

# Bayesian State Estimation Against Switching Unknown Inputs and Extra Packet Injections

Nicola Forti, Giorgio Battistelli, Luigi Chisci, and Bruno Sinopoli

**Abstract**—This technical note addresses state estimation of systems subject to switching unknown exogenous inputs and injection of false measurements. The random set paradigm is adopted in order to model, via *Random Finite Sets* (RFSs), the switching nature of the unknown input, which at each instance can be operating or not, as well as the uncertainty arising from the possible reception of extra false packets through the communication network. The problem of jointly detecting the unknown input and estimating the system state in the presence of random false measurements is then formulated and solved in the Bayesian framework leading to the analytical derivation of a *hybrid Bernoulli filter* that updates in real-time the joint posterior density of the unknown input Bernoulli RFS and of the state vector. The effectiveness of the developed tools for joint input detection and resilient state estimation is tested by simulating a cyber-physical attack on a standard IEEE power network.

**Index Terms**—Cyber-physical systems; Bayesian state estimation; Bernoulli filter; extra packet injections; random finite sets.

## I. INTRODUCTION

Due to its application to a wide range of real-world problems (such as fault detection and diagnosis, weather forecasting, tracking of maneuvering targets, etc.) state estimation in the presence of unknown inputs has been an area of intensive study over the last decades for both linear and nonlinear systems under different assumptions (e.g., see [1]–[6] and references therein). However, all this previous work is focused on jointly estimating the state and the unknown input under the assumption that the exogenous signal is always present, without considering the more challenging case of a *switching* external input, that at each time instant may or may not exist. This switching behavior seems to be, nowadays more than ever, appropriate to model unknown exogenous inputs that may represent, not only genuine faults/failures on the physical infrastructure, but also anomalous or even malicious activity (e.g. cyberattacks) affecting the computation and communication layers, i.e. any off-nominal behavior involving abnormal events that modern engineered systems must be able to detect, understand and, possibly, overcome in order to ensure reliability and security. Indeed, it is evident that next-generation cyber-physical systems (CPSs) integrating computation, communication, and physical components, will

be inevitably more vulnerable and prone to different types of misbehavior, unexplained by the available nominal models, against which they need to be *resilient*.

In the context of security of CPSs, preliminary studies addressed the problem of attack detection/identification for deterministic control systems [7] by modeling attacks as unknown inputs. Secure strategies have been studied for *replay* attacks [8], [9] where the adversary first records and then replays the observed data, as well as for *denial-of-service* (DoS) attacks [10], [11] disrupting the flow of data. Furthermore, active detection methods have been designed in order to reveal *stealthy* attacks via manipulation of e.g. control inputs [12] or dynamics [13]. In recent times, the problem of resilient state estimation, i.e. capable of reconstructing the state of the CPS even in the face of some misbehavior or attack, has gained considerable attention [14]–[22]. Initial work considered a worst-case approach for the special family of SISO systems [14]. Under the assumption of linear systems subject to an unknown but bounded number of *false-data injection* attacks on sensor outputs, the problem for a noise-free system has been cast into an  $\ell_0$ -optimization problem, which can be relaxed as a more efficient convex problem [15], and, in turn, adapted to systems with bounded noise [16]. Further advances tried to tackle the combinatorial complexity of the problem by resorting to satisfiability modulo theories [17] and investigated, in the same context, the case of Gaussian measurement noise [18] and the concept of observability under attacks [19]. Deterministic models of different attack policies have been presented based on adversary’s resources and system knowledge [20], and resilient strategies have been proposed for noisy systems with direct feedthrough under both data injection and switching mode attacks [21]. Most recently, in [22] resilient state estimation of CPSs has been addressed by modeling in a stochastic framework the attacker’s decision-making by assuming Markov (possibly uninformative) decision processes instead of unknown or worst-case models.

The present paper aims to address the problem of simultaneously detecting an unknown input while estimating the state of the system, in the presence of possible *extra packet injections*, i.e. multiple false observations (junk packets) added to the system-generated measurement in order to confuse the system monitor and create uncertainty about the origin of the received packets. This is a new type of *man-in-the-middle attack* against state estimation, already introduced in information security (see, e.g., [23], [24]), that can be captured by the proposed modeling framework. A random set approach is undertaken by representing the exogenous input presence/absence by means of a Bernoulli random set (i.e. a set that, with some probability,

N. Forti, G. Battistelli, and L. Chisci are with the Dipartimento di Ingegneria dell’Informazione, Università degli Studi di Firenze, Via Santa Marta 3, 50139 Firenze, Italy. Email: {nicola.forti,giorgio.battistelli,luigi.chisci}@unifi.it.

B. Sinopoli is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. Email: brunos@ece.cmu.edu.

can be either empty or a singleton depending on the presence or not of the external input) and by taking into account the possible injection of a random number of false packets by means of a measurement RFS. The joint input detection-state estimation problem is then formulated within the Bayesian framework as the recursive determination of the joint posterior density of the unknown input Bernoulli set and of the state vector at each time given all the measurement sets available up to that time. Strictly speaking, the posed Bayesian estimation problem is neither standard [25] nor *Bernoulli* filtering [26]–[29] but is rather a *hybrid* Bayesian filtering problem that aims to jointly estimate a Bernoulli random set for the unknown input and a random vector for the system state. An analytical solution of the hybrid filtering problem is found in terms of integral equations that generalize the Bayes and Chapman-Kolmogorov equations for the solution of joint input-and-state estimation (where, in this case, the external input is switching), and of the Bernoulli filter (for a system with unknown inputs). Preliminary results on this topic were presented in [30].

## II. PROBLEM SETUP AND PRELIMINARIES

### A. System description and input model

Let the discrete-time cyber-physical system of interest be modeled by

$$x_{k+1} = \begin{cases} f_k^0(x_k) + w_k, & \text{nominal} \\ f_k^1(x_k, a_k) + w_k, & \text{off-nominal} \end{cases} \quad (1)$$

where:  $k$  is the time index;  $x_k \in \mathbb{R}^n$  is the state vector to be estimated;  $a_k \in \mathbb{R}^m$  is an unknown exogeneous input affecting the system only when it is under off-nominal behavior. For instance, like in the simulation example of Section IV,  $a_k$  can model the effect of a non-strategic attack against the cyber-physical system, able to corrupt sensor/actuator data;  $f_k^0(\cdot)$  and  $f_k^1(\cdot, \cdot)$  are known state transition functions that describe the system evolution in the *nominal* and, respectively, *off-nominal* cases;  $w_k$  is a random process disturbance also affecting the system.

For monitoring purposes, the state of the above system is observed through the measurement model

$$y_k = \begin{cases} h_k^0(x_k) + v_k, & \text{nominal} \\ h_k^1(x_k, a_k) + v_k, & \text{off-nominal} \end{cases} \quad (2)$$

where:  $h_k^0(\cdot)$  and  $h_k^1(\cdot, \cdot)$  are known measurement functions that refer to the *nominal* and, respectively, *off-nominal* cases;  $v_k$  is a random measurement noise. It is assumed that the measurement  $y_k$  is actually delivered to the system monitor with probability  $p_d \in (0, 1]$ , where the non-unit probability might be due to a number of reasons (e.g. temporary denial of service, packet loss, sensor inability to detect or sense the system, etc.). For ease of presentation, we consider the case of a direct feedthrough from the unknown input to the output vector by assuming that, for the off-nominal output function, the Jacobian  $\partial h_k^1(x, a)/\partial a$  has full rank. However, the proposed approach could be extended also to account for an output function  $h_k^1$  depending only on some of the components of the vector  $a_k$ , by considering a one time unit

delay in the estimation of the part of unknown input not entering directly into the function  $h_k^1$  (see Section 4 of [4]). Further, while only the case of a single model of unknown input is taken into account here, multiple models [21] could be accommodated in the considered framework by letting (1)–(2) depend on a discrete variable, say  $\nu_k$ , which specifies the particular input model and has to be estimated together with  $a_k$ . Following [4], hereafter, the unknown input  $\{u_k\}$  is treated as a white process, independent of  $x_0, \{w_k\}$  and  $\{v_k\}$ , so as to model the fact that we cannot predict the value of  $a_k$  from the values of  $x_k$  and  $a_l$ , with  $l < k$ .

Besides the system-originated measurement  $y_k$  in (2), it is assumed that the system monitor might receive extra false measurements from, e.g., some cyber-attacker, which is able to send to the monitor one or multiple counterfeit measurements indistinguishable from the system-originated one (e.g. with same ID and timestamp). For the subsequent developments, it is convenient to introduce the *input set* at time  $k$ ,  $\mathcal{A}_k$ , which is either equal to the empty set if the system is under nominal behavior at time  $k$  or to the singleton  $\{a_k\}$  otherwise, i.e.

$$\mathcal{A}_k = \begin{cases} \emptyset, & \text{if the system is under nominal behavior} \\ \{a_k\}, & \text{otherwise.} \end{cases}$$

It is also convenient to define the *measurement set* at time  $k$ ,  $\mathcal{Z}_k$ , which in the presence of *extra packet injections* can be written as the union of two disjoint sets, i.e.

$$\mathcal{Z}_k = \mathcal{Y}_k \cup \mathcal{F}_k \quad (3)$$

where

$$\mathcal{Y}_k = \begin{cases} \emptyset, & \text{with probability } 1 - p_d \\ \{y_k\}, & \text{with probability } p_d \end{cases} \quad (4)$$

is the set of system-originated measurements and  $\mathcal{F}_k$  the finite set of false measurements.

The aim of this paper is to address the problem of joint input detection and state estimation, which amounts to jointly estimating, at each time  $k$ , the state  $x_k$  and the input RFS  $\mathcal{A}_k$  given the set of measurements  $\mathcal{Z}^k \triangleq \bigcup_{i=1}^k \mathcal{Z}_i$  up to time  $k$ .

### B. Joint input and state estimation

In this section we review the formulation of the *Joint Input and State Estimation* (JISE) problem in the Bayesian framework [4]. To this end, let us consider a system with direct feedthrough of the form

$$\begin{cases} x_{k+1} = f(x_k, u_k) + w_k \\ y_k = h(x_k, u_k) + v_k \end{cases} \quad (5)$$

where  $u_k$  is the unknown input vector. The goal of stochastic Bayesian filtering is to recursively estimate the time-varying posterior PDF of the unknown variables conditioned on all the information available up to that time. Hence, when the objective is the simultaneous input and state estimation, at each time instant  $k$ , the estimates of  $u_k$  and  $x_k$  can be obtained by solving the following problem.

**JISE problem:** For the system (5), given the measurement set  $y^k = \{y_1, y_2, \dots, y_k\}$ , sequentially compute the joint conditional PDF  $p(u_k, x_k | y^k)$  from  $p(u_{k-1}, x_{k-1} | y^{k-1})$ .

Assuming that the initial density  $p(x_0)$  is given, the solution can be described as a two-step procedure of prediction and correction. Let  $p(u_{k-1}, x_{k-1}|y^{k-1})$  denote the posterior PDF at  $k-1$ . The prediction step computes the conditional PDF  $p(x_k|y^{k-1})$  via the Chapman-Kolmogorov equation:

$$p(x_k|y^{k-1}) = \iint p(x_k|u_{k-1}, x_{k-1}) \times p(u_{k-1}, x_{k-1}|y^{k-1}) du_{k-1} dx_{k-1} \quad (6)$$

Then, at time instant  $k$ , the observed output  $y_k$  is available and can be used to update  $p(x_k|y^{k-1})$  and jointly estimate the conditional PDF of  $u_k$ , since  $y_k$  is the first measurement containing information about the unknown signal. The correction step can then be performed by applying the Bayes rule:

$$p(u_k, x_k|y^k) = \frac{p(y_k|u_k, x_k) p(x_k|y^{k-1}) p(u_k)}{p(y_k|y^{k-1})} \quad (7)$$

Notice that, in (7),  $p(u_k)$  is a PDF summarizing the prior knowledge on the input  $u_k$ . When no information on the unknown input  $u_k$  is supposed to be available,  $p(u_k)$  can be taken as an uninformative (flat) prior<sup>1</sup> so that (7) can be rewritten as

$$p(u_k, x_k|y^k) = \frac{p(y_k|u_k, x_k) p(x_k|y^{k-1})}{\int \int p(y_k|u, x) p(x|y^{k-1}) dx du} \quad (8)$$

With the derived Bayesian solution to JISE in the presence of direct feedthrough, optimal (with respect to any criterion) point estimates of the input and state can be obtained from this PDF. For instance, maximization of (8) with respect to  $x_k$  and  $u_k$  provides a Joint MAP-ML (Maximum A-Posteriori Maximum Likelihood) estimate of  $x_k$  and  $u_k$ , respectively [31]. This is a standard approach to address in a unitary framework the estimation of both stochastic quantities (in the considered setting, the system state) and uncertain parameters (the unknown input), without the need of any underlying model for the mechanism generating the latter [32].

### C. Random set estimation

An RFS (*Random Finite Set*)  $\mathcal{X}$  over  $\mathbb{X}$  is a random variable taking values in  $\mathcal{F}(\mathbb{X})$ , the collection of all finite subsets of  $\mathbb{X}$ . The mathematical background needed for Bayesian random set estimation can be found in [27]; here, the basic concepts needed for the subsequent developments are briefly reviewed. From a probabilistic viewpoint, an RFS  $\mathcal{X}$  is completely characterized by its *set density*  $f(\mathcal{X})$ , also called FISST (*Finite Set Statistics*) probability density. In fact, given  $f(\mathcal{X})$ , the cardinality *probability mass function*  $\rho(n)$  that  $\mathcal{X}$  have  $n \geq 0$  elements and the joint PDFs  $f(x_1, x_2, \dots, x_n|n)$  over  $\mathbb{X}^n$  given that  $\mathcal{X}$  have  $n$  elements, are obtained as follows:

$$\begin{aligned} \rho(n) &= \frac{1}{n!} \int_{\mathbb{X}^n} f(\{x_1, \dots, x_n\}) dx_1 \cdots dx_n \\ f(x_1, x_2, \dots, x_n|n) &= \frac{1}{n! \rho(n)} f(\{x_1, \dots, x_n\}). \end{aligned}$$

<sup>1</sup>When the unknown input  $u_k$  takes value in a bounded domain  $\mathbb{U}$  an uninformative prior is simply the uniform distribution over  $\mathbb{U}$ . Such a choice is rooted in the so-called *principle of indifference*. When instead  $\mathbb{U}$  is not bounded, the concept of uninformative prior has to be understood in a generalized sense as the (generalized or improper) distribution for which (8) holds.

In order to measure probability over subsets of  $\mathbb{X}$  or compute expectations of random set variables, Mahler [27] introduced the notion of *set integral* for a generic real-valued function  $g(\mathcal{X})$  of an RFS  $\mathcal{X}$  as

$$\int g(\mathcal{X}) \delta \mathcal{X} = g(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int g(\{x_1, \dots, x_n\}) dx_1 \cdots dx_n \quad (9)$$

In particular, in this work we will consider the Bernoulli RFS, i.e. a random set which can be either empty or, with some probability  $r \in [0, 1]$ , a singleton  $\{x\}$  whose element is distributed over  $\mathbb{X}$  according to the PDF  $p(x)$ . Accordingly, its set density is defined as follows:

$$f(\mathcal{X}) = \begin{cases} 1 - r, & \text{if } \mathcal{X} = \emptyset \\ r \cdot p(x), & \text{if } \mathcal{X} = \{x\} \end{cases} \quad (10)$$

### III. BAYESIAN RANDOM SET FILTER FOR JOINT INPUT DETECTION AND STATE ESTIMATION

Let the unknown input at time  $k$  be modeled as a Bernoulli random set  $\mathcal{A}_k \in \mathcal{B}(\mathbb{A})$ , where  $\mathcal{B}(\mathbb{A}) = \emptyset \cup \mathcal{S}(\mathbb{A})$  is a set of all finite subsets of the unknown input space  $\mathbb{A} \subseteq \mathbb{R}^m$ , and  $\mathcal{S}$  denotes the set of all singletons (i.e., sets with cardinality 1)  $\{a\}$  such that  $a \in \mathbb{A}$ . Further, let  $\mathbb{X} \subseteq \mathbb{R}^q$  denote the Euclidean space for the system state vector, then we can define the *Hybrid Bernoulli Random Set* (HBRS)  $\triangleq (\mathcal{A}, x)$ , as a new state variable which incorporates the input Bernoulli random set  $\mathcal{A}$  and the random state vector  $x$ , taking values in the hybrid space  $\mathcal{B}(\mathbb{A}) \times \mathbb{X}$ . A HBRS is fully specified by the probability  $r$  of  $\mathcal{A}$  being a singleton (i.e., of unknown input *existence*), the PDF  $p^0(x)$  defined on the state space  $\mathbb{X}$ , and the joint PDF  $p^1(a, x)$  defined on the joint space  $\mathbb{A} \times \mathbb{X}$ , i.e.

$$p(\mathcal{A}, x) = \begin{cases} (1 - r) p^0(x), & \text{if } \mathcal{A} = \emptyset \\ r \cdot p^1(a, x), & \text{if } \mathcal{A} = \{a\} \end{cases} \quad (11)$$

Moreover, since integration over  $\mathcal{B}(\mathbb{A}) \times \mathbb{X}$  takes the form

$$\int_{\mathcal{B}(\mathbb{A}) \times \mathbb{X}} p(\mathcal{A}, x) \delta \mathcal{A} dx = \int p(\emptyset, x) dx + \iint p(\{a\}, x) da dx \quad (12)$$

where the set integration with respect to  $\mathcal{A}$  is defined according to (9) while the integration with respect to  $x$  is an ordinary one, it is easy to see that  $p(\mathcal{A}, x)$  integrates to one by substituting (11) in (12), and noting that  $p^0(x)$  and  $p^1(a, x)$  are conventional probability density functions on  $\mathbb{X}$  and  $\mathbb{A} \times \mathbb{X}$ , respectively. This, in turn, guarantees that (11) is a FISST probability density for the HBRS  $(\mathcal{A}, x)$ , which will be referred to as *hybrid Bernoulli density* throughout the rest of the paper.

#### A. Measurement model and correction

Let us consider the *extra packet injection* model introduced in Section II-A, for which the measurement set defined in (3) is given by the union of two independent random sets. As it is clear from (4),  $\mathcal{Y}_k$  is a Bernoulli random set (with cardinality  $|\mathcal{Y}_k|$  at most 1) which depends on whether the system-originated measurement  $y_k$  is delivered or not. Conversely,

no prior knowledge on the cardinality distribution  $\rho(n)$  of the random set  $\mathcal{F}_k$  of false measurements is assumed. This means that  $\rho(n)$  is taken as an uninformative distribution and, hence, the FISST PDF of false-only measurements can be written as

$$\gamma(\mathcal{F}_k) \propto |\mathcal{F}_k|! \prod_{y_k \in \mathcal{F}_k} \kappa(y_k) \quad (13)$$

where  $\kappa(y_k)$  is a PDF describing the prior knowledge on the distribution of false measurements on the measurement space  $\mathbb{Y}$ . Clearly, if no prior knowledge on such a distribution can be assumed,  $\kappa(y_k)$  can be taken as an uninformative (i.e. uniform) PDF over  $\mathbb{Y}$  (see also the discussion in Section II-A). The following result holds.

*Lemma 1:* Let the cardinality of  $\mathcal{Z}_k$ , i.e. the number of received measurements, be equal to  $n$ . Then, the likelihood function  $\lambda(\mathcal{Z}_k | \mathcal{A}_k, x_k)$  can be written as

$$\lambda(\mathcal{Z}_k | \mathcal{A}_k, x_k) = \begin{cases} \gamma(\mathcal{Z}_k) \left[ 1 - p_d + \frac{p_d}{n} \sum_{y_k \in \mathcal{Z}_k} \frac{\ell(y_k | x_k)}{\kappa(y_k)} \right] & \text{if } \mathcal{A}_k = \emptyset \\ \gamma(\mathcal{Z}_k) \left[ 1 - p_d + \frac{p_d}{n} \sum_{y_k \in \mathcal{Z}_k} \frac{\ell(y_k | a_k, x_k)}{\kappa(y_k)} \right] & \text{if } \mathcal{A}_k = \{a_k\} \end{cases} \quad (14)$$

*Proof:* Let us first introduce the following FISST PDFs for  $\mathcal{A}_k = \emptyset$  and, respectively,  $\mathcal{A}_k = \{a_k\}$ :

$$\eta(\mathcal{Y}_k | \emptyset, x_k) = \begin{cases} 1 - p_d, & \text{if } \mathcal{Y}_k = \emptyset \\ p_d \ell(y_k | x_k), & \text{if } \mathcal{Y}_k = \{y_k\} \end{cases} \quad (15)$$

$$\eta(\mathcal{Y}_k | \{a_k\}, x_k) = \begin{cases} 1 - p_d, & \text{if } \mathcal{Y}_k = \emptyset \\ p_d \ell(y_k | a_k, x_k), & \text{if } \mathcal{Y}_k = \{y_k\} \end{cases} \quad (16)$$

Then, using the convolution formula [27, p. 385], one has

$$\lambda(\mathcal{Z}_k | \mathcal{A}_k, x_k) = \sum_{\mathcal{Y}_k \subseteq \mathcal{Z}_k} \eta(\mathcal{Y}_k | \mathcal{A}_k, x_k) \gamma(\mathcal{Z}_k \setminus \mathcal{Y}_k). \quad (17)$$

Hence, the likelihood corresponding to  $\mathcal{A}_k = \emptyset$  is given by

$$\lambda(\mathcal{Z}_k | \emptyset, x_k) = \eta(\emptyset | \emptyset, x_k) \gamma(\mathcal{Z}_k) + \sum_{y_k \in \mathcal{Z}_k} \eta(\{y_k\} | \emptyset, x_k) \gamma(\mathcal{Z}_k \setminus \{y_k\}). \quad (18)$$

Observe now that, in view of (13), we have

$$\gamma(\mathcal{Z}_k \setminus \{y_k\}) = \frac{\gamma(\mathcal{Z}_k)}{n \kappa(y_k)} \quad (19)$$

and, hence,  $\lambda(\mathcal{Z}_k | \emptyset, x_k)$  can be rewritten as in (14). Similarly, for  $\mathcal{A}_k = \{a_k\}$  we have

$$\lambda(\mathcal{Z}_k | \{a_k\}, x_k) = \eta(\emptyset | \{a_k\}, x_k) \gamma(\mathcal{Z}_k) + \sum_{y_k \in \mathcal{Z}_k} \eta(\{y_k\} | \{a_k\}, x_k) \gamma(\mathcal{Z}_k \setminus \{y_k\}) \quad (20)$$

which can be rewritten as in (14) by exploiting again (19). Notice that the first term on the RHS of (20) accounts for the case of no system-originated measurement, i.e.  $\mathcal{F}_k = \mathcal{Z}_k$ , while the subsequent term in the summation considers the union of disjoint events that one observation of  $\mathcal{Z}_k$  is authentic and the rest are false measurements, i.e.  $\mathcal{F}_k = \mathcal{Z}_k \setminus \{y_k\}$  for any  $y_k \in \mathcal{Z}_k$ . ■

Using the measurement model of Lemma 1, exact correction equations of the Bayesian random set filter for joint input detection and state estimation with extra packet injections are obtained as follows.

*Theorem 1:* Suppose that the prior density at time  $k$  is *hybrid Bernoulli* of the form

$$p(\mathcal{A}_k, x_k | \mathcal{Z}^{k-1}) = \begin{cases} (1 - r_{k|k-1}) p_{k|k-1}^0(x_k), & \text{if } \mathcal{A}_k = \emptyset \\ r_{k|k-1} \cdot p_{k|k-1}^1(a_k, x_k), & \text{if } \mathcal{A}_k = \{a_k\} \end{cases} \quad (21)$$

Then, given the measurement random set  $\mathcal{Z}_k$  defined in (3), also the posterior density at time  $k$  turns out to be *hybrid Bernoulli* of the form

$$p(\mathcal{A}_k, x_k | \mathcal{Z}^k) = \begin{cases} (1 - r_{k|k}) p_{k|k}^0(x_k), & \text{if } \mathcal{A}_k = \emptyset \\ r_{k|k} \cdot p_{k|k}^1(a_k, x_k), & \text{if } \mathcal{A}_k = \{a_k\} \end{cases} \quad (22)$$

completely specified by the triplet

$$r_{k|k} = \frac{1 - p_d (1 - \frac{1}{n} \Gamma_1)}{1 - p_d [1 - \frac{1}{n} (\Gamma_0 - r_{k|k-1} \Gamma)]} r_{k|k-1} \quad (23)$$

$$p_{k|k}^0(x_k) = \frac{1 - p_d + \frac{p_d}{n} \sum_{y_k \in \mathcal{Z}_k} \frac{\ell(y_k | x_k)}{\kappa(y_k)}}{1 - p_d (1 - \frac{1}{n} \Gamma_0)} p_{k|k-1}^0(x_k) \quad (24)$$

$$p_{k|k}^1(a_k, x_k) = \frac{1 - p_d + \frac{p_d}{n} \sum_{y_k \in \mathcal{Z}_k} \frac{\ell(y_k | a_k, x_k)}{\kappa(y_k)}}{1 - p_d (1 - \frac{1}{n} \Gamma_1)} p_{k|k-1}^1(a_k, x_k) \quad (25)$$

where

$$\Gamma_0 \triangleq \sum_{y_k \in \mathcal{Z}_k} \frac{\int \ell(y_k | x_k) p_{k|k-1}^0(x_k) dx_k}{\kappa(y_k)} \quad (26)$$

$$\Gamma_1 \triangleq \sum_{y_k \in \mathcal{Z}_k} \frac{\int \int \ell(y_k | a_k, x_k) p_{k|k-1}^1(a_k, x_k) da_k dx_k}{\kappa(y_k)} \quad (27)$$

and  $\Gamma \triangleq \Gamma_0 - \Gamma_1$ .

*Proof:* The correction equation of the Bayes random set filter for joint input detection and state estimation follows from a generalization of (7), which yields

$$p(\mathcal{A}_k, x_k | \mathcal{Z}^k) = \frac{\lambda(\mathcal{Z}_k | \mathcal{A}_k, x_k) p(\mathcal{A}_k, x_k | \mathcal{Z}^{k-1})}{p(\mathcal{Z}_k | \mathcal{Z}^{k-1})} \quad (28)$$

where  $\lambda(\mathcal{Z}_k | \mathcal{A}_k, x_k)$  is given by (18) and (20), while

$$\begin{aligned} p(\mathcal{Z}_k | \mathcal{Z}^{k-1}) &= \iint \lambda(\mathcal{Z}_k | \mathcal{A}_k, x_k) p(\mathcal{A}_k, x_k | \mathcal{Z}^{k-1}) \delta \mathcal{A}_k dx_k \\ &= \int \lambda(\mathcal{Z}_k | \emptyset, x_k) p(\emptyset, x_k | \mathcal{Z}^{k-1}) dx_k \\ &\quad + \iint \lambda(\mathcal{Z}_k | \{a_k\}, x_k) p(\{a_k\}, x_k | \mathcal{Z}^{k-1}) da_k dx_k. \end{aligned} \quad (29)$$

By using (18)-(20) and (21), and simply noting that  $\int p_{k|k-1}^0(x_k) dx_k = 1$  and  $\iint p_{k|k-1}^1(a_k, x_k) da_k dx_k = 1$ , (29) leads to

$$p(\mathcal{Z}_k | \mathcal{Z}^{k-1}) = \gamma(\mathcal{Z}_k) \left[ 1 - p_d + \frac{p_d}{n} (\Gamma_0 - r_{k|k-1} \Gamma) \right]. \quad (30)$$

For the case  $\mathcal{Z}_k = \emptyset$ , the above reduces to

$$p(\emptyset|\mathcal{Z}^{k-1}) = 1 - p_d \quad (31)$$

The posterior probability of unknown input existence  $r_{k|k}$  can be obtained from the posterior density (28) with  $\mathcal{A}_k = \emptyset$  via

$$r_{k|k} = 1 - \int p(\emptyset, x_k|\mathcal{Z}^k) dx_k \quad (32)$$

where - using (18), (21) and (31) in (28) - we have

$$p(\emptyset, x_k|\mathcal{Z}^k) = (1 - r_{k|k-1}) p_{k|k-1}^0(x_k). \quad (33)$$

Moreover,  $p_{k|k}^0(x_k) = p(\emptyset, x_k|\mathcal{Z}^k)/(1 - r_{k|k})$ , and the joint density for the system under off-nominal behavior can be easily derived from the posterior density with  $\mathcal{A}_k = \{a_k\}$  by recalling that  $p_{k|k}^1(a_k, x_k) = p(\{a_k\}, x_k|\mathcal{Z}^k)/r_{k|k}$ , where

$$p(\{a_k\}, x_k|\mathcal{Z}^k) = r_{k|k-1} \cdot p_{k|k-1}^1(a_k, x_k) \quad (34)$$

results from replacing (20), (21) and (31) in (28). Notice that from the set integral definition (9), and densities (33)-(34), it holds that  $\int p(\emptyset, x_k|\mathcal{Z}^k) dx_k + \iint p(\{a_k\}, x_k|\mathcal{Z}^k) da_k dx_k = 1$ . Hence, as stated, the Bayes correction (22) provides a hybrid Bernoulli density.

Next, for the case  $\mathcal{Z}_k = \{y_k\}$ , by substituting (18), (21) and (30) in (28), one gets  $p(\emptyset, x_k|\mathcal{Z}^k)$  which, in turn, is used to obtain (23) through (32). Once  $r_{k|k}$  is known, (24) and (25) easily follow as shown above for the case  $\mathcal{Z}_k = \emptyset$ . ■

From Theorem 1, it is evident that if  $p_d = 1$  and  $r_{k|k-1} = 1$ , then  $r_{k|k} = 1$  follows from (23). Moreover, if we further assume that no false measurements are collected at time  $k$ , i.e.  $\mathcal{Z}_k = \{y_k\}$ , then (25) simplifies to the standard Bayes filter correction of the JISE problem (7). In an analogous way, if  $r_{k|k-1} = 0$ , first we obtain  $r_{k|k} = 0$ , then, from (24), the standard Bayes filter correction for an input-free system:

$$p_{k|k}^0(x_k) = \frac{\ell(y_k|x_k) p_{k|k-1}^0(x_k)}{\int \ell(y_k|x_k) p_{k|k-1}^0(x_k) dx_k}. \quad (35)$$

## B. Dynamic model and prediction

Let us next introduce the dynamic model of the HBRIS  $(\mathcal{A}, x)$  essential to derive the prediction equations. To this end, it is reasonable to assume that the joint transitional density of  $(\mathcal{A}, x)$  at time  $k+1$  takes the form

$$\pi(\mathcal{A}_{k+1}, x_{k+1}|\mathcal{A}_k, x_k) = \pi(x_{k+1}|\mathcal{A}_k, x_k) \pi(\mathcal{A}_{k+1}|\mathcal{A}_k) \quad (36)$$

which ensues from considering the unknown input as independent of the system state, as supposed in Section II-B. Such an assumption is motivated by the fact that  $a_{k+1}$  may assume all possible values, being completely unknown (we consider the most general model for exogenous signals where any value can be injected, e.g., via the compromised actuators/sensors). Clearly, in accordance with (1), we have

$$\pi(x_{k+1}|\mathcal{A}_k, x_k) = \begin{cases} \pi(x_{k+1}|x_k), & \text{if } \mathcal{A}_k = \emptyset \\ \pi(x_{k+1}|a_k, x_k), & \text{if } \mathcal{A}_k = \{a_k\} \end{cases} \quad (37)$$

where  $\pi(x_{k+1}|x_k)$  and  $\pi(x_{k+1}|a_k, x_k)$  are known Markov transition PDFs.

Concerning instead the transitional density  $\pi(\mathcal{A}_{k+1}|\mathcal{A}_k)$ , it is reasonable to assume that the presence of an unknown input at time  $k+1$  is more probable when it is already present at time  $k$ . Accordingly, we can assume that: in the case of a system under normal operation at time  $k$ , an unknown input  $a_{k+1}$  will enter into action during the sampling interval with probability  $p_b$ ; if instead the system is already under off-nominal behavior (i.e.,  $\mathcal{A}_k$  is a singleton), it is supposed that the exogenous action will endure from time step  $k$  to time step  $k+1$  with probability  $p_s$ . Notice that the probabilities  $p_b$  and  $p_s$  can be seen as design parameters for the filter that can be tuned depending on the desired properties. For instance, the lower is  $p_b$  the more cautious will be the filter in declaring the presence of an unknown input. Similarly, the higher is  $p_s$  the more cautious will be the filter in declaring that the effect of the unknown input has disappeared. On the other hand, in accordance with the input model of Section II-B, we assume that the knowledge of  $a_k$  adds no information on  $a_{k+1}$ . Summing up, the dynamics of  $\mathcal{A}_k$  resulting from the aforesaid assumptions can be modeled as a Bernoulli Markov process described by the following densities:

$$\begin{aligned} \pi(\mathcal{A}_{k+1}|\emptyset) &= \begin{cases} 1 - p_b, & \text{if } \mathcal{A}_{k+1} = \emptyset \\ p_b p(a_{k+1}), & \text{if } \mathcal{A}_{k+1} = \{a_{k+1}\} \end{cases} \\ \pi(\mathcal{A}_{k+1}|\{a_k\}) &= \begin{cases} 1 - p_s, & \text{if } \mathcal{A}_{k+1} = \emptyset \\ p_s p(a_{k+1}), & \text{if } \mathcal{A}_{k+1} = \{a_{k+1}\} \end{cases} \end{aligned}$$

where  $p(a_{k+1})$  is a PDF representing the prior knowledge on the unknown input  $a_{k+1}$ . Clearly, when the external input is completely unknown, an uninformative PDF (e.g., uniform over the input space  $\mathbb{A}$ ) can be adopted for  $p(a_{k+1})$ , as usually done in the literature on unknown input estimation. As discussed in Section II-B, in the Bayesian framework this choice leads to a maximum-likelihood estimation of the unknown input value.

Under the above assumptions, an exact recursion for the prior density can be obtained, as stated in the following theorem.

*Theorem 2:* Given the posterior hybrid Bernoulli density  $p(\mathcal{A}_k, x_k|\mathcal{Z}^k)$  at time  $k$  of the form (22), fully characterized by the triplet  $(r_{k|k}, p_{k|k}^0(x_k), p_{k|k}^1(a_k, x_k))$ , also the predicted density turns out to be hybrid Bernoulli of the form

$$\begin{aligned} p(\mathcal{A}_{k+1}, x_{k+1}|\mathcal{Z}^k) & \quad (38) \\ &= \begin{cases} (1 - r_{k+1|k}) p_{k+1|k}^0(x_{k+1}), & \text{if } \mathcal{A}_{k+1} = \emptyset \\ r_{k+1|k} \cdot p_{k+1|k}^1(a_{k+1}, x_{k+1}), & \text{if } \mathcal{A}_{k+1} = \{a_{k+1}\} \end{cases} \end{aligned}$$

with

$$\begin{aligned} r_{k+1|k} &= (1 - r_{k|k}) p_b + r_{k|k} p_s \quad (39) \\ p_{k+1|k}^0(x_{k+1}) &= \frac{(1 - r_{k|k})(1 - p_b) p_{k+1|k}(x_{k+1}|\emptyset)}{1 - r_{k+1|k}} \end{aligned}$$

$$+ \frac{r_{k|k}(1 - p_s) p_{k+1|k}(x_{k+1}|\{a_k\})}{1 - r_{k+1|k}} \quad (40)$$

$$\begin{aligned} p_{k+1|k}^1(a_{k+1}, x_{k+1}) &= \frac{(1 - r_{k|k}) p_b p_{k+1|k}(x_{k+1}|\emptyset) p(a_{k+1})}{r_{k+1|k}} \\ &+ \frac{r_{k|k} p_s p_{k+1|k}(x_{k+1}|\{a_k\}) p(a_{k+1})}{r_{k+1|k}} \quad (41) \end{aligned}$$

where

$$p_{k+1|k}(x_{k+1}|\emptyset) = \int \pi(x_{k+1}|x_k) p_{k|k}^0(x_k) dx_k \quad (42)$$

$$p_{k+1|k}(x_{k+1}|\{a_k\}) = \iint \pi(x_{k+1}|a_k, x_k) p_{k|k}^1(a_k, x_k) da_k dx_k. \quad (43)$$

*Proof:* The prediction equation of the Bayes random set filter is given by the following generalization of (6)

$$\begin{aligned} p(\mathcal{A}_{k+1}, x_{k+1}|\mathcal{Z}^k) & \quad (44) \\ &= \iint \pi(\mathcal{A}_{k+1}, x_{k+1}|\mathcal{A}_k, x_k) p(\mathcal{A}_k, x_k|\mathcal{Z}^k) \delta \mathcal{A}_k dx_k \\ &= (1 - r_{k|k}) \int \pi(\mathcal{A}_{k+1}, x_{k+1}|\emptyset, x_k) p_{k|k}^0(x_k) dx_k \\ &+ r_{k|k} \iint \pi(\mathcal{A}_{k+1}, x_{k+1}|\{a_k\}, x_k) p_{k|k}^1(a_k, x_k) \delta \mathcal{A}_k dx_k \end{aligned}$$

where the set integral definition (9) and (22) have been used. Then, for  $\mathcal{A}_{k+1} = \emptyset$ , from (36), (37), and (38), one has

$$\begin{aligned} p(\emptyset, x_{k+1}|\mathcal{Z}^k) &= (1 - r_{k|k})(1 - p_b) \int \pi(x_{k+1}|x_k) p_{k|k}^0(x_k) dx_k \\ &+ r_{k|k}(1 - p_s) \iint \pi(x_{k+1}|a_k, x_k) p_{k|k}^1(a_k, x_k) da_k dx_k. \quad (45) \end{aligned}$$

Next, using (42) and (43), (45) becomes

$$\begin{aligned} p(\emptyset, x_{k+1}|\mathcal{Z}^k) &= (1 - r_{k|k})(1 - p_b) p_{k+1|k}(x_{k+1}|\emptyset) \\ &+ r_{k|k}(1 - p_s) p_{k+1|k}(x_{k+1}|\{a_k\}). \quad (46) \end{aligned}$$

Analogously, for  $\mathcal{A}_{k+1} = \{a_{k+1}\}$  we obtain

$$\begin{aligned} p(\{a_{k+1}\}, x_{k+1}|\mathcal{Z}^k) &= \left[ (1 - r_{k|k}) p_b p_{k+1|k}(x_{k+1}|\emptyset) \right. \\ &\left. + r_{k|k} p_s p_{k+1|k}(x_{k+1}|\{a_k\}) \right] p(a_{k+1}). \quad (47) \end{aligned}$$

Thus, the output of the prediction step given by (45)-(47) is of the form (38) under the settings (39)-(43). ■

It is clear from (39) that the system is predicted to be under off-nominal behavior at time  $k+1$  if either an existing unknown input persists from time  $k$ , or a novel external signal  $a_{k+1}$  starts affecting its dynamics. Similar to the standard Bernoulli filter, the prediction step of the proposed filter involves two separate terms, here accounting for the unknown input birth and input-survival. Notice that, if  $p_b = 0$ ,  $p_s = 1$  and  $r_{k|k} = 1$ , the prediction step (38) yields (6), which is the standard Chapman–Kolmogorov equation for the system under attack, since from (39)-(41) it follows that  $r_{k+1|k} = 1$ ,  $p_{k+1|k}^0(x_{k+1}) = 0$ , and  $p_{k+1|k}^1(a_{k+1}, x_{k+1}) = p_{k+1|k}(x_{k+1}|\{a_k\}) p(a_{k+1})$ .

*Remark 1:* As it happens in most nonlinear filtering problems, no exact closed-form solution to the proposed hybrid Bernoulli filter is admitted. However, for the special class of linear Gaussian models, this problem can be effectively mitigated by parameterizing the posterior densities  $p_{k|k}^0(\cdot)$  and  $p_{k|k}^1(\cdot, \cdot)$  via Gaussian mixtures and by limiting the growing number of components via simple pruning and merging procedures (see [33]) to allow for on-line computation. Detailed formulas of the Gaussian-mixture implementation of the hybrid Bernoulli filter (GM-HBF) can be found in [34].

*Remark 2:* Given the conditional density  $p(\mathcal{A}_k, x_k|\mathcal{Z}^k)$ , characterized by the triplet  $(r_{k|k}, p_{k|k}^0(\cdot), p_{k|k}^1(\cdot, \cdot))$ , the joint input detection and state estimation problem can be solved as follows. First of all, we perform unknown input detection using  $r_{k|k}$  from the available current hybrid Bernoulli density  $p(\mathcal{A}_k, x_k|\mathcal{Z}^k)$ . By using a MAP decision rule, given  $\mathcal{Z}_k$ , the detector will assign  $\hat{\mathcal{A}}_k \neq \emptyset$  (the system is under off-nominal behavior) if and only if  $\text{Prob}(\mathcal{A}_k \neq \emptyset|\mathcal{Z}^k) > \text{Prob}(\mathcal{A}_k = \emptyset|\mathcal{Z}^k)$ , i.e. if and only if  $r_{k|k} > 1/2$ . Then, if the unknown input has been detected, one can maximize  $p(\mathcal{A}_k, x_k|\mathcal{Z}^k)$  with respect to  $x_k$  and  $a_k$ . In this way it is possible to obtain a MAP estimate of  $x_k$  and a ML estimate of the unknown input  $a_k$ .

*Remark 3:* In the paper, the probability  $p_d$  of receiving the authentic measurements from sensors is assumed known based on the thought that this information might be available as an inherent property of the communication channel which can, for instance, be characterized from available historical trends. In the negative, the filter can be modified so as to estimate such a probability by incorporating an unknown variable, representing the probability  $p_d$ , into the augmented state variable (see [35] for a discussion in the context of multi-Bernoulli filtering).

#### IV. NUMERICAL EXAMPLE

The effectiveness of the Gaussian-mixture hybrid Bernoulli filter has been tested by simulating a *load altering attack* [36] on the IEEE 14-bus system (Fig. 1) consisting of 5 synchronous generators, 11 load buses, with parameters taken from [37]. The system dynamics is described by the linearized swing equation, obtained after Kron reduction [38], modeling the evolution of each phase angle  $\delta_i$  and angular frequency  $\omega_i$ ,  $i = 1, \dots, 5$  at generator buses ( $q = 10$  states). After discretization (with sampling interval  $T = 0.01s$ ), the model of the system takes the form (1)-(2), where the whole state is measured by a network  $\mathcal{S}$  of sensors. The DC state estimation model assumes 1 p.u. (per unit) voltage magnitudes in all buses and  $j1$  p.u. branch impedance. The system is assumed to be corrupted by additive zero mean Gaussian white process and measurement noises with variances  $\sigma_w^2 = 0.01$  and  $\sigma_v^2 = 0.01$ . At time  $k = 50$  an attack input  $a = [0.2, 0.1]^T$  p.u. is injected into the system to abruptly increase the real power demand of the two victim load buses 3 and 9 with an additional loading of 21.23% and 33.9%, respectively. This type of attack can provoke a loss of synchrony of the rotor angles, and hence a deviation of the rotor speeds of all generators from the nominal value. We also fixed the following parameters:  $p_b = 0.05$ ,  $p_s = 0.95$ ,  $p_d = 0.95$ , pruning and merging thresholds  $\gamma_p = 10^{-2}$  and  $\gamma_m = 3$  for the Gaussian-mixture implementation. The guessed distribution of the unknown input was taken as normally distributed. To get an uninformative prior, we set  $p(a_{k+1}) \sim \mathcal{N}(\hat{a}, P^a)$ , with  $\hat{a} = [0, 0]^T$  and  $P^a = 10^5 \mathbf{I}_2$ . The distribution of false measurements  $\kappa(\cdot)$  is modeled as uninformative (i.e. uniform) over the bounded interval  $[-5, 5]$  by the filter, while the actual extra packets are generated from a uniform distribution over  $[-2, 2.5]$  for the rotor angles and over  $[-0.6, 1]$  for the frequencies. Fig. 2 shows the true and estimated probability of attack existence (a) and the Root

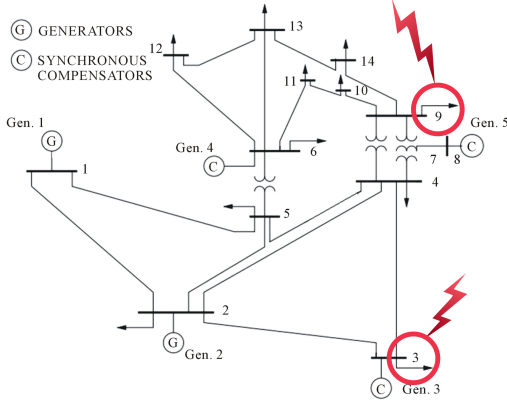


Fig. 1: Single-line model of the IEEE 14-bus system. The true victim load buses 3 and 9 are circled in red.

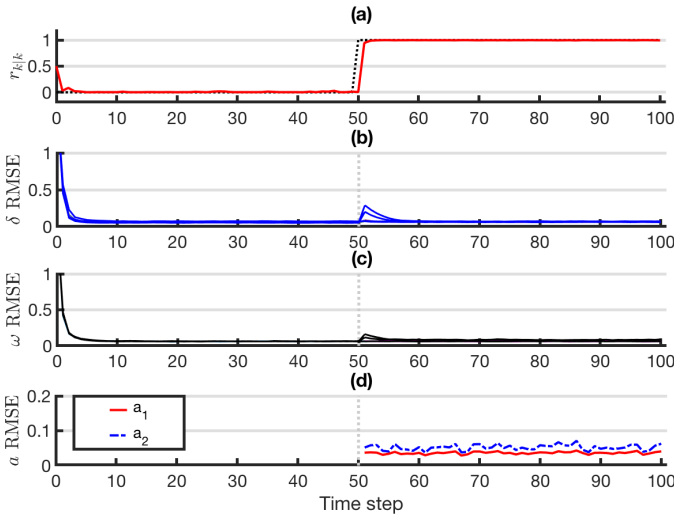


Fig. 2: Performance of the GM-HBF in terms of unknown input detection (a), estimation of rotor angles  $\delta_i$ ,  $i = 1, \dots, 5$  (b), frequencies  $\omega_i$ ,  $i = 1, \dots, 5$  (c), and attack signal (d).

Mean Square Error (RMSE), averaged over 1000 Monte Carlo runs, relative to the rotor angle (b) and frequency (c) estimates. Fig. 2 (d) shows the RMSE of the estimated components of the attack input, extracted from  $p_{k|k}^1(a, x)$ . As shown in the results (a)-(d), the proposed filter succeeds in promptly detecting the exogenous input altering the nominal power system behavior, and in being simultaneously resilient to attack signals on power demand, as well as robust to extra false packets and undelivered measurements. Fig. 3 (a) provides a comparison between the true and the estimated values of the two rotor angles mainly affected by the victim load buses, and clearly shows how  $\delta_1$  and  $\delta_3$  lose synchrony once the load altering attack enters into action. Nevertheless, the proposed estimator keeps tracking the state evolution with high accuracy even after time  $k = 50$ , once recognized that the system is under attack. Finally, Fig. 3 (b) shows the performance of the GM-HBF in estimating the generator frequencies  $\omega_1$  and  $\omega_3$ , before and after the appearance of the unknown input.

Next, we further tested the robustness of the proposed filter to a possible mismatch between the assumed and the actual

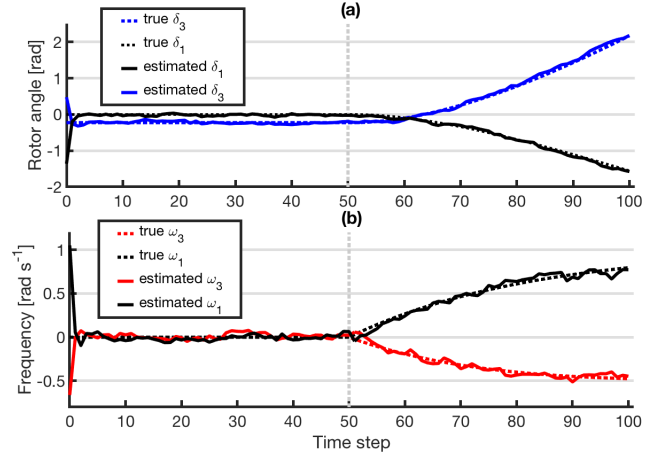


Fig. 3: (a) Estimated vs. true trajectory of rotor angles  $\delta_j$ ,  $j = 1, 3$ . Note that, if  $|\delta_j|$  is sufficiently large (i.e., values close to  $\pi/2$ ), the linear small signal approximation significantly deviates from the nonlinear dynamics of the system, and hence the assumed dynamic model becomes inaccurate. (b) Estimated vs. true trajectory of frequencies  $\omega_1$  and  $\omega_3$ .

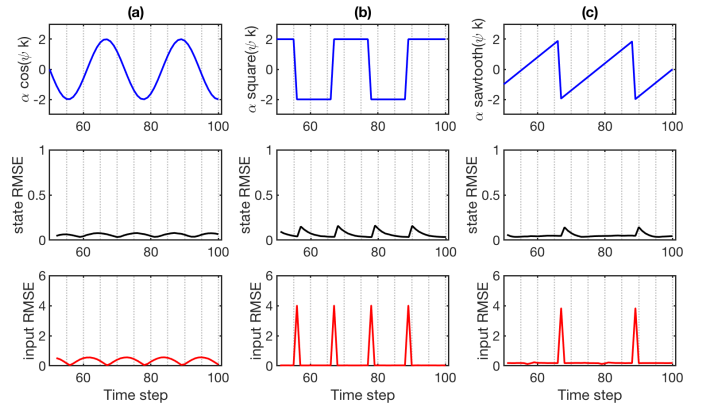


Fig. 4: Performance in terms of state (frequencies) and input reconstruction (after  $k = 50$ ) under distribution mismatch:  $a_k$  generated as a cosine (a), square (b), and sawtooth (c) wave.

distributions of the unknown input and of the extra packets. In particular, we evaluated the performance of the GM-HBF estimator via Monte Carlo simulations using different shapes of time-varying functions for the unknown signal (see [39] for a similar attack model on power grids). We considered three different scenarios with  $a_k$  on the victim load bus 9 generated by cos, square, and sawtooth functions in MATLAB, with amplitude  $\alpha = 2$  and angular frequency  $\psi = 3$ , respectively. In contrast, the guessed distribution of the unknown input was taken as normally distributed in all scenarios. To get an uninformative prior, we set  $p(a_{k+1}) \sim \mathcal{N}(\hat{a}, P^a)$ , with  $\hat{a} = 0$  and  $P^a = 10^5$ . For the extra packet injections, recall that the cardinality distribution is modeled as an uninformative distribution in (13), while the actual number of false measurements follows a Poisson distribution with average number  $\xi = 10$ . As shown in Fig. 4, the hybrid Bernoulli filter is found to be robust to possible distribution mismatches, since it can successfully track the unknown signal, and hence guarantee solid performance in terms of state estimation even under

different mismatch conditions involving both the unknown input and the false measurements. The results about input detection and rotor angles estimation are omitted since they are similar to the ones shown in Figs. 2 (a)-(b).

## V. CONCLUSIONS

This note proposed a general Bayesian framework to solve state estimation for (linear/nonlinear) systems in the face of switching unknown inputs and extra packet injections. Random finite sets have been exploited in order to model the switching nature of the exogenous signals as well as the possible presence of false measurements. A Bayes-optimal hybrid Bernoulli filter has been proposed for jointly detecting the unknown inputs and estimating the system state. The promising results exhibited in a realistic power system application, even under different distribution mismatch conditions, motivate future work on hybrid Bernoulli filtering, such as worst-case performance analysis and the extension to distributed settings.

## REFERENCES

- [1] S. Gillijns and B. D. Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough," *Automatica*, vol. 43, no. 5, pp. 934–937, 2007.
- [2] S. Gillijns and B. D. Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, no. 1, pp. 111–116, 2007.
- [3] Y. Cheng, H. Ye, Y. Wang, and D. Zhou, "Unbiased minimum-variance state estimation for linear systems with unknown input," *Automatica*, vol. 45, no. 2, pp. 485–491, 2009.
- [4] H. Fang, R. A. De Callafon, and J. Cortés, "Simultaneous input and state estimation for nonlinear systems with applications to flow field estimation," *Automatica*, vol. 49, no. 9, pp. 2805–2812, 2013.
- [5] S. Yong, M. Zhu, and E. Frazzoli, "Simultaneous input and state estimation with a delay," in *Proc. 54th IEEE Conference on Decision and Control*, pp. 468–475, Osaka, Japan, 2015.
- [6] S. Yong, M. Zhu, and E. Frazzoli, "A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems," *Automatica*, vol. 63, no. 1, pp. 321–329, 2016.
- [7] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.
- [8] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Allerton Conference on Communication, Control, and Computing*, pp. 911–918, 2009.
- [9] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Conference on Decision and Control*, pp. 1854–1859, 2013.
- [10] C. D. Persis and P. Tesi, "Input-to-state stabilizing control under denial-of-service," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2930–2944, 2015.
- [11] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal denial-of-service attack scheduling with energy constraint," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3023–3028, 2015.
- [12] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.
- [13] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *Proc. 54th IEEE Conference on Decision and Control*, pp. 5820–5826, Osaka, Japan, 2015.
- [14] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, 2015.
- [15] 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.
- [16] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2017.
- [17] Y. Shoukry, A. Puggelli, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving," in *Proc. American Control Conference*, pp. 3818–3823, Chicago, Illinois, USA, 2015.
- [18] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 49–59, 2017.
- [19] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proc. American Control Conference*, pp. 2439–2444, 2015.
- [20] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, no. 1, pp. 135–148, 2015.
- [21] S. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *Proc. 54th IEEE Conference on Decision and Control*, pp. 5162–5169, Osaka, Japan, 2015.
- [22] D. Shi, R. J. Elliott, and T. Chen, "On finite-state stochastic modeling and secure estimation of cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 65–80, 2017.
- [23] Q. Gu, P. Liu, S. Zhu, and C.-H. Chu, "Defending against packet injection attacks in unreliable ad hoc networks," in *Proc. IEEE Global Telecommunications Conference*, vol. 3, pp. 1837–1841, St. Louis, Missouri, USA, 2005.
- [24] X. Zhang, H. Chan, A. Jain, and A. Perrig, "Bounding packet dropping and injection attacks in sensor networks," Tech. Rep. 07-019, CMU-CyLab, Pittsburgh, PA, USA, 2007. [Online]. Available: [https://www.cylab.cmu.edu/files/pdfs/tech\\_reports/cmucylab07019.pdf](https://www.cylab.cmu.edu/files/pdfs/tech_reports/cmucylab07019.pdf).
- [25] Y. Ho and R. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, 1964.
- [26] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on Bernoulli filters: theory, implementation and applications," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3406–3430, 2013.
- [27] R. P. S. Mahler, *Statistical multisource multitarget information fusion*. Norwood, MA, USA: Artech House, Inc., 2007.
- [28] B.-T. Vo, D. Clark, B.-N. Vo, and B. Ristic, "Bernoulli forward-backward smoothing for joint target detection and tracking," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4473–4477, 2011.
- [29] B.-T. Vo, C. M. See, N. Ma, and W. T. Ng, "Multi-sensor joint detection and tracking with the Bernoulli filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1385–1402, 2012.
- [30] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "A Bayesian approach to joint attack detection and resilient state estimation," in *Proc. 55th IEEE Conference on Decision and Control*, pp. 1192–1198, Las Vegas, Nevada, USA, 2016.
- [31] A. Yeredor, "The joint MAP-ML criterion and its relation to ML and to Extended Least-Squares," *IEEE Transaction on Signal Processing*, vol. 48, no. 12, pp. 3484–3492, 2000.
- [32] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters – Part II: estimation methods," *IEEE Transaction on Signal Processing*, vol. 63, no. 24, pp. 6647–6658, 2015.
- [33] B.-N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.
- [34] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "Joint attack detection and secure state estimation of cyber-physical systems," Tech. Rep., 2016 [Online]. Available: <http://arxiv.org/abs/1612.08478>.
- [35] B.-T. Vo, B.-N. Vo, R. Hoseinnezhad, and R. P. S. Mahler, "Robust multi-bernoulli filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 399–409, 2013.
- [36] S. Amini, H. Mohsenian-Rad, and F. Pasqualetti, "Dynamic load altering attacks in smart grid," in *Proc. Innovative Smart Grid Technologies Conference*, pp. 1–5, 2015.
- [37] R. Zimmerman, C. Murillo-Sanchez, and R. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.
- [38] F. Pasqualetti, A. Bicchi, and F. Bullo, "A graph-theoretical characterization of power network vulnerabilities," in *Proc. American Control Conference*, pp. 3918–3923, 2011.
- [39] A. F. Taha, J. Qi, J. Wang, and J. H. Panchal, "Risk mitigation for dynamic state estimation against cyber attacks and unknown inputs," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 886–899, 2018.