Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 50-1 (2017) 4979–4986

Secure State Estimation of Cyber-Physical Systems Under Switching Attacks

N. Forti * **G. Battistelli** * **L. Chisci** * **B. Sinopoli** **

* *Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Firenze, Firenze, Italy (e-mail: nicola.forti,giorgio.battistelli,luigi.chisci@unifi.it)*

** *Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA (e-mail: brunos@ece.cmu.edu)*

Abstract: This paper deals with secure state estimation against switching mode and signal attacks on cyber-physical systems, possibly affected by adversarial extra fake measurement injection. A stochastic Bayesian approach is undertaken by exploiting Bernoulli and Poisson random sets for modeling the attack existence and, respectively, fake measurements, as well as multiple models for handling the various attack modes. A Gaussian mixture implementation of the resulting random set Bayesian filter is presented and a simulation case-study concerning an electrical power network is worked out in order to demonstrate the effectiveness of the proposed approach.

Keywords: Cyber-physical systems; secure state estimation; switching attack; Bayesian state estimation; multiple-model filtering; Bernoulli filter.

1. INTRODUCTION

Cyber-Physical Systems (CPSs) are complex systems integrating computation, networking and physical processes. Notable examples of CPSs include next-generation systems in electric power grids, transportation and mobility, building and environmental monitoring/control, healthcare, and industrial process control. While advances in CPS technology will provide enhanced autonomy, efficiency, seamless interoperability and cooperation, the tighter interaction between cyber and physical realms is unavoidably introducing novel security vulnerabilities, which make CPSs subject to non-standard malicious threats. Recent real-world attacks, such as the Maroochy Shire sewage spill, the Stuxnet worm targeting an industrial control system, and the lately reported massive power outage against Ukrainian electric grid, have brought the attention of the engineering community towards the urgency of designing secure CPSs. There exists a wide range of cyber-attacks and a variety of approaches to handle them, i.e. to detect the attack outbreak as well as to correctly estimate the system state even in presence of the attack. Pasqualetti et al. (2013) addressed the problem of attack detection/identification, and proposed attack monitors for deterministic control systems. Active detection methods have been designed in order to reveal stealthy attacks via manipulation of e.g. control inputs in Mo et al. (2015) and dynamics in Weerakkody and Sinopoli (2015). Recent works of Mo and Sinopoli (2015), Yong et al. (2015) have focused on the problem of secure state estimation, i.e. of reconstructing the state even when the CPS of interest is under attack. Under the assumption of linear systems subject to an unknown, but bounded, number of *false-data injection* attacks, the problem for a noise-free system has been cast into an ℓ_0-optimization problem, and relaxed into a more efficient convex problem in Fawzi et al. (2014), and later adapted to systems with bounded noise in Pajic et al. (2015). The combinatorial complexity of this problem is tackled by Shoukry et al. (2015). Recently, Teixeira et al. (2015) modeled the most popular types of attack, based on the notions of adversary's resources and system knowledge. This paper specifically focuses on *switching mode attacks* by which the cyber-attacker can switch the currently operating mode of the CPS within a finite set of possible attack modes. Farraj et al. (2016) described how this can be achieved by altering the network's topology in a power system through breaker control signals, while Amin et al. (2010) studied the same type of attack on water distribution systems, where the water outflow can be influenced via boundary control actions. In particular, the paper aims to address the problem of jointly detecting a signal attack and estimating both the attack mode and system state from the available observations. The overall problem is formulated in a stochastic random set Bayesian framework by exploiting Bernoulli modeling for the signal attack presence/absence and multiple models to account for the different attack modes. It is worth to highlight that the adopted approach exhibits the following positive features: 1) can deal with nonlinear systems; 2) takes into account the presence of disturbances and noise; 3) can encompass in a unique framework different types of attacks (switching signal and mode attacks, extra packet injection, packet substitution, etc.); 4) provides (discrete or continuous) probability distributions of the attack existence, attack mode, attack signal and system state which are very useful for taking decisions.

The rest of the paper is organized as follows. Section 2 introduces the considered attack models and provides the

10.1016/j.ifacol.2017.08.765

necessary background. Section 3 formulates and solves the joint *attack detection and mode-state estimation* problem of interest in the Bayesian framework. Section 4 discusses the Gaussian-mixture implementation of the joint attack detector and mode-state estimator derived in Section 3. Then, Section 5 demonstrates the effectiveness of the proposed approach via a simulation example concerning a power network. Finally, Section 6 ends the paper with concluding remarks and perspectives for future work.

2. PROBLEM SETUP AND PRELIMINARIES

2.1 System and Attack Model

Let the discrete-time cyber-physical system under switching attacks be modeled by

$$x_{k+1} = \begin{cases} f_k^0(\nu_k, x_k) + w_k, & \text{under no signal attack} \\ f_k^1(\nu_k, x_k, a_k) + w_k, & \text{under signal attack} \end{cases} \tag{1}$$

where: k is the time index; $\nu_k \in \mathcal{M} = \{1, 2, ..., \mathfrak{m}\}$ the mode in operation at time k; $x_k \in \mathbb{R}^n$ is the state vector to be estimated; $a_k \in \mathbb{R}^m$, called attack vector, is an unknown input affecting the system only when it is under attack; $f_k^0(\nu_k, \cdot)$ and $f_k^1(\nu_k, \cdot, \cdot)$ are known mode-matched state transition functions that describe the system evolution under a specific mode ν_k, in the *no signal attack* and, respectively, *signal attack* cases; w_k is a random process disturbance, with probability density function (PDF) $p_w(\nu_k, \cdot)$, also affecting the system. For monitoring purposes, the state of the above system is observed through the measurement model

$$y_k = \begin{cases} h_k^0(\nu_k, x_k) + v_k, & \text{under no signal attack} \\ h_k^1(\nu_k, x_k, a_k) + v_k, & \text{under signal attack} \end{cases} \tag{2}$$

where: $h_k^0(\nu_k, \cdot)$ and $h_k^1(\nu_k, \cdot, \cdot)$ are known mode-matched measurement functions that refer to the *no signal attack* and, respectively, *signal attack* cases; v_k is a random measurement noise with PDF $p_v(\nu_k, \cdot)$. 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.).

Jump Markov models (1)-(2) allow to describe cyber-physical systems subject to two different types of switching attacks, as considered in Yong et al. (2015): (i) switching mode attacks, and (ii) switching signal attacks. The former class of attacks is capable of switching the ongoing mode of the system between a finite set of possible models $\mathcal{M}$, by e.g. altering the state transition of the system (in Farraj et al. (2016) the topology of a power network). Moreover, a change in the system mode might represent a modification of the set of corruptible actuators/sensors, i.e. a change of the structure under which the signal attack enters the system. In other words, switching mode attacks model every possible cyber-physical adversary's action causing a change of the functions f^0, f^1 governing the system dynamics and/or of the functions h^0, h^1 describing the observation process. On the other hand, a signal attack (ii), modeled in (1)-(2) via the attack vector a_k, is a time-varying signal of arbitrary magnitude and location injected into the system to corrupt sensor/actuator data (also known as *false-data injection attack*), here modeled as an unknown input. Specifically, as in Fang et al. (2013) for unknown inputs, a_k is treated as a white stochastic process $\{a_k\}$, independent of $x_0, \{w_k\}$ and $\{v_k\}$. This means that a_k and a_l are independent random variables for $k \neq l$, and a_k is independent of x_k and y^{k-1}. Such an assumption accounts for the fact that a_k may assume all possible values, being completely unknown (we consider the most general model for signal attacks where any value can be injected via the compromised actuators/sensors), and the knowledge of a_k adds no information on a_l, if $k \neq l$. At each time instant k, the signal attack can be present or not, according to the binary hypothesis 1 or 0, respectively, in (1)-(2).

Besides the above switching attacks (i) and (ii), the proposed attack model takes into account the presence of malicious *extra packet injections*, already addressed in Gu et al. (2005), Zhang et al. (2007), and Forti et al. (2016). This means that, in addition to the system-originated measurement y_k in (2), it is assumed that the system monitor might receive from some cyber-attacker one or multiple extra fake measurements indistinguishable (e.g. with same time stamp and sender id) from the system-originated one. For the subsequent developments, it is convenient to introduce the *attack set* at time k, $\mathcal{A}_k$, which is either equal to the empty set if the system is not under signal attack at time k or to the singleton $\{a_k\}$ otherwise, i.e.

$$\mathcal{A}_k = \begin{cases} \emptyset, & \text{if the system is not under signal attack} \\ \{a_k\}, & \text{otherwise.} \end{cases}$$

Due to the possible presence of the *extra packet injection* attack, it is also convenient to define the *measurement set* at time k

$$\mathcal{Z}_k = \mathcal{Y}_k \cup \mathcal{F}_k \tag{3}$$

where

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

is the set of system-originated measurements and $\mathcal{F}_k$ the finite set of fake measurements. It is worth mentioning that the above attack model could be extended to include the case of *packet substitution*, see Forti et al. (2016).

2.2 Multiple Model Approach

In order to handle switching attacks that can change the model in effect of the cyber-physical system (1)-(2), single-model approaches to secure state estimation, like the one proposed in Forti et al. (2016), need to be accomodated for switched systems. To this end, the idea is to rely on the Multiple Model (MM) approach [Bar-Shalom et al. (2004)]. For state estimation problems in jump Markov systems with known inputs, the MM framework provides, in theory, Bayes-optimal solutions by running in parallel a bank of $\mathfrak{m}$ mode-matched Bayesian filters. In simple terms, each filter, at each time instant, provides the mode-conditioned PDFs of the state given the observations, and recursively computes the modal probabilities for each mode ν_k. In this way, the MM approach can infer the *best* model of the current system's mode of operation as well as estimate the state of the system, based on the mode estimate. The MM approach commonly assumes that the true mode of the system switches according to

a (homogeneous) Markov chain with known transition probabilities

$$\pi_{ji} = \text{prob}\left(\nu_k = i \,|\, \nu_{k-1} = j\right), \quad i,j \in \mathcal{M} \tag{5}$$

where $\sum_{i=1}^{\mathtt{m}} \pi_{ji} = 1$. This assumption leads to the so called *Dynamic MM estimator*. A particular case of the aforementioned filter is the *Static MM estimator* which conversely assumes a constant mode variable $\nu_k \in \mathcal{M}$, i.e. $\nu_k = \nu$, and hence $\pi_{ji} = 0 \; \forall j, i \in \mathcal{M}$ with $j \neq i$. In the special case of joint mode, state, and attack estimation in cyber-physical systems of the form (1)-(2), the simultaneous presence of both the unknown mode and input affecting the system poses new challenges. In particular, the signal attack vector a_k can be considered as either a mode-dependent vector with possibly different dimension within distinct system modes, or a vector with fixed dimension for each possible mode (as assumed in Yong et al. (2015)). Furthermore, two possible approaches can be undertaken for solving the above problem, depending on the available knowledge of mode transitions. In this respect, although in adversarial environments it is usually more realistic to assume no prior knowledge of the mode transition model, and hence the *Static MM* approach provides the most suitable tool, there also exist cases where a *Dynamic MM* approach turns out to be preferable. A typical example of such a case is the problem of detecting and localizing an unknown malicious source (e.g. a biochemical attack) inducing and/or affecting the field to be monitored (a similar problem in a non-malicious setting has been presented by Battistelli et al. (2015)). The objective of secure estimation becomes the joint task of detecting the presence of the source, localizing it, estimating its intensity, and monitoring the induced field (for further information on dynamic field estimation, the interested reader is referred to Battistelli et al. (2016)). Since the unknown location of the source corresponds to a specific mode of the system, and the source intensity can be treated as an unknown signal attack, this problem can be recast as a joint mode, state, and attack estimation in jump Markov systems (1)-(2). Notice that in this case, the attack vector (intensity) has fixed dimension for each mode, and prior knowledge of the modal transitions can be assumed (since at each time instant they depend on the current location of the moving source, and hence modes corresponding to locations close to this position will clearly have higher probability to become the active mode of the system at the next step). This is the reason why in problems as the one described above, a *Dynamic MM* approach is preferable to undertake with respect to a static filtering. In this paper, under the assumption of a signal attack vector with fixed dimension $a_k \in \mathbb{R}^m$, we present a Bayesian recursion based on the *Dynamic MM* filtering, although in the numerical example a *Static MM* approach will be adopted. The aim of this paper is to address the problem of joint signal attack detection and simultaneous mode–state estimation, which amounts to jointly estimating, at each time k, the mode ν_k modeling switching mode attacks, the state x_k, and the signal attack set $\mathcal{A}_k$ given the set of measurements $\mathcal{Z}^k \triangleq \cup_{i=1}^k \mathcal{Z}_i$ up to time k.

2.3 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 Mahler (2007); 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* $p(n)$ that $\mathcal{X}$ have $n \geq 0$ elements and the joint PDFs $f\left(x_1, x_2, \ldots, x_n | n\right)$ over $\mathbb{X}^n$ given that $\mathcal{X}$ have n elements, are obtained as follows:

$$p(n) = \int_{\mathbb{X}^n} f(\{x_1, \ldots, x_n\}) \, dx_1 \cdots dx_n$$

$$f\left(x_1, x_2, \ldots, x_n | n\right) = \frac{1}{n!\, p(n)} \; f(\{x_1, \ldots, x_n\})$$

In order to measure probability over subsets of $\mathbb{X}$ or compute expectations of random set variables, Mahler (2007) introduced the notion of *set integral* for a generic real-valued function $g(\mathcal{X})$ of an RFS $\mathcal{X}$ as

$$\int g(\mathcal{X}) \, \delta\mathcal{X} = g(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int g(\{x_1, \ldots, x_n\}) \, dx_1 \cdots dx_n \tag{6}$$

Two specific types of RFSs, i.e. Bernoulli and Poisson RFSs, will be considered in this work.

Bernoulli RFS. A Bernoulli RFS is a random set which can be either empty or, with some probability $r \in [0, 1]$, a singleton $\{x\}$ distributed over $\mathbb{X}$ according to the PDF $p(x)$. Accordingly, its set density is defined as follows:

$$f(\mathcal{X}) = \begin{cases} 1 - r, & \text{if } \mathcal{X} = \emptyset \\ r \cdot p(x), & \text{if } \mathcal{X} = \{x\} \end{cases} \tag{7}$$

Poisson RFS. A Poisson RFS is a random finite set with Poisson-distributed cardinality, i.e.

$$p(n) = \frac{e^{-\xi} \xi^n}{n!}, \;\; n = 0, 1, 2, \ldots \tag{8}$$

and elements independently distributed over $\mathbb{X}$ according to a given spatial density $p(\cdot)$. Accordingly, its set density is defined as follows:

$$f(\mathcal{X}) = e^{-\xi} \prod_{x \in \mathcal{X}} \xi \, p(x). \tag{9}$$

3. SECURE ESTIMATION AGAINST SWITCHING ATTACKS

Let the signal attack 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 attack probability 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 $\mathbb{X} \subseteq \mathbb{R}^n$ denote the Euclidean space for the system state vector. In Forti et al. (2016) we defined the *Hybrid Bernoulli Random Set* (HBRS) $(\mathcal{A}, x)$ as a state variable which incorporates the Bernoulli attack random set $\mathcal{A}$ and the random state vector x. The HBRS is then augmented in order to include the hidden mode (or discrete state) in the new state variable $(\mathcal{A}, x, \nu)$, that we refer to as *Multiple Model Hybrid Bernoulli Random Set* (MM-HBRS), which takes values in the hybrid space $\mathcal{B}(\mathbb{A}) \times \mathbb{X} \times \mathcal{M}$. An MM-HBRS is fully specified by the (signal attack) probability r of $\mathcal{A}$ being a singleton, the mode-conditioned PDF

$p^0(x,\nu)$, and the mode-conditioned joint PDF $p^1(a,x,\nu)$, i.e.

$$p(\mathcal{A},x,\nu)=\begin{cases}(1-r)\,p^0(x,\nu), & \text{if } \mathcal{A}=\emptyset\\ r\cdot p^1(a,x,\nu), & \text{if } \mathcal{A}=\{a\}\end{cases} \tag{10}$$

with integration over the new state space

$$\sum_{i=1}^{\mathfrak{m}}\mu^i\int_{\mathcal{F}(\mathbb{B})\times\mathbb{X}} p(\mathcal{A},x|\nu_k=i)\,\delta\mathcal{A}\,dx \tag{11}$$

where

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

and $\mu^i \triangleq \text{prob}(\nu_k=i|\mathcal{Z})$ is the mode probability of mode i, given the measurement set $\mathcal{Z}$. The set integration with respect to $\mathcal{A}$ is defined according to (6) while the integration with respect to x is an ordinary one. Notice that in (11) $p(\mathcal{A},x,\nu)$ integrates to one, since integration with respect to $\mathcal{A}$ and x equals 1, $p^0(x)$ and $p^1(a,x)$ being conventional probability density functions, and $\sum_{i=1}^{\mathfrak{m}}\mu^i=1$. Thus, (10) turns out to be a FISST probability density for the MM-HBRS $(\mathcal{A},x,\nu)$, which will be referred to as *multiple model hybrid Bernoulli density* throughout the rest of the paper.

Similar to the single-model filter described in Forti et al. (2016), a MM-HBRS can be corrected and predicted in a recursive fashion so as to form a novel *Multiple Model Hybrid Bernoulli Filter* (MM-HBF). Note that proofs are omitted due to lack of space.

3.1 Measurement Model and Correction

The measurement model is the same described in (Forti et al., 2016, Section III.A) for the single-model HBF in the case of *extra packet injection*, with the only difference that, since (1)-(2) is now a jump Markov system, the measurement likelihood becomes dependent on the discrete-valued mode variable ν. Due to the possible presence of *extra packet injection*, whose attack model has been introduced in Section 2.1, 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, $\mathcal{F}_k$ is the random set of fake measurements that will be modeled hereafter as a Poisson random set, such that the number of counterfeit measurements is Poisson-distributed according to (8) and the FISST PDF of fake-only measurements $\gamma(\cdot)$ is given by (9) with spatial distribution $\kappa(\cdot)$ in place of $p(\cdot)$. The likelihood corresponding to $\mathcal{A}_k=\emptyset$ is given by

$$\begin{aligned}\lambda(\mathcal{Z}_k|\emptyset,x_k,\nu_k)&=\eta(\emptyset|\emptyset,x_k,\nu_k)\,\gamma(\mathcal{Z}_k)\\ &\quad+\sum_{y_k\in\mathcal{Z}_k}\eta(\{y_k\}|\emptyset,x_k,\nu_k)\,\gamma(\mathcal{Z}_k\setminus\{y_k\})\\ &=\gamma(\mathcal{Z}_k)\left[1-p_d+p_d\sum_{y_k\in\mathcal{Z}_k}\frac{\ell(y_k|x_k,\nu_k)}{\xi\,\kappa(y_k)}\right]\end{aligned}$$

while for $\mathcal{A}_k=\{a_k\}$ we have

$$\begin{aligned}\lambda(\mathcal{Z}_k|\{a_k\},x_k,\nu_k)&=\eta(\emptyset|\{a_k\},x_k,\nu_k)\,\gamma(\mathcal{Z}_k)\\ &\quad+\sum_{y_k\in\mathcal{Z}_k}\eta(\{y_k\}|\{a_k\},x_k,\nu_k)\,\gamma(\mathcal{Z}_k\setminus\{y_k\})\\ &=\gamma(\mathcal{Z}_k)\left[1-p_d+p_d\sum_{y_k\in\mathcal{Z}_k}\frac{\ell(y_k|a_k,x_k,\nu_k)}{\xi\,\kappa(y_k)}\right].\end{aligned} \tag{13}$$

Using the above measurement model, exact correction equations of the multiple-model Bayesian random set filter for joint signal attack detection, mode and state estimation in the case of extra packet injection attack are obtained as follows.

Note that from now on the notation $\langle\alpha,\beta\rangle=\int\alpha(x)\beta(x)\mathrm{d}x$ will be used for the inner product of two functions.

Theorem 1. Assume that the prior density at time k is multiple model hybrid Bernoulli of the form

$$p(\mathcal{A}_k,x_k,\nu_k|\mathcal{Z}^{k-1})=\begin{cases}(1-r_{k|k-1})p^0_{k|k-1}(x_k,\nu_k) & \text{if } \mathcal{A}_k=\emptyset\\ r_{k|k-1}\cdot p^1_{k|k-1}(a_k,x_k,\nu_k) & \text{if } \mathcal{A}_k=\{a_k\}\end{cases} \tag{14}$$

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

$$p(\mathcal{A}_k,x_k,\nu_k|\mathcal{Z}^{k})=\begin{cases}(1-r_{k|k})\,p^0_{k|k}(x_k,\nu_k) & \text{if } \mathcal{A}_k=\emptyset\\ r_{k|k}\cdot p^1_{k|k}(a_k,x_k,\nu_k) & \text{if } \mathcal{A}_k=\{a_k\}\end{cases} \tag{15}$$

with parameters

$$r_{k|k}=\frac{1-p_d\,(1-\Gamma_1)}{1-p_d(1-\Gamma_0+r_{k|k-1}\Gamma)}\,r_{k|k-1} \tag{16}$$

$$\begin{aligned}p^0_{k|k}(x_k,\nu_k)&=\frac{1-p_d\left[1-\sum_{y_k\in\mathcal{Z}_k}\frac{\ell(y_k|x_k,\nu_k)}{\xi\,\kappa(y_k)}\right]}{1-p_d\,(1-\Gamma_0)}\\ &\quad\times p^0_{k|k-1}(x_k,\nu_k)\end{aligned} \tag{17}$$

$$\begin{aligned}p^1_{k|k}(a_k,x_k,\nu_k)&=\frac{1-p_d\left[1-\sum_{y_k\in\mathcal{Z}_k}\frac{\ell(y_k|a_k,x_k,\nu_k)}{\xi\,\kappa(y_k)}\right]}{1-p_d\,(1-\Gamma_1)}\\ &\quad\times p^1_{k|k-1}(a_k,x_k,\nu_k)\end{aligned} \tag{18}$$

where

$$\Gamma_0\triangleq\sum_{y_k\in\mathcal{Z}_k}\frac{\left\langle\ell(y_k|x_k,\nu_k),p^0_{k|k-1}(x_k,\nu_k)\right\rangle}{\xi\,\kappa(y_k)} \tag{19}$$

$$\Gamma_1\triangleq\sum_{y_k\in\mathcal{Z}_k}\frac{\left\langle\ell(y_k|a_k,x_k,\nu_k),p^1_{k|k-1}(a_k,x_k,\nu_k)\right\rangle}{\xi\,\kappa(y_k)} \tag{20}$$

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

3.2 Dynamic Model and Prediction

The joint transitional density, described in (Forti et al., 2016, Section III.B) for the single-model HBF, is also affected by the jump Markov model, so that in the new MM framework it also depends on the specific modal state ν_k as follows

$$p(\mathcal{A}_{k+1}, x_{k+1}, \nu_{k+1}|\mathcal{A}_k, x_k, \nu_k) = p(x_{k+1}, \nu_{k+1}|\mathcal{A}_k, x_k, \nu_k)\, p(\mathcal{A}_{k+1}|\mathcal{A}_k) \quad (21)$$

where

$$p(x_{k+1}, \nu_{k+1}|\mathcal{A}_k, x_k, \nu_k) = \begin{cases} p(\nu_{k+1}|\nu_k)\, p(x_{k+1}|x_k, \nu_k) & \text{if } \mathcal{A}_k = \emptyset \\ p(\nu_{k+1}|\nu_k)\, p(x_{k+1}|a_k, x_k, \nu_k) & \text{if } \mathcal{A}_k = \{a_k\} \end{cases} \quad (22)$$

Note that the dynamics of the Markov process $\mathcal{A}_k$ in (21) is Bernoulli, i.e.

$$p(\mathcal{A}_{k+1}|\emptyset) = \begin{cases} 1 - p_b & \text{if } \mathcal{A}_{k+1} = \emptyset \\ p_b\, p(a_{k+1}) & \text{if } \mathcal{A}_{k+1} = \{a_{k+1}\} \end{cases} \quad (23)$$

$$p(\mathcal{A}_{k+1}|\{a_k\}) = \begin{cases} 1 - p_s & \text{if } \mathcal{A}_{k+1} = \emptyset \\ p_s\, p(a_{k+1}) & \text{if } \mathcal{A}_{k+1} = \{a_{k+1}\} \end{cases} \quad (24)$$

where $p(a_{k+1})$ is the PDF of the attack input vector. Clearly, when the attack vector is completely unknown, a non-informative PDF (e.g., uniform in the attack space) can be used as $p(a_{k+1})$. 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 multiple model hybrid Bernoulli density $p(\mathcal{A}_k, x_k, \nu_k|\mathcal{Z}^k)$ at time k of the form (15), fully characterized by the parameter triplet $\big(r_{k|k}, p^0_{k|k}(x_k, \nu_k), p^1_{k|k}(a_k, x_k, \nu_k)\big)$, also the predicted density turns out to be multiple model hybrid Bernoulli of the form

$$p(\mathcal{A}_{k+1}, x_{k+1}, \nu_{k+1}|\mathcal{Z}^k) = \begin{cases} (1 - r_{k+1|k})\, p^0_{k+1|k}(x_{k+1}, \nu_{k+1}) & \text{if } \mathcal{A}_{k+1} = \emptyset \\ r_{k+1|k} \cdot p^1_{k+1|k}(a_{k+1}, x_{k+1}, \nu_{k+1}) & \text{if } \mathcal{A}_{k+1} = \{a_{k+1}\} \end{cases} \quad (25)$$

with parameters

$$r_{k+1|k} = (1 - r_{k|k})\, p_b + r_{k|k}\, p_s \quad (26)$$

$$p^0_{k+1|k}(x_{k+1}, \nu_{k+1}) = \frac{(1 - r_{k|k})(1 - p_b)\, p_{k+1|k}(x_{k+1}|\emptyset)}{1 - r_{k+1|k}} + \frac{r_{k|k}(1 - p_s)\, p_{k+1|k}(x_{k+1}|\{a_k\})}{1 - r_{k+1|k}} \quad (27)$$

$$p^1_{k+1|k}(a_{k+1}, x_{k+1}, \nu_{k+1}) = \frac{(1 - r_{k|k})\, p_b\, p_{k+1|k}(x_{k+1}|\emptyset)\, p(a_{k+1})}{r_{k+1|k}} + \frac{r_{k|k}\, p_s\, p_{k+1|k}(x_{k+1}|\{a_k\})\, p(a_{k+1})}{r_{k+1|k}} \quad (28)$$

where

$$p_{k+1|k}(x_{k+1}|\emptyset) = \Big\langle p(x_{k+1}|x_k, \nu_k), p^0_{k|k}(x_k, \nu_k) \Big\rangle \quad (29)$$

$$p_{k+1|k}(x_{k+1}|\{a_k\}) = \Big\langle p(x_{k+1}|a_k, x_k, \nu_k), p^1_{k|k}(a_k, x_k, \nu_k) \Big\rangle . \quad (30)$$

□

4. GAUSSIAN-MIXTURE IMPLEMENTATION

Although no closed-form solution to the Bayes optimal recursion is admitted in general, for the special class of linear Gaussian models it is possible to analytically propagate in time the posterior densities $p^0_{k|k}(\cdot)$ and $p^1_{k|k}(\cdot)$ in the form of Gaussian mixtures (weights, means and covariances), and the probability of a signal attack. Note that in the case of nonlinear models and/or non-Gaussian noises, the solution can be obtained via nonlinear extensions of the GM approximation (e.g. Unscented/Extended GM) or sequential Monte Carlo methods (i.e. particle filter).

Denoting by $\mathcal{N}(x; m, P)$ a Gaussian PDF in the variable x, with mean m and covariance P, the closed-form solution assumes linear Gaussian observation and transition models conditioned on the modal state, i.e. for each mode $i \in \mathcal{M}$ one has

$$\ell(y_k|x_k, \nu_k = i) = \mathcal{N}(y; C^i x_k, R^i) \quad (31)$$

$$\ell(y_k|a_k, x_k, \nu_k = i) = \mathcal{N}(y; C^i x_k + H^i a_k, R^i) \quad (32)$$

$$p(x_{k+1}|x_k, \nu_k = i) = \mathcal{N}(x; A^i x_k, Q^i) \quad (33)$$

$$p(x_{k+1}|a_k, x_k, \nu_k = i) = \mathcal{N}(x; A^i x_k + G^i a_k, Q^i) \quad (34)$$

In addition, the signal attack model can be expressed as a Gaussian mixture of the form

$$p(a) = \sum_{j=1}^{J^a} \tilde{\omega}^{a,j} \mathcal{N}(a; \tilde{a}^j, \tilde{P}^{a,j}) \quad (35)$$

and the probabilities of signal attack survival p_s and measurement delivery p_d are assumed independent of both the system state and mode. In the GM implementation, each probability density at time k conditioned on mode $\nu_k = i$ is represented by the following set of parameters

$$\Big(r_{k|k}, p^{0,i}_{k|k}(x_k), p^{1,i}_{k|k}(a_k, x_k)\Big) = \Big(r_{k|k}, \big\{\omega^{0,ij}_{k|k}, m^{0,ij}_{k|k}, P^{0,ij}_{k|k}\big\}_{j=1}^{J^0_{k|k}}, \big\{\omega^{1,ij}_{k|k}, m^{1,ij}_{k|k}, P^{1,ij}_{k|k}\big\}_{j=1}^{J^1_{k|k}}\Big),\ i \in \mathcal{M}$$

where symbols ω and J denote, respectively, weights and number of mixture components. In the above equation we defined $m^0_{k|k} = \hat{x}^0_{k|k}$, $m^1_{k|k} = [\hat{x}^{1\,T}_{k|k}, \hat{a}^T_k]^T$, $P^0_{k|k} \triangleq \mathbb{E}[(x_k - \hat{x}^0_{k|k})(x_k - \hat{x}^0_{k|k})^T]$, $P^1_{k|k} = \begin{bmatrix} P^{1x}_{k|k} & P^{xa}_k \\ P^{ax}_k & P^a_k \end{bmatrix}$, and $P^{1x}_{k|k} \triangleq \mathbb{E}[(x_k - \hat{x}^1_{k|k})(x_k - \hat{x}^1_{k|k})^T]$, $(P^{xa}_k)^T = P^{ax}_k \triangleq \mathbb{E}[(a_k - \hat{a}_k)(x_k - \hat{x}^1_{k|k})^T]$, $P^a_k \triangleq \mathbb{E}[(a_k - \hat{a}_k)(a_k - \hat{a}_k)^T]$. The weights are such that $\sum_{j=1}^{J^0_{k|k}} \omega^{0,j}_{k|k} = 1$, and $\sum_{j=1}^{J^1_{k|k}} \omega^{1,j}_{k|k} = 1$. The Gaussian Mixture implementation of the *Multiple Model Hybrid Bernoulli Filter* (GM-MM-HBF) is described as follows.

4.1 GM-MM-HBF correction

Proposition 3. Suppose assumptions (31)-(35) hold, the predicted FISST density at time k is fully specified by the triplet $\big(r_{k|k-1}, p^0_{k|k-1}(x_k, \nu_k), p^1_{k|k-1}(a_k, x_k, \nu_k)\big)$, and $p^0_{k|k-1}(\cdot)$, $p^1_{k|k-1}(\cdot)$ for each $i \in \mathcal{M}$ are Gaussian mixtures of the form

$$p^{0,i}_{k|k-1}(x_k) = \sum_{j=1}^{J^{0,i}_{k|k-1}} \omega^{0,ij}_{k|k-1} \mathcal{N}(m^{0,ij}_{k|k-1}, P^{0,ij}_{k|k-1}) \quad (36)$$

$$p^{1,i}_{k|k-1}(a_k, x_k) = \sum_{j=1}^{J^{1,i}_{k|k-1}} \omega^{1,ij}_{k|k-1} \mathcal{N}(m^{1,ij}_{k|k-1}, P^{1,ij}_{k|k-1}) \quad (37)$$

where $m^{0,ij}_{k|k-1} = \hat{x}^{0,ij}_{k|k-1}$, $m^{1,ij}_{k|k-1} = [(\hat{x}^{1,ij}_{k|k-1})^T, (\hat{a}^j_k)^T]^T$, $\sum_{j=1}^{J^{0,i}_{k|k-1}} \omega^{0,ij}_{k|k-1} = 1$, and $\sum_{j=1}^{J^{1,i}_{k|k-1}} \omega^{1,ij}_{k|k-1} = 1$. Then the posterior FISST density $\big(r_{k|k}, p^0_{k|k}(x_k,\nu_k), p^1_{k|k}(a_k,x_k,\nu_k)\big)$ for each mode i is given by

$$r_{k|k} = \frac{1-p_d+p_d\,\Gamma_1}{1-p_d+p_d(1-r_{k|k-1})\Gamma_0+p_d\,r_{k|k-1}\Gamma_1}\,r_{k|k-1} \tag{38}$$

$$\begin{aligned} p^{0,i}_{k|k}(a_k,x_k) &= \sum_{j=1}^{J^{0,i}_{k|k}} \omega^{0,ij}_{k|k}\mathcal{N}(m^{0,ij}_{k|k}, P^{0,ij}_{k|k}) \\ &= \sum_{j=1}^{J^{0,i}_{k|k-1}} \omega^{0,ij}_{\bar{D},k|k}\mathcal{N}(m^{0,ij}_{k|k-1}, P^{0,ij}_{k|k-1}) \\ &\quad + \sum_{y_k\in\mathcal{Z}_k}\sum_{j=1}^{J^{0,i}_{k|k-1}} \omega^{0,ij}_{D,k|k}\mathcal{N}(m^{0,ij}_{k|k}, P^{0,ij}_{k|k}) \end{aligned} \tag{39}$$

$$\begin{aligned} p^{1,i}_{k|k}(a_k,x_k) &= \sum_{j=1}^{J^{1,i}_{k|k}} \omega^{1,ij}_{k|k}\mathcal{N}(m^{1,ij}_{k|k}, P^{1,ij}_{k|k}) \\ &= \sum_{j=1}^{J^{1,i}_{k|k-1}} \omega^{1,ij}_{\bar{D},k|k}\mathcal{N}(m^{1,ij}_{k|k-1}, P^{1,ij}_{k|k-1}) \\ &\quad + \sum_{y_k\in\mathcal{Z}_k}\sum_{j=1}^{J^{1,i}_{k|k-1}} \omega^{1,ij}_{D,k|k}\mathcal{N}(m^{1,ij}_{k|k}, P^{1,ij}_{k|k}) \end{aligned} \tag{40}$$

where we denoted, for $b = 0, 1$:

$$\omega^{b,ij}_{\bar{D},k|k} = \frac{(1-p_d)\,\omega^{b,ij}_{k|k-1}}{\Delta_b}, \quad \omega^{b,ij}_{D,k|k} = \frac{p_d\,\omega^{b,ij}_{k|k-1} q^{b,ij}_k(y_k)}{\Delta_b\,\xi\,\kappa(y_k)}$$

$$\Delta_b = 1-p_d+p_d\sum_{y_k\in\mathcal{Z}_k}\sum_{h\in\mathcal{M}}\sum_{l=1}^{J^{1,h}_{k|k-1}} \frac{\omega^{b,hl}_{k|k-1}}{\xi\,\kappa(y_k)} q^{b,hl}_k(y_k)$$

and

$$\begin{aligned} q^{0,ij}_k(y_k) &= \mathcal{N}(y_k; C^i m^{0,ij}_{k|k-1}, C^i P^{0,ij}_{k|k-1} C^{i^T} + R^i) \\ q^{1,ij}_k(y_k) &= \mathcal{N}(y_k; \tilde{C}^i m^{1,ij}_{k|k-1}, \tilde{C}^i P^{1,ij}_{k|k-1} \tilde{C}^{i^T} + R^i) \end{aligned}$$

with $\tilde{C}^i \triangleq [C^i, H^i]$. □

4.2 GM-MM-HBF prediction

Proposition 4. Suppose assumptions (31)-(35) hold, the posterior FISST density at time k is fully specified by the triplet $\big(r_{k|k}, p^0_{k|k}(x_k,\nu_k), p^1_{k|k}(a_k,x_k,\nu_k)\big)$, and $p^0_{k|k}(\cdot)$, $p^1_{k|k}(\cdot)$ for each $i \in \mathcal{M}$ are Gaussian mixtures of the form

$$p^{0,i}_{k|k}(x_k) = \sum_{j=1}^{J^{0,i}_{k|k}} \omega^{0,ij}_{k|k}\mathcal{N}(m^{0,ij}_{k|k}, P^{0,ij}_{k|k}) \tag{41}$$

$$p^{1,i}_{k|k}(a_k,x_k) = \sum_{j=1}^{J^{1,i}_{k|k}} \omega^{1,ij}_{k|k}\mathcal{N}(m^{1,ij}_{k|k}, P^{1,ij}_{k|k}). \tag{42}$$

Then the predicted FISST density $\big(r_{k+1|k}, p^0_{k+1|k}(x_{k+1},\nu_{k+1}), p^1_{k+1|k}(a_{k+1},x_{k+1},\nu_{k+1})\big)$ for each mode i is given by

$$r_{k+1|k} = (1-r_{k|k})p_b + r_{k|k}p_s \tag{43}$$

$$p^{0,i}_{k+1|k}(x_{k+1}) = \sum_{j=1}^{J^{0,i}_{k+1|k}} \omega^{0,ij}_{k+1|k}\mathcal{N}(m^{0,ij}_{k+1|k}, P^{0,ij}_{k+1|k}) \tag{44}$$

$$p^{1,i}_{k+1|k}(a_{k+1},x_{k+1}) = \sum_{j=1}^{J^{1,i}_{k+1|k}} \omega^{1,ij}_{k+1|k}\mathcal{N}(m^{1,ij}_{k+1|k}, P^{1,ij}_{k+1|k}) \tag{45}$$

where (44) can be written as

$$\begin{aligned} p^{0,i}_{k+1|k}(x_{k+1}) &= \underbrace{\sum_{h\in\mathcal{M}}\sum_{j=1}^{J^{0,h}_{k|k}} \omega^{0,hj}_{\bar{B},k+1|k}\mathcal{N}(m^{0,hj}_{\bar{B},k+1|k}, P^{0,hj}_{\bar{B},k+1|k})}_{\text{no attack-birth}} \\ &\quad + \underbrace{\sum_{h\in\mathcal{M}}\sum_{j=1}^{J^{1,h}_{k|k}} \omega^{0,hj}_{\bar{S},k+1|k}\mathcal{N}(m^{0,hj}_{\bar{S},k+1|k}, P^{0,hj}_{\bar{S},k+1|k})}_{\text{no attack-survival}} \end{aligned}$$

with

$$\begin{aligned} m^{0,hj}_{\bar{B},k+1|k} &= A^h m^{0,hj}_{k|k} \\ P^{0,hj}_{\bar{B},k+1|k} &= A^h P^{0,hj}_{k|k} A^{h^T} + Q^h \\ \omega^{0,hj}_{\bar{B},k+1|k} &= \frac{(1-r_{k|k})(1-p_b)}{1-r_{k+1|k}}\,\pi_{hi}\,\omega^{0,hj}_{k|k} \\ m^{0,hj}_{\bar{S},k+1|k} &= \tilde{A}^h m^{1,hj}_{k|k} \\ P^{0,hj}_{\bar{S},k+1|k} &= \tilde{A}^h P^{1,hj}_{k|k} \tilde{A}^{h^T} + Q^h \\ \omega^{0,hj}_{\bar{S},k+1|k} &= \frac{r_{k|k}(1-p_s)}{1-r_{k+1|k}}\,\pi_{hi}\,\omega^{1,hj}_{k|k}. \end{aligned}$$

Moreover, (45) can be written as

$$\begin{aligned} &p^{1,i}_{k+1|k}(a_{k+1},x_{k+1}) \\ &= \underbrace{\sum_{h\in\mathcal{M}}\sum_{j=1}^{J^{0,h}_{k|k}}\sum_{l=1}^{J^a} \omega^{1,hjl}_{B,k+1|k}\mathcal{N}(m^{1,hjl}_{B,k+1|k}, P^{1,hjl}_{B,k+1|k})}_{\text{attack-birth}} \\ &\quad + \underbrace{\sum_{h\in\mathcal{M}}\sum_{j=1}^{J^{1,h}_{k|k}}\sum_{l=1}^{J^a} \omega^{1,hjl}_{S,k+1|k}\mathcal{N}(m^{1,hjl}_{S,k+1|k}, P^{1,hjl}_{S,k+1|k})}_{\text{attack-survival}} \end{aligned}$$

where

$$\begin{aligned} m^{1,hjl}_{B,k+1|k} &= \begin{bmatrix} A^h m^{0,hj}_{k|k} \\ \tilde{a}^l \end{bmatrix} \\ P^{1,hjl}_{B,k+1|k} &= \begin{bmatrix} A^h P^{0,hj}_{k|k} A^{h^T} + Q^h & 0 \\ 0 & \tilde{P}^{a,l} \end{bmatrix} \\ \omega^{1,hjl}_{B,k+1|k} &= \frac{(1-r_{k|k})\,p_b}{r_{k+1|k}}\,\pi_{hi}\,\omega^{0,hj}_{k|k}\,\tilde{\omega}^{a,l} \\ m^{1,hjl}_{S,k+1|k} &= \begin{bmatrix} \tilde{A}^h m^{1,hj}_{k|k} \\ \tilde{a}^l \end{bmatrix} \end{aligned}$$

$$P^{1,hjl}_{S,k+1|k} = \begin{bmatrix} \tilde{A}^h P^{1,hj}_{k|k} \tilde{A}^{h^T} + Q^h & 0 \\ 0 & \tilde{P}^{a,l} \end{bmatrix}$$
$$\omega^{1,hjl}_{S,k+1|k} = \frac{r_{k|k}\, p_s}{r_{k+1|k}}\, \pi_{hi}\, \omega^{1,hj}_{k|k}\, \tilde{\omega}^{a,l}.$$

□

5. NUMERICAL EXAMPLE: POWER SYSTEM

In this section, we demonstrate the effectiveness of the proposed Bayesian random-set approach for secure CPS state estimation in the presence of mode/signal switching attacks, extra packet injection attacks as well as uncertainty on measurement delivery. We consider the Western System Coordinating Council (WSCC) 9-bus test case in Fig. 1, consisting of 3 synchronous generators, 3 generator terminal buses, and 3 load buses. The transmission lines' parameters, the inertia and the damping coefficients of generators are taken from Sauer and Pai (1998). The dynamics of the system can be described by the linearized swing equation for the $n = 6$ active buses, derived through the Kron reduction by Pasqualetti et al. (2011) of the linear small-signal power network model. The n-dimensional state of the system comprises both the rotor angles and the frequencies of each generator. After discretization (with sampling interval $T = 0.01s$), the power system model takes the form (1)-(2), where each mode corresponds to one of the $\mathfrak{m} = 3$ different hypotheses on the set of vulnerable load buses $\mathcal{V}_1 = \{6, 8\}$, $\mathcal{V}_2 = \{5, 6\}$, and $\mathcal{V}_3 = \{5, 8\}$. At time $k = 50$ a signal attack vector $a_k = [0.05, 0.04]^T$ per-unit is injected into the system to abruptly increase the real power demand of the two victim load buses 6 and 8 with an additional loading of 5.56% and, respectively, 4%. This type of attack, referred to as *load altering attack* by Amini et al. (2015), 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. In this numerical study, the probabilities of attack-birth and attack-survival are fixed, respectively, at $p_b = 0.2$ and $p_s = 0.8$. A network of 6 sensors is deployed to measure the state of the system. The system-generated measurement vector is supposed to be delivered at the monitor/control center with probability $p_d = 0.98$. The extra fake measurements injected into the sensor channel are modeled as a Poisson RFS with average number $\xi = 40$ and probability density uniformly distributed over the interval $[-10, 0]$, suitably chosen to emulate system-originated observations. Fig. 2 shows the resulting number of fake measurements maliciously injected at each time step and the cases of undelivered system-originated measurement. For the joint task of signal attack detection and mode-state estimation, here we relied on the *Static* version (introduced in Section 2.2) of the GM-MM-HBF (described in Section 4). We notice from Fig. 3 that the proposed secure state estimation algorithm succeeds in detecting the switching mode attack, and hence in estimating the true system's mode of operation (characterized by the highest mode probability) $i = 1$, corresponding to a load altering attack on $\mathcal{V}_1$. Note that the posterior mode probabilities shown in Fig. 3 are determined as follows:

$$\bar{\mu}^i_{k|k} = \frac{\omega^{0,i}_{k|k}(1 - r^i_{k|k}) + \omega^{1,i}_{k|k}\, r^i_{k|k}}{\sum_{i=1}^{\mathfrak{m}} \omega^{0,i}_{k|k}(1 - r^i_{k|k}) + \omega^{1,i}_{k|k}\, r^i_{k|k}}, \quad i = 1, 2, 3.$$

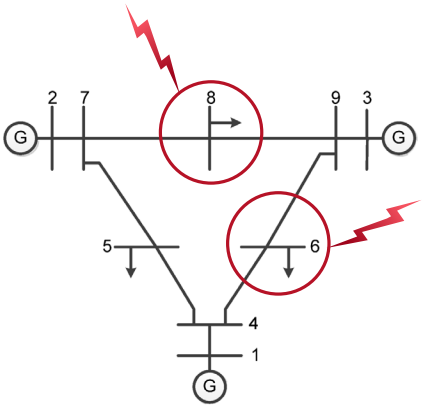

Fig. 1. Single-line model of the WSCC 9-bus system. The *true* victim load buses 6 and 8 are circled in red.

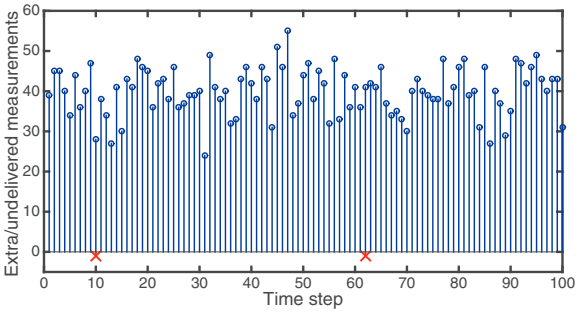

Fig. 2. Number of extra fake measurements injected (blue circles), and cases of undelivered system-originated measurement (red cross in −1) vs time. The proposed approach turns out to be particularly robust to *extra packet injections*.

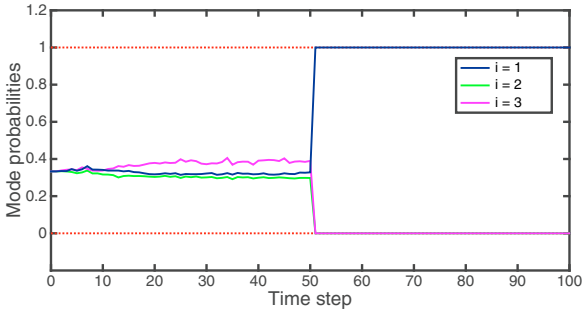

Fig. 3. Mode probabilities $\bar{\mu}^i_{k|k}$, $i = 1, 2, 3$. The three possible attack modes of the system share similar probabilities within the time interval $[0, 49]$ when there is no signal attack. The different behavior is revealed once a_k enters into action at time $k = 50$ and the unknown mode $i = 1$ is correctly estimated.

Moreover, the proposed filter promptly detects the unknown signal attack, as we can see from the attack probability $r^*_{k|k}$ in Fig. 4 which takes the unitary value after time $k = 50$. At each time instant k the estimated attack probability $r^*_{k|k} = r^{i^*}_{k|k}$ can be computed from the estimated mode $i^* = \arg\max \bar{\mu}^i_{k|k}$.

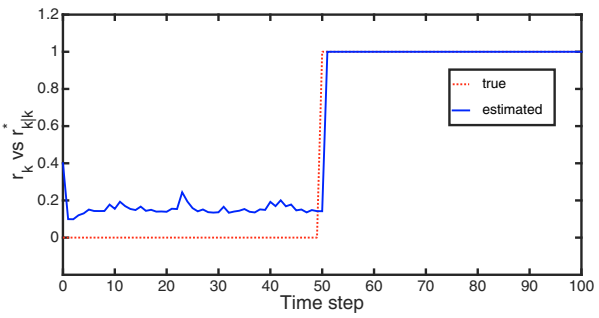

Fig. 4. *True* (r_k) and estimated ($r^*_{k|k}$) probability of existence of the signal attack a_k.

6. CONCLUSIONS

It has been shown how to securely estimate the state of a cyber-physical system in presence of attacks of various types by which the cyber-attacker can simultaneously switch an attack signal and the attack mode, and can also inject fake measurements. All these ingredients have been incorporated in a random set stochastic Bayesian filtering problem where Bernoulli and Poisson random sets have been used to model the attack signal switching and, respectively, measurement injection while multiple models have been exploited to account for different attack modes. A recursive Bayesian filter solving the formulated problem has been derived and its Gaussian mixture implementation has been developed and tested on a power network case study, exhibiting promising results in terms of prompt attack detection and resilient state estimation. Future work will concern application of the proposed technique to attack detection-system monitoring of power networks and to detection-localization of malicious sources.

ACKNOWLEDGEMENTS

This material is based upon work partly supported by the Department of Energy under Award Number DE-OE0000779.

REFERENCES

Amin, S., Litrico, X., Sastry, S., and Bayen, A. (2010). Stealthy deception attacks on water scada systems. In *Proc. of the 13th ACM Int. Conference on Hybrid Systems: Computation and Control (HSCC)*, 161–170.

Amini, S., Mohsenian-Rad, H., and Pasqualetti, F. (2015). Dynamic load altering attacks in smart grid. In *Innovative Smart Grid Technologies Conference (ISGT)*, 1–5.

Bar-Shalom, Y., Li, X., and Kirubarajan, T. (2004). *Estimation with applications to tracking and navigation: Theory algorithms and software*. John Wiley and Sons.

Battistelli, G., Chisci, L., Forti, N., Pelosi, G., and Selleri, S. (2015). Point source estimation via finite element multiple-model kalman filtering. In *Proc. of the IEEE 54th Conference on Decision and Control (CDC)*, 4984–4989.

Battistelli, G., Chisci, L., Forti, N., Pelosi, G., and Selleri, S. (2016). Distributed finite-element Kalman filter for field estimation. *IEEE Transactions on Automatic Control*.

Fang, H., Callafon, R.D., and Cortés, J. (2013). Simultaneous input and state estimation for nonlinear systems with applications to flow field estimation. *Automatica*, 49(9), 2805–2812.

Farraj, A., Hammad, E., Daoud, A., and Kundur, D. (2016). A game-theoretic analysis of cyber switching attacks and mitigation in smart grid systems. *IEEE Transactions on Smart Grid*, 7(4), 1846–1855.

Fawzi, H., Tabuada, P., and Diggavi, S. (2014). Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6), 1454–1467.

Forti, N., Battistelli, G., Chisci, L., and Sinopoli, B. (2016). A Bayesian approach to joint attack detection and resilient state estimation. In *Proc. of the IEEE 55th Conference on Decision and Control (CDC)*, 1192–1198.

Gu, Q., Liu, P., Zhu, S., and Chu, C.H. (2005). Defending against packet injection attacks in unreliable ad hoc networks. In *IEEE Global Telecommunications Conference (GLOBECOM)*, volume 3, 1837–1841.

Mahler, R. (2007). *Statistical multisource multitarget information fusion*. Artech House, Norwood, MA, USA.

Mo, Y. and Sinopoli, B. (2015). Secure estimation in the presence of integrity attacks. *IEEE Transactions on Automatic Control*, 60(4), 1145–1151.

Mo, Y., Weerakkody, S., and Sinopoli, B. (2015). Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Syst. Mag.*, 35(1), 93–109.

Pajic, M., Tabuada, P., Lee, I., and Pappas, G. (2015). Attack-resilient state estimation in the presence of noise. In *Proc. of the IEEE 54th Conference on Decision and Control (CDC)*, 5827–5832.

Pasqualetti, F., Bicchi, A., and Bullo, F. (2011). A graph-theoretical characterization of power network vulnerabilities. In *Proc. of the American Control Conference (ACC)*, 3918–3923.

Pasqualetti, F., Dörfler, F., and Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Trans. on Automatic Control*, 58(11), 2715–2729.

Sauer, P. and Pai, M. (1998). *Power System Dynamics and Stability*. Prentice Hall.

Shoukry, Y., Puggelli, A., Nuzzo, P., Sangiovanni-Vincentelli, A., Seshia, S., and Tabuada, P. (2015). Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving. In *Proc. of the American Control Conference (ACC)*, 3818–3823.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K. (2015). A secure control framework for resource-limited adversaries. *Automatica*, 51(1), 135–148.

Weerakkody, S. and Sinopoli, B. (2015). Detecting integrity attacks on control systems using a moving target approach. In *Proc. of the IEEE 54th Conference on Decision and Control (CDC)*, 5820–5826.

Yong, S., Zhu, M., and Frazzoli, E. (2015). Resilient state estimation against switching attacks on stochastic cyber-physical systems. In *Proc. of the IEEE 54th Conference on Decision and Control (CDC)*, 5162–5169.

Zhang, X., Chan, H., Jain, A., and Perrig, A. (2007). Bounding packet dropping and injection attacks in sensor networks. Tech. Rep. `https://www.cylab.cmu.edu/files/pdfs/tech_reports/cmucylab07019.pdf`.