{J}}), \\ \xi_{\mathcal{J}} &= H\hat{x}_{\mathcal{J}} + w + \bar{o}(K_{\mathcal{J}}, y_{\mathcal{J}}, \hat{x}_{\mathcal{J}}, w_{\mathcal{J}}), \end{aligned} \tag{11}$$

where $\bar{o}(K_{\mathcal{J}}, y_{\mathcal{J}}, \hat{x}_{\mathcal{J}}, w_{\mathcal{J}}) = K_{\mathcal{J}}\left(y_{\mathcal{J}} - (H_{\mathcal{J}}\hat{x}_{\mathcal{J}} + w_{\mathcal{J}})\right)$ is an output injection term employing outputs $y_{\mathcal{J}}$, known inputs $w_{\mathcal{J}}$ and an observer matrix $K_{\mathcal{J}}$ to be designed. Note that the first two terms in $\xi_{\mathcal{J}}$ use the full H from system (9) and all the known inputs w, respectively.

Proposition 1: Consider system (9) under Assumption 2. Suppose $N > 2M$ and, for every set $\mathcal{J} \subset \mathbb{N}_{[1,N]}$ with $\#(\mathcal{J}) \geq N - 2M$, there exist a matrix $P_{\mathcal{J}} = P_{\mathcal{J}}^T > 0$, scalars $\nu_{\mathcal{J}} \geq 0$, $\mu_{\mathcal{J}} \geq 0$ and observer matrices $K_{\mathcal{J}}$ and $L_{\mathcal{J}}$ such that the following holds

$$\begin{bmatrix} \mathcal{A}(P_{\mathcal{J}}, P_{\mathcal{J}}L_{\mathcal{J}}, \nu_{\mathcal{J}}) & \mathcal{B}(P_{\mathcal{J}}, K_{\mathcal{J}}) & -P_{\mathcal{J}} \\ \mathcal{B}(P_{\mathcal{J}}, K_{\mathcal{J}})^T & \mathcal{D}(\bar{d}) & 0 \\ -P_{\mathcal{J}} & 0 & -\mu_{\mathcal{J}}\mathbb{I}_{n_{\mathcal{J}}} \end{bmatrix} \leq 0, \tag{12}$$

where

- $\mathcal{A}(P_{\mathcal{J}}, P_{\mathcal{J}}L_{\mathcal{J}}, \nu_{\mathcal{J}}) := P_{\mathcal{J}}(A - L_{\mathcal{J}}H_{\mathcal{J}}) + (A - L_{\mathcal{J}}H_{\mathcal{J}})^T P_{\mathcal{J}} + \nu_{\mathcal{J}}\mathbb{I}_{n_x}$,
- $\mathcal{B}(P_{\mathcal{J}}, K_{\mathcal{J}}) := P_{\mathcal{J}} + (H - K_{\mathcal{J}}H_{\mathcal{J}})^T$,
- $\mathcal{D}(\bar{d}) := -2\,\mathrm{diag}\left(\bar{d}_1^{-1}, \bar{d}_2^{-1}, \ldots, \bar{d}_N^{-1}\right)$,
- $n_{\mathcal{J}} := \sum_{i \in \mathcal{J}} n_i$.

Then, (ii) of Theorem 1 holds. □

Inequality (12) is a linear matrix inequality (LMI) in $P_{\mathcal{J}}$, $P_{\mathcal{J}}L_{\mathcal{J}}$, $\nu_{\mathcal{J}}$, $K_{\mathcal{J}}$ and $\mu_{\mathcal{J}}$, which can be solved efficiently using computational tools. The design we have used here was first introduced as the circle criterion observer in [20], which can be tuned to attenuate measurement noise and input disturbances according to the design in [21]. Here, we have adapted the design such that the observer (11) is robust with respect to the attack vector $a_{\mathcal{J}}$ in the sense of (4).

VII. Application: Secure Monitoring for the Voltage Regulation of a Power Distribution Network

A typical low voltage power distribution network (shown in Figure 2) would consist of N customers feeding into the distribution network in a line configuration, with the smart secondary substation at the head of the line. The substation functions as a monitoring center, sending the desired set-point voltage $\bar{v}$ to each local controller Σ_i, such that the voltages received by each customer v_i is regulated to operate in a safe operating range, i.e. for a given $\delta > 0$,

$$\bar{v} - \delta \leq v_i(t) \leq \bar{v} + \delta, \qquad \forall t \geq 0. \tag{13}$$

For each customer $i \in \mathbb{N}_{[1,N]}$, the received voltage level is v_i and the voltage level at the point of connection with the distribution line is v_i', with a corresponding line impedance $Z_i' = R_i' + jX_i'$ in between customer i and the connection point on the distribution line, where $R_i' \in \mathbb{R}_{\geq 0}$ is the resistance and $X_i' \in \mathbb{R}_{\geq 0}$ is the reactance. In between each connection point, the corresponding line impedance is $Z_i = R_i + jX_i$, where $R_i \in \mathbb{R}_{\geq 0}$ is the resistance and $X_i \in \mathbb{R}_{\geq 0}$ is the reactance. Each customer has a load which can consume reactive $q_{c,i}$ and active powers $\rho_{c,i}$, independently of the generated reactive $q_{g,i}$ and active powers $\rho_{g,i}$.

In [22], a class of sector-bounded droop controllers Σ_i which uses local measurements v_i were shown to regulate the voltages v_i such that the safety constraint (13) is satisfied. This is achieved via appropriate injection of reactive power $q_{g,i}$ by each local controller Σ_i to regulate the flow of active P_i and reactive Q_i powers, under the assumption that the net injected active power ρ_i and the reactive power $q_{c,i}$ consumed by customer i is bounded and the bounds are known.

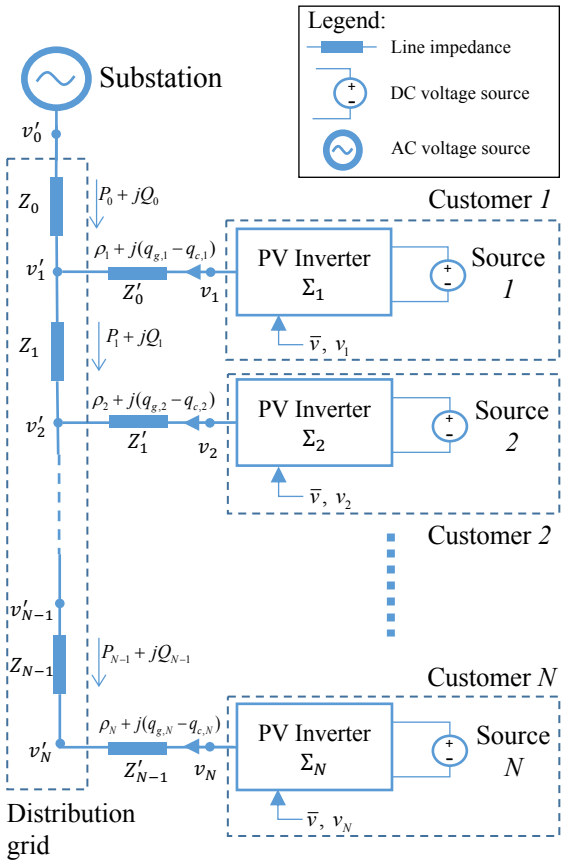

Fig. 2. Infrastructure of a low-voltage grid. Voltage regulation is achieved via local controllers Σ_i, for $i \in \mathbb{N}_{[1,N]}$, with the monitoring center (situated at the substation) receiving potentially corrupted measurements of v_i.

We are now concerned with the *security* problem where the measurements of the voltages v_i received at the monitoring center situated at the substation has been maliciously corrupted. We model this measurement corruption with an additive attack signal $\alpha_i : \mathbb{R}_{\geq 0} \to \mathbb{R}$ which is potentially unbounded, as follows

$$\hat{y}_i = v_i + \alpha_i. \tag{14}$$

This is a major issue as the presence of the attack signal

α_i would mislead the monitoring center into thinking that the safety constraint (13) has been violated and thus a false alarm is raised and possibly triggering unwarranted operator actions.

Our solution is to employ the results in the previous sections to estimate the voltages v_i, given that M out of N of the measurements y_i are maliciously manipulated. To this end, we model the power flow in the distribution grid as done in [22] using the linearized DistFlow model [23] and assume that the droop controllers Σ_i have been designed using the methodology presented in [22]. The relationship between the power flow and voltages between key nodes is

$$\begin{array}{rcl} P_{i+1} & = & P_i + \rho_{i+1}, \\ Q_{i+1} & = & Q_i + q_{i+1}, \\ {v'_{i+1}}^2 & = & {v'_i}^2 - 2\beta_i(P_i, Q_i), \\ {v'_i}^2 & = & {v_i}^2 - 2\beta'_{i-1}(\rho_i, q_i), \end{array} \tag{15}$$

where P_i and Q_i are the respective total active and reactive powers flowing from customer i to customer $i+1$; $\rho_i := \rho_{g,i} - \rho_{c,i}$ and $q_i := q_{g,i} - q_{c,i}$ are the net injection of the respective active and reactive power into the distribution line from customer i; $\beta_i(r,s) := R_i r + X_i s$ and $\beta'_i(r,s) := R'_i r + X'_i s$ with $\beta'_{-1}(r,s) = 0$ for all $r, s \in \mathbb{R}$.

Each local controller Σ_i actuated by the inverter can generate reactive power $q_{g,i}$ as follows

$$\dot{q}_{g,i} = -\frac{1}{\tau_i} q_{g,i} + \frac{1}{\tau_i} K_i(\bar{v}^2 - v_i^2), \tag{16}$$

where $\tau_i \in \mathbb{R}_{>0}$ is the time-constant of the inverter's response, $\bar{v} \in \mathbb{R}$ is the reference voltage communicated to each customer i and the droop function $K_i(w)$ is a static mapping from the difference of the squared voltages w to the set-point for the reactive power. We choose the droop function $K_i(w)$ to be a piecewise saturation function considered in [24] which takes the following form:

$$K_i(w) := \begin{cases} -\bar{Q}_i, & w \le w_{\min,i}, \\ -\left(1 - \frac{w - w_{\min,i}}{w_{m,i} - w_{\min,i}}\right) \bar{Q}_i & w \in (w_{\min,i}, w_{m,i}], \\ 0, & w \in (w_{m,i}, w_{n,i}], \\ \left(\frac{w - w_n}{w_{\max,i} - w_{n,i}}\right) \bar{Q}_i & w \in (w_{n,i}, w_{\max,i}], \\ \bar{Q}_i, & w > w_{\max,i}, \end{cases} \tag{17}$$

where $w_{\min,i} \le w_{m,i} \le 0 \le w_{n,i} \le w_{\max,i}$ are design parameters, $\bar{Q}_i \in \mathbb{R}_{\ge 0}$ is the saturation limit of the i-th inverter satisfying $\bar{Q}_i = \sqrt{\bar{s}_i^2 - \rho_{g,i}^2}$, where $\bar{s}_i \in \mathbb{R}$ is the maximum apparent power of the i-th inverter. The design parameters $w_{\min,i}$, $w_{\max,i}$, $w_{m,i}$, $w_{n,i}$ are chosen such that

$$d_i := \min\left\{ \frac{\bar{Q}_i}{w_{\max,i} - w_{n,i}}, \frac{\bar{Q}_i}{w_{m,i} - w_{\min,i}} \right\} \tag{18}$$

satisfies [22, Theorem 6] such that the safety constraint (13) is met. We employ the same change in state coordinates as done in [22] such that the distribution model (15), controllers (16), and measurements received at the monitoring center (14) can be written in the form of (9) by choosing

- the state $x = (q_{g,1}, q_{g,2}, \ldots, q_{g,N})$,
- $z_i := \bar{v}^2 - v_i^2$,
- the known input $w_i = \phi_i(\rho, q_c) + \bar{v}^2 - {v'}_0^2$, where $\phi_i(\rho, q_c) := \sum_{j \in \mathbb{N}_{[0,i-1]}} \psi_j(\rho, q_c) + \sum_{j \in \mathbb{N}_{[0,i-2]}} 2\beta'_j(\rho_{j+1}, q_{c,j+1})$ and

$$\psi_j(\rho, q_c) := 2X_j \sum_{k \in \mathbb{N}_{[j+1,N]}} q_{c,k} - 2R_j \sum_{k \in \mathbb{N}_{[j+1,N]}} \rho_k - 2\beta'_j(\rho_{j+1}, q_{c,j+1}),$$

  where β'_j is from (15),
- the attack signal $a_i = 2v_i\alpha_i - \alpha_i^2$, where α_i comes from (14),
- H_i to be the rows of the matrix

$$H = -2 \begin{pmatrix} X_0 & X_0 & \cdots & X_0 \\ \star & (X_0 + X_1) & \cdots & (X_0 + X_1) \\ \vdots & \ddots & \ddots & \vdots \\ \star & \cdots & \star & \sum_{i \in \mathbb{N}_{[0,N-1]}} X_i \end{pmatrix} - 2\mathrm{diag}\left(X'_0, \ldots, X'_{N-1}\right),$$

  where $\star$ denotes a block component of a symmetric matrix,
- $A = \mathrm{diag}\left(-1/\tau_1, -1/\tau_2, \ldots, -1/\tau_N\right)$,
- $\phi_i(z_i) = \tau_i^{-1} K_i(z_i)$, which satisfies Assumption 2 with $\underline{d}_i = 0$ and $\bar{d}_i = d_i/\tau_i$, where d_i is defined in (18).

To recapitulate, given that the monitoring center sits remotely at the substation, the objective is to estimate the voltages v_i, given that the monitoring center only has access to the measurements y_i, $i \in \mathbb{N}_{[1,N]}$, where M out of N of these measurements may be corrupted. The main idea is to first estimate the states x of all the controllers Σ_i, $i \in \mathbb{N}_{[1,N]}$, then estimate the voltages v_i via

$$\hat{v}_i(t)^2 = -H_i \hat{x}(t) - \psi_i(\rho, q_c) + {v'}_0^2, \tag{19}$$

where $\hat{x}$ is the state estimate provided by the secure estimation algorithm described in Section V; ψ and ${v'}_0^2$ are known. We have kept the squared form of the voltage $\hat{v}_i^2$ due to the distribution model (15) used. By application of Proposition 1 and Theorem 2, we provide the following guarantee.

Corollary 1: Consider the distribution model (15) and controllers (16) with measurements (14) where $a_{\mathcal{I}}$ belongs to $\mathcal{N}_{\mathcal{I}}$ for some unknown set $\mathcal{I} \subset \mathbb{N}_{[1,N]}$ with at most M elements. Suppose $N > 2M$ and for every $\mathcal{J} \subset \mathbb{N}_{[1,N]}$ with $\#(\mathcal{J}) \ge N - 2M$, there exists a matrix $P_{\mathcal{J}} = P_{\mathcal{J}}^T > 0$, scalars $\nu \ge 0$, $\mu_a \ge 0$ and observer matrices $K_{\mathcal{J}}$ and $L_{\mathcal{J}}$ such that (12) holds. Then, using the secure state estimation algorithm (5), (6), (7), the estimated squared voltages $\hat{v}_i^2$ computed according to (19) converges to to the true squared voltages v_i^2 as follows

$$\left| v_i(t)^2 - \hat{v}_i(t)^2 \right| \le \beta_v\left(\left|x(0) - \hat{x}(0)\right|, t\right), \tag{20}$$

for all $t \ge 0$, $i \in \mathbb{N}_{[1,N]}$, initial conditions $q_{g,i}(0) \in \mathbb{R}$, where $x(0) = (q_{g,1}(0), q_{g,2}(0), \ldots, q_{g,N}(0))$, $\hat{x}(0) = (\hat{q}_{g,1}(0), \hat{q}_{g,2}(0), \ldots, \hat{q}_{g,N}(0))$ and β_v is a class $\mathcal{KL}$ function. □

VIII. Conclusions and future work

We introduced a new definition of observability for nonlinear systems in which a subset of the outputs can be manipulated maliciously. A sufficient condition such that asymptotic state reconstruction can be achieved in the presence of sensor attacks is provided, which requires building a bank of observers with an ISS property. A secure state estimation algorithm is proposed which shows that the framework used for linear continuous time systems in our earlier work [1] can be used for nonlinear systems as well. A systematic method for designing the observers is proposed for a class of nonlinear systems and we showed the relevance of this work in the monitoring of a power distribution network. Future work includes reducing the computational resources needed in terms of time and number of observers required.

References

[1] M. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *2015 American Control Conference (ACC)*, pp. 2439–2444, July 2015.

[2] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Systems Magazine*, vol. 35, pp. 20–23, Feb 2015.

[3] M. S. Chong, H. Sandberg, and A. M. H. Teixeira, "A tutorial introduction to security and privacy for cyber-physical systems," in *2019 18th European Control Conference (ECC)*, pp. 968–978, June 2019.

[4] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[5] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2015.

[6] Y. Shoukry, M. Chong, M. Wakaiki, P. Nuzzo, A. Sangiovanni-Vincentelli, S. Seshia, J. Hespanha, and P. Tabuada, "SMT-based observer design for cyber-physical systems under sensor attacks," *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 1, p. 5, 2018.

[7] C.-H. Xie and G.-H. Yang, "Secure estimation for cyber-physical systems with adversarial attacks and unknown inputs: An L_2-gain method," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 6, pp. 2131–2143, 2018.

[8] A.-Y. Lu and G.-H. Yang, "Switched projected gradient descent algorithms for secure state estimation under sparse sensor attacks," *Automatica*, vol. 103, pp. 503–514, 2019.

[9] A.-Y. Lu and G.-H. Yang, "Secure switched observers for cyber-physical systems under sparse sensor attacks: a set cover approach," *IEEE Transactions on Automatic Control*, vol. 64, no. 9, pp. 3949–3955, 2019.

[10] L. An and G.-H. Yang, "State estimation under sparse sensor attacks: A constrained set partitioning approach," *IEEE Transactions on Automatic Control*, vol. 64, no. 9, pp. 3861–3868, 2018.

[11] L. An and G.-H. Yang, "Secure state estimation against sparse sensor attacks with adaptive switching mechanism," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2596–2603, 2017.

[12] Y. Shoukry, P. Nuzzo, N. Bezzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving," in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 3804–3809, IEEE, 2015.

[13] Q. Hu, D. Fooladivanda, Y. H. Chang, and C. J. Tomlin, "Secure state estimation and control for cyber security of the nonlinear power systems," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1310–1321, 2017.

[14] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "A robust circle-criterion observer-based estimator for discrete-time nonlinear systems in the presence of sensor attacks," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 571–576, IEEE, 2018.

[15] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo, "Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems having redundant sensors," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1162–1169, 2018.

[16] S. Nateghi, Y. Shtessel, J.-P. Barbot, and C. Edwards, "Cyber attack reconstruction of nonlinear systems via higher-order sliding-mode observer and sparse recovery algorithm," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 5963–5968, IEEE, 2018.

[17] E. D. Sontag, "Input to state stability: Basic concepts and results," in *Nonlinear and optimal control theory*, pp. 163–220, Springer, 2008.

[18] M. S. Chong, H. Sandberg, and J. P. Hespanha, "A secure state estimation algorithm for nonlinear systems under sensor attacks," 2020.

[19] G. Bornard and H. Hammouri, "A high gain observer for a class of uniformly observable systems," in *Proceedings of the 30th IEEE Conference on Decision and Control*, pp. 1494–1496, IEEE, 1991.

[20] M. Arcak and P. Kokotović, "Nonlinear observers: a circle criterion design and robustness analysis," *Automatica*, vol. 37, no. 12, pp. 1923 – 1930, 2001.

[21] M. Chong, R. Postoyan, D. Nešić, L. Kuhlmann, and A. Varsavsky, "A robust circle criterion observer with application to neural mass models," *Automatica*, vol. 48, no. 11, pp. 2986–2989, 2012.

[22] M. Chong, D. Umsonst, and H. Sandberg, "Local voltage control of an inverter-based power distribution network with a class of slope-restricted droop controllers," in *Procceedings of the 8th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, 2019. To appear.

[23] M. Baran and F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.

[24] F. Andren, B. Bletterie, S. Kadam, P. Kotsampopoulos, and C. Bucher, "On the stability of local voltage control in distribution networks with a high penetration of inverter-based generation," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2519–2529, 2015.