This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

Distributed Joint Attack Detection and Secure State Estimation

Nicola Forti, Giorgio Battistelli, Luigi Chisci, *Senior Member, IEEE*, Suqi Li, Bailu Wang, and Bruno Sinopoli

***Abstract*—The joint task of detecting attacks and securely monitoring the state of a cyber-physical system is addressed over a cluster-based network wherein multiple fusion nodes collect data from sensors and cooperate in a neighborwise fashion in order to accomplish the task. The attack detection-state estimation problem is formulated in the context of random set theory by representing joint information on the attack presence/absence, on the system state and on the attack signal in terms of a *Hybrid Bernoulli Random Set* (HBRS) density. Then, combining previous results on HBRS recursive Bayesian filtering with novel results on Kullback-Leibler averaging of HBRSs, a novel distributed HBRS filter is developed and its effectiveness is tested on a case-study concerning wide-area monitoring of a power network.**

***Index Terms*—Cyber-physical systems; distributed detection and estimation; signal attack; Bayesian state estimation; Bernoulli filter.**

I. INTRODUCTION

CYBER-physical systems (CPSs) arise from the integration of computational and physical resources, interconnected via a communication network. Typical examples of CPSs include next-generation systems in electric power grids, transportation and mobility, building and environmental monitoring/control, health-care, and industrial process control. While on one hand, advances in CPS technology will enable growing autonomy, efficiency, seamless interoperability and cooperation, on the other hand the increased 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 sabotaging an industrial control system, and the lately reported massive power outage against Ukrainian electric grid [1], have brought into particularly sharp focus the urgency of designing secure CPSs. In presence of malicious threats against CPSs, standard approaches extensively used for systems subject to benign faults and failures need to be rethought. This is why recent advances on the design of secure systems have explored different paths, e.g., [2]–[10]. Recent work on attack detection and secure state estimation for CPSs [11], [12] has posed the problem in a Bayesian framework [13] and has exploited the powerful tools of random set theory [14], [15] for modelling various types of cyber-attacks, specifically: (i) *signal* attack, i.e. signal of arbitrary magnitude and location injected (with known structure) to corrupt sensor/actuator data, (ii) *packet substitution* attack, i.e., an intruder possibly intercepts and then replaces the system-generated measurement with a fake (unstructured) one, and (iii) *extra packet injection* [16], [17]) in which multiple counterfeit observations (junk packets) are possibly added to the system-generated measurement. In particular, the signal attack presence/absence is modelled by means of a Bernoulli random set (i.e., a set that can be either empty or a singleton depending on the presence or not of the attack) while the possible injection of fake measurements is modelled by a Bernoulli or Poisson random set for the *packet substitution* or, respectively, *extra packet injection* attack. Accordingly, joint attack detection-state estimation has been formulated as a recursive Bayesian filtering problem wherein the joint posterior density of the signal attack Bernoulli set and of the state vector, called *Hybrid Bernoulli Random Set* (HBRS) density in [11], is updated in time and whenever new data become available. The resulting *centralized* HBRS filter developed in [11] can timely detect signal attacks as well as reliably estimate the system state even in presence of the aforementioned (signal, packet substitution and extra packet injection) cyber-attacks provided that a fusion center receives all sensor data and stores/processes the aggregated data. Due to the geographically dispersed nature of CPSs, a distributed approach, wherein multiple fusion nodes communicate and cooperate to perform the joint attack detection and state estimation task, is by far preferable. However, devising distributed solutions becomes particularly challenging when the correlations between estimates from different fusion nodes are not known. The optimal solution to this problem was developed in [18], but the computational cost of calculating the common information can make the solution intractable in many real-world applications. A number of suboptimal solutions with demonstrated tractability have been formulated based on the Kullback-Leibler average (KLA) or generalized Covariance Intersection rule proposed by Mahler [19]. KLA is the generalization of Covariance Intersection [20] which only utilizes the mean and covariance and is limited to Gaussian posteriors. The KLA fusion rule relaxes the Gaussian constraint, and can be used to fuse multi-object distributions with completely unknown correlations, since it intrinsically

This work was supported in part by the U.S. Department of Energy under Award Number DE-OE0000779 and in part by Ente Cassa di Risparmio di Firenze under Grant 2015-0772.

N. Forti, G. Battistelli, and L. Chisci are with the Department of Information Engineering (DINFO), University of Florence, Via Santa Marta 3, 50139 Florence, Italy (e-mail: {nicola.forti,giorgio.battistelli,luigi.chisci}@unifi.it).

S. Li and B. Wang are with the School of Electronic Engineering, University of Electronic Science and Technology, Chengdu 611731, China, currently on leave at the DINFO, University of Florence (e-mail: qi_qi_zhu1210@163.com; w_b_l3020@163.com).

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

avoids any double counting of common information [21].

In this respect, a novel distributed HBRS filter is developed in the present paper to cope with a sensor network with cluster-based configuration. More precisely, the considered network consists of multiple fusion nodes (cluster heads or system monitors) each receiving measurements from multiple remote sensors via non-secure links and exchanging information with a subset of neighbors via secure links. The main contributions of this paper can be summarized as follows.

i) The attack detection-state estimation problem is formulated in the context of random set theory by representing the joint information on the attack presence/absence, on the system state and on the signal attack in terms of a HBRS density.
ii) We derive a closed-from solution for the KLA of HBRS densities. This novel result is provided as a key ingredient to derive the proposed distributed HBRS filter for joint attack detection and secure state estimation.
iii) We prove the immunity of the KLA fusion of HBRSs to double counting of information.
iv) We exploit consensus on the average [22], [23] to perform the collective KLA computation of HBRS densities over the whole network.
v) We test the proposed distributed HBRS filter for joint attack detection & secure state estimation on a benchmark wide-area monitoring system and we verify the efficiency of the Gaussian-mixture implementation of the proposed KLA fusion algorithm.

The rest of the paper is organized as follows. Section II provides motivations and the problem setup. Section III reviews HBRSs used to represent information about the CPS. Then Section IV deals with distributed fusion of HBRSs. Section V presents the novel distributed HBRS filter for joint attack detection and secure state estimation. Section VI provides a simulation case-study concerning wide-area monitoring of a smart grid to demonstrate the potentials of the proposed approach. Finally, Section VII ends the paper with concluding remarks and perspectives for future work.

II. PROBLEM SETUP

A. Motivating example: Wide-area monitoring systems

To motivate the problem of distributed secure state estimation of CPSs we consider a *wide-area monitoring system* (WAMS) [24], [25], i.e. a power system with multiple control areas consisting of local generators and loads, connected by inter-area tie-lines (see Fig. 1 for the example of the IEEE 14-bus wide-area monitoring system). WAMSs have recently attracted considerable attention due to the deregulation of modern power networks which have led to the introduction of numerous regional transmission organizations (RTOs) conceived to operate smaller portions of a large interconnected power network. This motivates the interest on decentralized strategies to monitor the electric grid over large geographical areas, each requiring the overall interconnection's state information to be available via a regional communication structure.

As illustrated in Fig. 1, WAMSs are partitioned power systems (on a geographical basis) where each non-overlapping

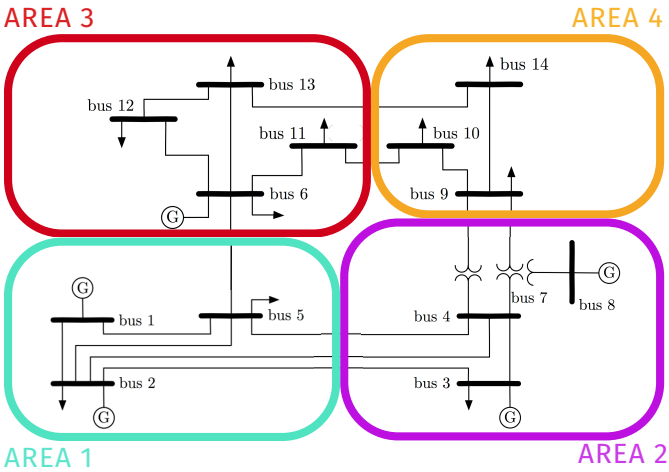


Fig. 1: IEEE 14-bus wide-area monitoring system (partitioned into four different areas).

area is assigned to a local fusion node, which only has access to its own local measurements and is dedicated to the reconstruction of the overall state by exchanging data with a small number of neighboring areas through wireless communication channels. The considered communication scheme is shown in Fig. 2. Due to both power and bandwidth constraints imposed by such communication systems, a centralized setup collecting multi-area measurements may not be practically feasible in wide-area monitoring operations. All these reasons call for a distributed approach in order to minimize the communication overhead and better manage coordination across geographically separated areas.

B. System description and attack model

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

$$x_{t+1} = \begin{cases} f_t^0(x_t) + w_t, & \text{under no attack} \\ f_t^1(x_t, a_t) + w_t, & \text{under attack} \end{cases} \quad (1)$$

where: t is the time index; $x_t \in \mathbb{R}^n$ is the state vector to be estimated; $a_t \in \mathbb{R}^m$, called attack vector, is an unknown input affecting the system only when it is under attack; $f_t^0(\cdot)$ and $f_t^1(\cdot,\cdot)$ are known state transition functions that describe the system evolution in the *no attack* and, respectively, *attack* cases; w_t is a random process disturbance also affecting the system, independent identically distributed (IID) according to the probability density function (PDF) $p_w(\cdot)$.

The attack modeled in (1)-(2) via the attack vector a_t is usually referred to as *signal* attack. While for ease of presentation only the case of a single attack model is taken into account, multiple attack models [10], [12] could be accommodated in the considered framework by letting (1)-(2) depend on a discrete variable, say ν_t, which specifies the particular attack model and has to be estimated together with a_t.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

For monitoring purposes, a sensor network with cluster-based configuration is taken into account. More specifically, it is supposed that the state of the above system is observed through a set of N_s remote sensors, each one characterized by a measurement equation of the form

$$y_t^i = \begin{cases} h_t^{0,i}(x_t) + v_t^i, & \text{under no attack} \\ h_t^{1,i}(x_t, a_t) + v_t^i, & \text{under attack} \end{cases} \quad (2)$$

for $i = 1, \ldots, N_s$, where $h_t^{0,i}(\cdot)$ and $h_t^{1,i}(\cdot,\cdot)$ are the known measurement functions of sensor i that refer to the *no attack* and, respectively, *attack* cases; the random measurement noises v_t^i, mutually independent as well as independent of the process disturbance w_t, are also IID with PDFs $p_v^i(\cdot)$.

Besides the N_s remote sensors, the network consists of a set $\mathcal{N} = \{1, \ldots, N_f\}$ of N_f fusion nodes (cluster heads or system monitors). Each fusion node $j \in \mathcal{N}$ receives the measurements y_t^i of a subset $\mathcal{S}_j$ of the sensor set $\mathcal{S} = \{1 \ldots, N_s\}$ and exchanges information with a subset $\mathcal{N}_j \subseteq \mathcal{N}$ of fusion nodes. The set $\mathcal{N}_j$ is called the set of in-neighbors of fusion node j. Hence, the set of fusion nodes define a (possibly directed) network (or graph) with node set $\mathcal{N}$ and link set $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ given by $\mathcal{L} = \{(\ell, j) : \ell \in \mathcal{N}_j\}$. For the reader's convenience, an example of a sensor network with cluster-based configuration is depicted in Fig. 2. Clearly, it is supposed that $\bigcup_{j\in\mathcal{N}} \mathcal{S}_j = \mathcal{S}$, i.e. each sensor sends its local measurements to at least one fusion node. On the other hand, in order to allow for redundancy in the communication topology, the sets $\mathcal{S}_j$, $j \in \mathcal{N}$, need not be mutually disjoint.

In accordance with the considered hierarchical topology, the network nodes are supposed to be characterized by different levels of security. More specifically, the fusion nodes are considered as trusted nodes, i.e. they cannot be compromised by adversarial attacks, and the communication between them is supposed to be secure, for instance because it is carried out through dedicated wired communication channels. On the contrary, the communication between the sensors and the fusion nodes is supposed to be non-secure. This scenario reflects the practical situation in which several low-cost remote sensors are deployed in the area of interest and data exchange occurs via a non-secure wireless channel, or when a malicious agent can take control of some of the remote sensors. Accordingly, it is assumed that the measurement y_t^i, $i \in \mathcal{S}_j$, is actually delivered to the fusion node $j \in \mathcal{N}$ with probability $p_d^{i,j} \in (0, 1]$, 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.). Further, besides the system-originated measurement y_t^i in (2), it is assumed that the fusion node might receive *fake* measurements from some cyber-attacker. In this respect, the following two cases will be considered.

1) *Packet substitution* - With some probability $p_f^{i,j} \in [0, 1)$, the attacker replaces the system-originated measurement y_t^i with a fake one $\tilde{y}_t^{i,j}$.
2) *Extra packet injection* - The attacker sends to the fusion node one or multiple fake measurements indistinguishable from the system-originated one.

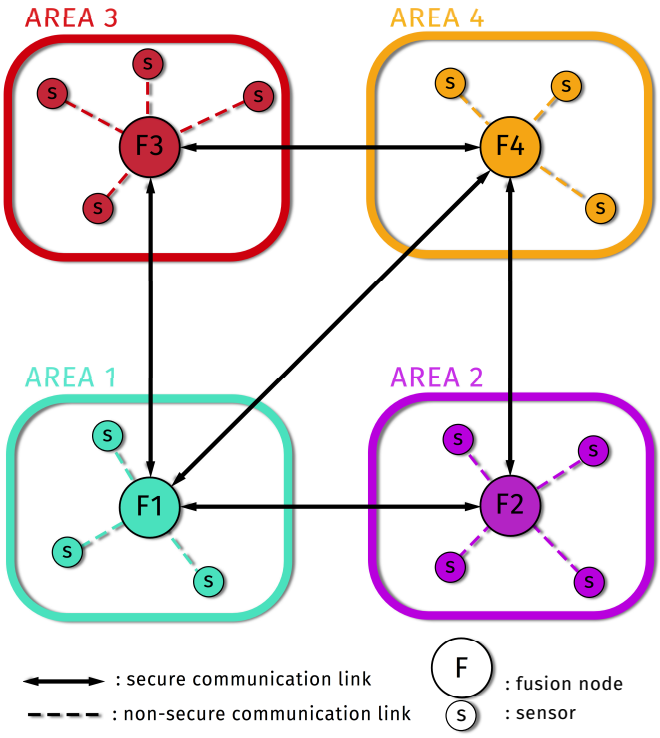


Fig. 2: Cluster-based configuration of a sensor network with local fusion nodes and sensors, secure and non-secure links.

For the subsequent developments, it is also convenient to define the set $\mathcal{Z}_t^{i,j}$ representing the set of measurements (either true or false) received by fusion node j from sensor i at time t. For the *packet substitution* attack:

$$\mathcal{Z}_t^{i,j} = \begin{cases} \emptyset, & \text{with probability } 1 - p_d^{i,j} \\ \{y_t^i\}, & \text{with probability } p_d^{i,j}(1 - p_f^{i,j}) \\ \{\tilde{y}_t^{i,j}\}, & \text{with probability } p_d^{i,j}\, p_f^{i,j} \end{cases} \quad (3)$$

where y_t^i is given by (2) and $\tilde{y}_t^{i,j}$ is a fake measurement provided by the attacker in place of y_t^i. Conversely, for the *extra packet injection* attack the definition (3) is replaced by

$$\mathcal{Z}_t^{i,j} = \mathcal{Y}_t^{i,j} \cup \mathcal{F}_t^{i,j} \quad (4)$$

where

$$\mathcal{Y}_t^{i,j} = \begin{cases} \emptyset, & \text{with probability } 1 - p_d^{i,j} \\ \{y_t^{i,j}\}, & \text{with probability } p_d^{i,j} \end{cases} \quad (5)$$

is the set of system-originated measurements and $\mathcal{F}_t^{i,j}$ the finite set of fake measurements.

Then, the aim of this paper is to address the problem of distributed joint attack detection and state estimation, which amounts to jointly estimating, at each time t and in each fusion node $j \in \mathcal{N}$, the state x_t and, when present, the attack signal a_t only on the basis of: the measurement sets $\mathcal{Z}_1^{i,j} \ldots, \mathcal{Z}_t^{i,j}$ received up to time t from the sensor nodes i belonging to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

cluster $\mathcal{S}_j$; the data received from all adjacent fusion nodes $\ell \in \mathcal{N}_j$.

III. Hybrid Bernoulli random set for joint attack detection and state estimation

In order to address the joint state/attack estimation problem, it is convenient to introduce the *attack set* at time t, $\mathcal{A}_t$, which is either equal to the empty set if the system is not under signal attack at time t or to the singleton $\{a_t\}$ otherwise, i.e.

$$\mathcal{A}_t = \begin{cases} \emptyset, & \text{if the system is not under signal attack} \\ \{a_t\}, & \text{otherwise.} \end{cases} \tag{6}$$

In this paper, the estimation problem is addressed in a Bayesian framework by exploiting the concept of *Random Finite Sets* (RFSs), i.e. variables which are random in both the number of elements and the values of the elements. In fact, as shown hereafter, RFSs represent a convenient way to model both the attack set $\mathcal{A}_t$ and the measurement sets $\mathcal{Z}_t^{i,j}$ within a common framework.

A. Random set estimation

An RFS $\mathcal{X}$ 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 [14]; 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 $f(x_1, x_2, \ldots, x_n|n)$ over $\mathbb{X}^n$ given that $\mathcal{X}$ have n elements, are obtained as follows:

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

where $f(\mathcal{X}) = f(\{x_1, \ldots, x_n\}) = n!\, f(x_1, \ldots, x_n)$ denotes, using the set and, respectively, vector notation, the FISST probability density of RFS $\mathcal{X}$. In fact, the multi-object distribution $f(\{x_1, \ldots, x_n\})$ (in set notation) can also be expressed in vector notation, noting that the probability assigned to the finite set $\{x_1, \ldots, x_n\}$ must be equally distributed among the $n!$ possible permutations of the same elements. In order to measure probability over subsets of $\mathbb{X}$ or compute expectations of random set variables, Mahler [14] introduced the notion of *set integral* for a generic real-valued function $g(\mathcal{X})$ of an RFS $\mathcal{X}$ as

$$\int g(\mathcal{X})\, \delta\mathcal{X} = g(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int g(\{x_1, \ldots, x_n\})\, dx_1 \cdots dx_n. \tag{7}$$

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

1) 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:

$$f(\mathcal{X}) = \begin{cases} 1 - r, & \text{if } \mathcal{X} = \emptyset \\ r \cdot p(x), & \text{if } \mathcal{X} = \{x\}. \end{cases} \tag{8}$$

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

$$\rho(n) = \frac{e^{-\mu}\mu^n}{n!}, \quad n = 0, 1, 2, \ldots \tag{9}$$

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

$$f(\mathcal{X}) = e^{-\mu} \prod_{x \in \mathcal{X}} \mu\, p(x). \tag{10}$$

B. Hybrid Bernoulli random set

We now consider the problem of simultaneous detection and estimation of the signal attack and of the state of the system under monitoring, given a set of observations. The key idea is to use the random set paradigm to model the switching nature of the signal attack (presence/absence) by means of a Bernoulli random set $\mathcal{A}$ defined in (6) (i.e. a set that, with some probability r, can be either empty or a singleton depending on the presence or not of the attack) and the possible injection of fake measurements by means of a random measurement set $\mathcal{Z}$ defined in (3) and (4) as a Bernoulli or Poisson RFS for the packet substitution or, respectively, extra packet injection attack. It is worth pointing out that the posed Bayesian estimation problem is neither standard [13] nor Bernoulli filtering [14], [15], [26] but is rather a hybrid Bayesian filtering problem that aims to jointly estimate a Bernoulli random set $\mathcal{A}$ for the signal attack and a random vector x for the system state. An analytical solution of the hybrid filtering problem has been found in [11] in terms of integral equations that generalize the Bayes and Chapman-Kolmogorov equations of the Bernoulli filter [26]. The key feature of this *hybrid Bernoulli filter* is that it jointly estimates the posterior PDF of the system state and of the signal attack (when the system is assumed under attack) as well as the probability of attack existence. This is made possible thanks to the following definition of *hybrid Bernoulli random set*. Let the signal attack input be modeled as a Bernoulli random set $\mathcal{A} \in \mathcal{B}(\mathbb{A})$, 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 x. Then, we can define the HBRS $(\mathcal{A}, x)$, as a new state variable which incorporates the Bernoulli attack 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 (signal attack) probability r of $\mathcal{A}$ being a singleton, the PDF

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

$p^0(x)$ defined on the state space $\mathbb{X}$, and the joint PDF $p^1(a,x)$ defined on the joint attack input-state space $\mathbb{A}\times\mathbb{X}$, i.e.

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

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

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

where the set integration with respect to $\mathcal{A}$ is defined according to (7) 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) into (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 HBRS density throughout the rest of the paper. Note that, in order to model the signal attack presence/absence, it is convenient to introduce a binary random variable $\epsilon_t\in\{0,1\}$, referred to as the *attack existence*. By convention, $\epsilon_t=1$ means that the system is under signal attack at time t, i.e. $\mathcal{A}_t\neq\emptyset$. By contrast, if $\epsilon_t=0$ the system is not under signal attack at time t, i.e. $\mathcal{A}_t=\emptyset$. Thus, the signal attack is effectively modeled by a Bernoulli random set $\mathcal{A}_t$ which is either empty (if $\epsilon_t=0$) or a singleton $\mathcal{A}_t=\{a_t\}$ when $\epsilon_t=1$. The notion of attack existence is used to detect the presence (existence) of a signal attack and thus initiate its estimation by means of the posterior probability of attack existence $r_t=\text{Prob}(\mathcal{A}_t\neq\emptyset|\mathcal{Z}_t)$. In particular, the centralized hybrid Bernoulli Bayesian filter proposed in [11] for joint attack detection-state estimation propagates in time, via a two-step prediction-correction procedure, a joint posterior density completely characterized by a triplet consisting of: (1) the probability of existence of the signal attack r; (2) the PDF $p^0(x)$ in the state space for the system under no signal attack; (3) the PDF $p^1(a,x)$ in the joint attack input-state space for the system under signal attack. The triplet $(r,p^0(\cdot),p^1(\cdot,\cdot))$ provides useful information for attack detection, state estimation and attack reconstruction. Specifically, the estimated probability of attack existence is used to take a decision about attack existence. Decision rules can be based on different criteria, such as the Maximum A posteriori Probability (MAP) which compares the *a posteriori* probabilities $\text{Prob}(\mathcal{A}\neq\emptyset|\mathcal{Z})$ and $\text{Prob}(\mathcal{A}=\emptyset|\mathcal{Z})$ of the two hypotheses on attack existence $\hat{\mathcal{A}}\neq\emptyset$ and $\hat{\mathcal{A}}=\emptyset$ (the system is under signal attack or not) via a simple binary hypothesis test [27]. Given the decision about signal attack existence, then secure state estimation can be performed on the basis of the available posteriors, either $p^0(\cdot)$ or $p^1(\cdot,\cdot)$. In fact, optimal point estimates of the state of the system under nominal operation or, respectively, of the attack input and of the state of the system under attack can be obtained from $p^0(\cdot)$ or $p^1(\cdot,\cdot)$ according to some criterion, e.g., MAP or Minimum Mean Squared Error (MMSE). The resulting HBRS filter, as a sequential Bayesian estimator, recursively estimates the triplet $(r,p^0(\cdot),p^1(\cdot,\cdot))$ through the prediction and correction steps, by using the received observation set as well as the measurement and dynamic models described below.

C. Measurement models and likelihood functions

1) Packet substitution: Let us consider the *packet substitution* attack model introduced in Section II-A and denote by $\lambda(\mathcal{Z}_t^{i,j}|\mathcal{A}_t,x_t)$ the likelihood function of the measurement set defined in (3), which has obviously two possible forms, $\mathcal{A}_t$ being a Bernoulli random set. In particular, for $\mathcal{A}_t=\emptyset$:

$$\lambda(\mathcal{Z}_t^{i,j}|\emptyset,x_t)= \tag{13}$$
$$\begin{cases}1-p_d^{i,j}, & \text{if } \mathcal{Z}_t^{i,j}=\emptyset\\ \\ p_d^{i,j}\big[(1-p_f^{i,j})\,\ell^i(z|x_t)+p_f^{i,j}\,\kappa^{i,j}(z)\big], & \text{if } \mathcal{Z}_t^{i,j}=\{z\}\end{cases}$$

where $\{z\}$ denotes the singleton whose element represents a delivered measurement, i.e. $\lambda(\{z\}|\mathcal{A}_t,x_t)$ is the likelihood that a single measurement z will be collected. Furthermore, $\ell^i(z|x_t)$ is the standard likelihood function of the system-generated measurement z when no signal attack is present, whereas $\kappa^{i,j}(\cdot)$ is a PDF modeling the fake measurement $\tilde{y}_t^{i,j}$, assumed to be independent of the system state. Conversely, for $\mathcal{A}_t=\{a_t\}$:

$$\lambda(\mathcal{Z}_t^{i,j}|\{a_t\},x_t)= \tag{14}$$
$$\begin{cases}1-p_d^{i,j}, & \text{if } \mathcal{Z}_t^{i,j}=\emptyset\\ \\ p_d^{i,j}\big[(1-p_f^{i,j})\,\ell^i(z|a_t,x_t)+p_f^{i,j}\,\kappa^{i,j}(z)\big], & \text{if } \mathcal{Z}_t^{i,j}=\{z\}\end{cases}$$

where $\ell^i(z|a_t,x_t)$ denotes the conventional likelihood of measurement z, due to the system under attack a_t in state x_t. Notice that, by using the definition of set integral (7), it is easy to check that both forms (13) and (14) of the likelihood function $\lambda(\mathcal{Z}_t^{i,j}|\mathcal{A}_t,x_t)$ integrate to one.

2) Extra packet injection: Let us now consider the *extra packet injection* attack model introduced in Section II-A, for which the measurement set defined in (4) is given by the union of two independent random sets. As it is clear from (5), $\mathcal{Y}_t^{i,j}$ is a Bernoulli random set (with cardinality $|\mathcal{Y}_t^{i,j}|$ at most 1) which depends on whether the sensor-originated measurement y_t^i is delivered or not. Conversely, $\mathcal{F}_t^{i,j}$ 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 (9) and the FISST PDF of fake-only measurements $\gamma(\mathcal{F}_t^{i,j})$ is given by (10) with spatial distribution $\kappa^{i,j}(\cdot)$ in place of $p(\cdot)$. For the measurement set (4), the aim is to find the expression of the likelihood function $\lambda(\mathcal{Z}_t^{i,j}|\mathcal{A}_t,x_t)$. To this end, let us first introduce the following FISST PDF for $\mathcal{A}_t=\emptyset$:

$$\eta(\mathcal{Y}_t^{i,j}|\emptyset,x_t)=\begin{cases}1-p_d^{i,j}, & \text{if } \mathcal{Y}_t^{i,j}=\emptyset\\ \\ p_d^{i,j}\,\ell^i(z|x_t), & \text{if } \mathcal{Y}_t^{i,j}=\{z\}\end{cases} \tag{15}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

and for $\mathcal{A}_t = \{a_t\}$:

$$\eta(\mathcal{Y}_t^{i,j}|\{a_t\}, x_t) = \begin{cases} 1 - p_d^{i,j}, & \text{if } \mathcal{Y}_t^{i,j} = \emptyset \\ p_d^{i,j}\, \ell^i(z|a_t, x_t), & \text{if } \mathcal{Y}_t^{i,j} = \{z\} \end{cases}. \tag{16}$$

Then, using the convolution formula [14, p. 385], it follows that

$$\lambda(\mathcal{Z}_t^{i,j}|\mathcal{A}_t, x_t) = \sum_{\mathcal{Y}_t^{i,j} \subseteq \mathcal{Z}_t^{i,j}} \eta(\mathcal{Y}_t^{i,j}|\mathcal{A}_t, x_t)\, \gamma(\mathcal{Z}_t^{i,j} \backslash \mathcal{Y}_t^{i,j}). \tag{17}$$

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

$$\begin{aligned} &\lambda(\mathcal{Z}_t^{i,j}|\emptyset, x_t) \\ &= \eta(\emptyset|\emptyset, x_t)\, \gamma(\mathcal{F}_t^{i,j}) + \sum_{z \in \mathcal{Z}_t^{i,j}} \eta(\{z\}|\emptyset, x_t)\, \gamma(\mathcal{Z}_t^{i,j} \setminus \{z\}) \\ &= \gamma(\mathcal{F}_t^{i,j}) \left[1 - p_d^{i,j} + p_d^{i,j} \sum_{z \in \mathcal{Z}_t^{i,j}} \frac{\ell^i(z|x_t)}{\mu\, \kappa^{i,j}(z)}\right] \end{aligned} \tag{18}$$

where (15) and (10) have been used, while for $\mathcal{A}_t = \{a_t\}$ we have

$$\begin{aligned} &\lambda(\mathcal{Z}_t^{i,j}|\{a_t\}, x_t) \\ &= \eta(\emptyset|\{a_t\}, x_t)\, \gamma(\mathcal{F}_t^{i,j}) + \sum_{z \in \mathcal{Z}_t^{i,j}} \eta(\{z\}|\{a_t\}, x_t)\, \gamma(\mathcal{Z}_t^{i,j} \setminus \{z\}) \\ &= \gamma(\mathcal{F}_t^{i,j}) \left[1 - p_d^{i,j} + p_d^{i,j} \sum_{z \in \mathcal{Z}_t^{i,j}} \frac{\ell^i(z|a_t, x_t)}{\mu\, \kappa^{i,j}(z)}\right]. \end{aligned} \tag{19}$$

D. Dynamic model

Let us finally introduce the dynamic model of the HBRS $(\mathcal{A}, x)$. First, it is assumed that, in the case of a system under normal operation at time t, an attack a_{t+1} will be launched to the system by an adversary during the sampling interval with probability (of attack-birth) p_b. On the other hand, if the system is under attack (i.e., $\mathcal{A}_t$ is a singleton), it is supposed that the adversarial action will endure from time step t to time step $t+1$ with probability (of attack-survival) p_s. It is further assumed that $(\mathcal{A}, x)$ is a Markov process with joint transitional density

$$m(\mathcal{A}_{t+1}, x_{t+1}|\mathcal{A}_t, x_t) = m(x_{t+1}|\mathcal{A}_t, x_t)\, m(\mathcal{A}_{t+1}|\mathcal{A}_t) \tag{20}$$

which ensues from considering the attack as a stochastic process independent of the system state. Such an assumption is motivated by the fact that (i) a_t 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 (ii) the knowledge of a_t adds no information on a_s, if $s \neq t$. In addition, note that

$$m(x_{t+1}|\mathcal{A}_t, x_t) = \begin{cases} m(x_{t+1}|x_t), & \text{if } \mathcal{A}_t = \emptyset \\ m(x_{t+1}|a_t, x_t), & \text{if } \mathcal{A}_t = \{a_t\} \end{cases} \tag{21}$$

are known Markov transition PDFs, while the dynamics of the Markov process $\mathcal{A}_t$ resulting from the aforestated assumptions is Bernoulli, described by the following densities:

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

$$m(\mathcal{A}_{t+1}|\{a_t\}) = \begin{cases} 1 - p_s, & \text{if } \mathcal{A}_{t+1} = \emptyset \\ p_s\, p(a_{t+1}), & \text{if } \mathcal{A}_{t+1} = \{a_{t+1}\} \end{cases}$$

where $p(a_{t+1})$ 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 $p(a_{t+1})$.

In [11], it was shown that, when the above-described measurement and dynamic models are used and a centralized setting is taken into account, HBRSs are closed with respect to both the prediction and correction steps of the Bayes filter recursion and the resulting filter can be derived in closed-form. In order to make it possible to extend such results to the considered distributed setting, the problem of how to fuse HBRSs needs to be addressed in the next section.

IV. Distributed fusion of hybrid Bernoulli random sets

The focus of this section is on how to fuse local HBRS densities coming from multiple fusion nodes. A key issue is how to consistently fuse such densities taking into account that the agents may share common information and that such common information is impossible to single out. Hence, optimal (Bayes) fusion [18], [19] has to be ruled out and some robust suboptimal fusion approach has to be undertaken. In this respect, the paradigm of Kullback-Leibler fusion (average) has been successfully introduced in [28] for single-object PDFs and has been extended to FISST densities in [29]. From a notational point of view, please notice that in this section the fusion agent j is indicated as subscript while in the other parts of the paper where also the time t appears, j is indicated as superscript (and t as subscript).

A. Kullback-Leibler fusion

Given two FISST probability densities $f(\mathcal{X})$ and $g(\mathcal{X})$, let us first define the *Kullback-Leibler divergence* (KLD) from $g(\cdot)$ to $f(\cdot)$ as

$$D_{KL}\left(f \parallel g\right) \triangleq \int f\left(\mathcal{X}\right) \log \frac{f(\mathcal{X})}{g(\mathcal{X})}\, \delta\mathcal{X} \tag{22}$$

where the integral in (22) must be interpreted as a set integral according to the definition (7). Then, the *weighted KLA* $\overline{f}$ of the agent multi-object densities f_j, $j \in \mathcal{N}$, is defined as follows

$$\overline{f} = \arg\inf_{f} \sum_{j \in \mathcal{N}} \omega_j\, D_{KL}\left(f \parallel f_j\right) \tag{23}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

with weights ω_j satisfying

$$\omega_j \geq 0, \qquad \sum_{j\in\mathcal{N}} \omega_j = 1. \tag{24}$$

Notice from (23) that the weighted KLA of the agent densities is the one that minimizes the weighted sum of distances from such densities. In particular, the choice $\omega_j = 1/|\mathcal{N}|$ for any $j \in \mathcal{N}$ in (23) provides the (uniformly weighted) KLA which averages the agent densities giving to all of them the same level of confidence. An interesting interpretation of such a notion can be given recalling that, in Bayesian statistics, the KLD (22) can be seen as the information gain achieved when moving from a prior $g(\mathcal{X})$ to a posterior $f(\mathcal{X})$. Thus, according to (23), the average PDF is the one that minimizes the sum of the information gains from the initial multi-object densities. This choice is also coherent with the *Principle of Minimum Discrimination Information* (PMDI) according to which the probability density which best represents the current state of knowledge is the one which produces an information gain as small as possible (see [30], [31]). The adherence to the PMDI is important in order to counteract the so-called *data incest* phenomenon, i.e. the unaware reuse of the same piece of information due to the presence of loops within the network.

The following fundamental result holds.

Theorem 1: (*Kullback-Leibler fusion of general multi-object densities [29]*) - The weighted KLA defined in (23) turns out to be given by

$$\overline{f}\left(\mathcal{X}\right) = \frac{\prod_{j\in\mathcal{N}} \left[f_j\left(\mathcal{X}\right)\right]^{\omega_j}}{\int \prod_{j\in\mathcal{N}} \left[f_j\left(\mathcal{X}\right)\right]^{\omega_j} \delta\mathcal{X}}. \tag{25}$$

Notice that (25) states that the fused density $\overline{f}$ is nothing but the normalized weighted geometric mean of the agent densities. It must be pointed out that the fusion rule (25), which has been derived as KLA of the local multi-object densities, coincides with the *Generalized Covariance Intersection* for multi-object fusion first proposed by Mahler [19] and also known as *Exponential Mixture Density* [21].

When the agent densities are HBRS densities, the following result holds.

Theorem 2: The weighted KLA of agent HBRSs, with densities

$$p_j(\mathcal{A},x) = \begin{cases} (1-r_j)\, p_j^0(x), & \text{if } \mathcal{A}=\emptyset \\ r_j\, p_j^1(a,x), & \text{if } \mathcal{A}=\{a\} \end{cases}, \quad j\in\mathcal{N} \tag{26}$$

and fusion weights ω_j satisfying (24), is a HBRS with density given by

$$\overline{p}(\mathcal{A},x) = \begin{cases} (1-\overline{r})\, \overline{p}^0(x), & \text{if } \mathcal{A}=\emptyset \\ \overline{r}\, \overline{p}^1(a,x), & \text{if } \mathcal{A}=\{a\} \end{cases} \tag{27}$$

where

$$\overline{r} = \frac{\prod_{j\in\mathcal{N}} [r_j]^{\omega_j} \overline{\eta}^1}{\prod_{j\in\mathcal{N}} [r_j]^{\omega_j} \overline{\eta}^1 + \prod_{j\in\mathcal{N}} [1-r_j]^{\omega_j} \overline{\eta}^0} \tag{28}$$

$$\overline{p}^0(x) = \frac{\prod_{j\in\mathcal{N}} [p_j^0(x)]^{\omega_j}}{\overline{\eta}^0} \tag{29}$$

$$\overline{p}^1(a,x) = \frac{\prod_{j\in\mathcal{N}} [p_j^1(a,x)]^{\omega_j}}{\overline{\eta}^1} \tag{30}$$

with

$$\overline{\eta}^0 = \int \prod_{j\in\mathcal{N}} [p_j^0(x)]^{\omega_j} dx \tag{31}$$

$$\overline{\eta}^1 = \int \prod_{j\in\mathcal{N}} [p_j^1(a,x)]^{\omega_j} da\, dx. \tag{32}$$

Proof: First, we compute the numerator of the KLA fusion in (25), i.e., $\prod_{j\in\mathcal{N}} [f_j(\mathcal{X})]^{\omega_j}$. Substitution of the hybrid Bernoulli densities (26) of each node into this term yields, if $\mathcal{A}=\emptyset$,

$$\begin{aligned} &\prod_{j\in\mathcal{N}} [p_j(\mathcal{A},x)]^{\omega_j} \\ &= \prod_{j\in\mathcal{N}} [(1-r_j) p_j^0(x)]^{\omega_j} \\ &= \prod_{j\in\mathcal{N}} [1-r_j]^{\omega_j} \prod_{j\in\mathcal{N}} [p_j^0(x)]^{\omega_j}, \end{aligned} \tag{33}$$

otherwise if $\mathcal{A}=\{a\}$,

$$\begin{aligned} &\prod_{j\in\mathcal{N}} [p_j(\mathcal{A},x)]^{\omega_j} \\ &= \prod_{j\in\mathcal{N}} [r_j\, p_j^1(a,x)]^{\omega_j} \\ &= \prod_{j\in\mathcal{N}} [r_j]^{\omega_j} \prod_{j\in\mathcal{N}} [p_j^1(a,x)]^{\omega_j}. \end{aligned} \tag{34}$$

Then, we compute the denominator of (25), i.e., $\int \prod_{j\in\mathcal{N}} [f_j(\mathcal{X})]^{\omega_j} \delta\mathcal{X}$. According to the integral of hybrid Bernoulli densities (12), we have

$$\begin{aligned} &\int \prod_{j\in\mathcal{N}} [p_j(\mathcal{A},x)]^{\omega_j} \delta\mathcal{A}\, dx \\ &= \prod_{j\in\mathcal{N}} [1-r_j]^{\omega_j} \int \prod_{j\in\mathcal{N}} [p_j^0(x)]^{\omega_j} dx \\ &\quad + \prod_{j\in\mathcal{N}} [r_j]^{\omega_j} \int \prod_{j\in\mathcal{N}} [p_j^1(a,x)]^{\omega_j} da\, dx \\ &= \prod_{j\in\mathcal{N}} [1-r_j]^{\omega_j} \overline{\eta}^0 + \prod_{j\in\mathcal{N}} [r_j]^{\omega_j} \overline{\eta}^1 \end{aligned} \tag{35}$$

where $\overline{\eta}^0$ and $\overline{\eta}^1$ are given by (31) and (32) respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

Hence, if $\mathcal{A}=\emptyset$, we can obtain the weighted KLA $\overline{p}(\mathcal{A},x)$ by combining (33) and (35), i.e.,

$$\begin{aligned}&\overline{p}(\mathcal{A},x)\\ =&\frac{\prod_{j\in\mathcal{N}}[1-r_j]^{\omega_j}\prod_{j\in\mathcal{N}}[p_j^0(x)]^{\omega_j}}{\prod_{j\in\mathcal{N}}[1-r_j]^{\omega_j}\overline{\eta}^0+\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}\\ =&\frac{\prod_{j\in\mathcal{N}}[1-r_j]^{\omega_j}\overline{\eta}^0}{\prod_{j\in\mathcal{N}}[1-r_j]^{\omega_j}\overline{\eta}^0+\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}\cdot\frac{\prod_{j\in\mathcal{N}}[p_j^0(x)]^{\omega_j}}{\overline{\eta}^0}\\ =&\left(1-\frac{\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}{\prod_j[1-r_j]^{\omega_j}\overline{\eta}^0+\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}\right)\cdot\frac{\prod_{j\in\mathcal{N}}[p_j^0(x)]^{\omega_j}}{\overline{\eta}^0}\\ =&(1-\overline{r})\overline{p}^0(x)\end{aligned} \tag{36}$$

where $\overline{r}$ and $\overline{p}^0(x)$ are given in (28) and (29) respectively.

Similarly, we can also get the weighted KLA $\overline{p}(\mathcal{A},x)$ when $\mathcal{A}=\{a\}$ by combining (34) and (35), i.e.,

$$\begin{aligned}&\overline{p}(\mathcal{A},x)\\ =&\frac{\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\prod_{j\in\mathcal{N}}[p_j^1(a,x)]^{\omega_j}}{\prod_{j\in\mathcal{N}}[1-r_j]^{\omega_j}\overline{\eta}^0+\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}\\ =&\frac{\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}{\prod_{j\in\mathcal{N}}[1-r_j]^{\omega_j}\overline{\eta}^0+\prod_{j\in\mathcal{N}}[r_j]^{\omega_j}\overline{\eta}^1}\cdot\frac{\prod_{j\in\mathcal{N}}[p_j^1(a,x)]^{\omega_j}}{\overline{\eta}^1}\\ =&\overline{r}\,\overline{p}^1(a,x)\end{aligned} \tag{37}$$

where $\overline{p}^1(a,x)$ is given in (30).

As a result, the weighted KLA of agent HBRS densities is still a HBRS density characterized by the quantities $\overline{r}$, $\overline{p}^0(x)$ and $\overline{p}^1(a,x)$ in (28)-30). ■

B. Immunity to data incest

The KLA fusion guarantees immunity to double counting of information [32] and, further, the consensus approach always gives rise to densities which avoid data incest irrespectively of the number of consensus iterations being carried out. In this section, we illustrate this property of KLA fusion in the specific case of HBRS densities.

Example: Consider only two fusion nodes with associated conditional HBRS densities $p_j\left(\mathcal{A},x|\mathcal{Z}_j\right)$ for $j=1,2$, where $\mathcal{Z}_j=\cup_{i\in\mathcal{S}_j}\mathcal{Z}^{i,j}$ denotes the set of measurements received by node j. The collective information set $\mathcal{Z}_1\cup\mathcal{Z}_2$ can be decomposed into the union of the three disjoint information sets as follows $\mathcal{Z}_1\cup\mathcal{Z}_2=(\mathcal{Z}_1\setminus\mathcal{Z}_2)\cup(\mathcal{Z}_2\setminus\mathcal{Z}_1)\cup(\mathcal{Z}_1\cap\mathcal{Z}_2)$. Hence, the optimal fusion of $p_1(\cdot,\cdot)$ and $p_2(\cdot,\cdot)$ could be obtained as

$$\begin{aligned}&\overline{p}^o(\mathcal{A},x)\\ \propto&\,p(\mathcal{A},x|\mathcal{Z}_1\cup\mathcal{Z}_2)\\ \propto&\,p(\mathcal{A},x|\mathcal{Z}_1\setminus\mathcal{Z}_2)\,p(\mathcal{A},x|\mathcal{Z}_2\setminus\mathcal{Z}_1)\,p(\mathcal{A},x|\mathcal{Z}_1\cap\mathcal{Z}_2)\\ \propto&\,\frac{p_1(\mathcal{A},x)\,p_2(\mathcal{A},x)}{p(\mathcal{A},x|\mathcal{Z}_1\cap\mathcal{Z}_2)}\end{aligned} \tag{38}$$

if the HBRS densities $p(\mathcal{A},x|\mathcal{Z}_1\cap\mathcal{Z}_2)$ conditioned to the common information $\mathcal{Z}_1\cap\mathcal{Z}_2$ were known (the symbol $\propto$ stands for "proportional to" and the proportionality factor is determined by imposing that $\overline{p}^o(\cdot,\cdot)$ has unit integral over $\mathcal{B}(\mathbb{A})\times\mathbb{X}$). However, in the considered framework wherein nodes repeatedly fuse information from their neighbors without any knowledge about the network topology, it is impossible to single out the common information and thus apply (38).

Hence, some robust suboptimal fusion strategy has to be adopted. The simplest distributed averaging algorithm obtained via convex combination [22], [23] of the posteriors is the so-called "naive" distributed fusion and is given by

$$\begin{aligned}\overline{p}^{\text{naive}}(\mathcal{A},x)&\triangleq\frac{p_1(\mathcal{A},x)\,p_2(\mathcal{A},x)}{\int p_1(\cdot)\,p_2(\cdot)\,\delta\mathcal{A}\,dx}\\ &\propto p(\mathcal{A},x|\mathcal{Z}_1\setminus\mathcal{Z}_2)\,p(\mathcal{A},x|\mathcal{Z}_2\setminus\mathcal{Z}_1)[p(\mathcal{A},x|\mathcal{Z}_1\cap\mathcal{Z}_2)]^2.\end{aligned} \tag{39}$$

It can be observed from (39) that the naive distributed fusion involves double counting of common information compared with the optimal fusion in (38).

Another alternative is the KLA fusion of HBRS densities, adopted in this paper, which provides

$$\begin{aligned}\overline{p}(\mathcal{A},x)&\triangleq\frac{[p_1(\mathcal{A},x)]^{\omega_1}[p_2(\mathcal{A},x)]^{\omega_2}}{\int[p_1(\mathcal{A},x)]^{\omega_1}[p_2(\mathcal{A},x)]^{\omega_2}\delta\mathcal{A}\,dx}\\ &\propto[p(\mathcal{A},x|\mathcal{Z}_1\setminus\mathcal{Z}_2)]^{\omega_1}[p(\mathcal{A},x|\mathcal{Z}_2\setminus\mathcal{Z}_1)]^{\omega_2}p(\mathcal{A},x|\mathcal{Z}_2\cap\mathcal{Z}_1)\end{aligned} \tag{40}$$

where it can be seen that no double counting of common information occurs. The price to be paid is a conservative flattening $[p(\mathcal{Z}_1\setminus\mathcal{Z}_2|\mathcal{A},x)]^{\omega_1}[p(\mathcal{Z}_2\setminus\mathcal{Z}_1|\mathcal{A},x)]^{\omega_2}$ of exclusive information. Hence the fusion (27) under (24), turns out to be robust with respect to data incest. The interested reader is referred to [33] for a comparison between optimal and suboptimal distributed fusion rules (including naive and KLA) in terms of state estimation performance for a general linear-Gaussian model with two fusion nodes.

For the subsequent developments, it is convenient to introduce the operators $\oplus$ and $\odot$ defined as follows:

$$\begin{aligned}p\left(\mathcal{A},x\right)\oplus q\left(\mathcal{A},x\right)&\triangleq\frac{p\left(\mathcal{A},x\right)\,q\left(\mathcal{A},x\right)}{\int p(\mathcal{A},x)\,q(\mathcal{A},x)\,\delta\mathcal{A}\,dx}\\ \omega\odot p\left(\mathcal{A},x\right)&\triangleq\frac{[p(\mathcal{A},x)]^{\omega}}{\int[p(\mathcal{A},x)]^{\omega}\,\delta\mathcal{A}\,dx},\end{aligned}$$

the above integrals being HBRS integrals as in (12).

The previous example shows that no double counting occurs in the distributed KLA fusion in the case of two fusion nodes. Hereafter, it is mathematically proved that, in the general case with an arbitrary number of fusing nodes, the KLA distributed fusion ensures immunity to the double counting of information irrespective of the unknown common information in the densities p_j. To this end, for each $\mathcal{I}\in\mathcal{F}(\mathcal{N})$, let $Y_{\mathcal{I}}$ denote the information (i.e. measurements and/or prior information) which is available to all, and only, the nodes belonging to $\mathcal{I}$. Thus, $Y_{\{j\}}$ is the information available uniquely to node j. Accordingly, $Y_{\mathcal{N}}$ represents the information shared by the entire network. For instance, if the number of fusing nodes is $N_f=3$, we have $Y_{\{1,2\}}=(\mathcal{Z}_1\cap\mathcal{Z}_2)\setminus\mathcal{Z}_3$ and $Y_{\{1\}}=\mathcal{Z}_1\setminus(\mathcal{Z}_2\cup\mathcal{Z}_3)$.

By construction, the pieces of information $Y_{\mathcal{I}}$ are taken as mutually independent so that $Y_{\mathcal{I}}\cap Y_{\mathcal{I}'}=\emptyset$ for any $\mathcal{I},\mathcal{I}'\in$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

$\mathcal{F}(\mathcal{N})$ with $\mathcal{I} \neq \mathcal{I}'$. In other words, $\{Y_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{F}(\mathcal{N})}$ provides a (non-overlapping) partition of the collective information. Let $p\left(\mathcal{A}, x|Y_{\mathcal{I}}\right)$ now be the HBRS density conditioned to the information $Y_{\mathcal{I}}$. Then, in each node $i \in \mathcal{N}$, the HBRS density $p_i(\mathcal{A}, x)$ can be factorized as

$$p_i\left(\mathcal{A}, x\right) = \bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N}):\, i\in\mathcal{I}} p(\mathcal{A}, x|Y_{\mathcal{I}}) \tag{41}$$

Theorem 3: Let all the HBRS densities in (26) be factorized as in (41). Then, the distributed KLA fusion $\overline{p}$ in (27) turns out to be equal to

$$\overline{p}(\mathcal{A}, x) = \bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N})} \left[\left(\sum_{i\in\mathcal{I}} \omega_i\right) \odot p(\mathcal{A}, x|Y_{\mathcal{I}})\right]. \tag{42}$$

Proof: By substituting (41) into (25), and thanks to the properties of the operators $\oplus$ and $\odot$ in [29], we have

$$\begin{aligned}
\overline{p}(\mathcal{A}, x) &= \bigoplus_{i\in\mathcal{N}} \left[\omega_i \odot p_i\left(\mathcal{A}, x\right)\right] \\
&= \bigoplus_{i\in\mathcal{N}} \left[\omega_i \odot \left(\bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N}):\, i\in\mathcal{I}} p(\mathcal{A}, x|Y_{\mathcal{I}})\right)\right] \\
&= \bigoplus_{i\in\mathcal{N}} \left[\bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N}):\, i\in\mathcal{I}} \left(\omega_i \odot p(\mathcal{A}, x|Y_{\mathcal{I}})\right)\right] \\
&= \bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N})} \left[\bigoplus_{i\in\mathcal{I}} \left(\omega_i \odot p(\mathcal{A}, x|Y_{\mathcal{I}})\right)\right] \\
&= \bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N})} \left[\left(\sum_{i\in\mathcal{I}} \omega_i\right) \odot p(\mathcal{A}, x|Y_{\mathcal{I}})\right].
\end{aligned}$$

■

It can be seen from (42) that each independent piece of information $Y_{\mathcal{I}}$ is counted only once with weight $\sum_{i\in\mathcal{I}} \omega_i$. Since by construction $0 \leq \sum_{i\in\mathcal{I}} \omega_i \leq 1$, one can conclude that no double counting of common information occurs. In other words, the KLA fusion turns out to be inherently robust with respect to data incest. The price to be paid is a conservative flattening of independent information which occurs whenever $\sum_{i\in\mathcal{I}} \omega_i < 1$.

Remark 1: Similar to the case of two fusion nodes, the optimal Bayesian fusion of all HBRS densities p_i, $i \in \mathcal{N}$, could be obtained as

$$\overline{p}^o(\mathcal{A}, x) = \bigoplus_{\mathcal{I}\in\mathcal{F}(\mathcal{N})} p(\mathcal{A}, x|Y_{\mathcal{I}}) \tag{43}$$

where each independent piece of information $Y_{\mathcal{I}}$ is counted only once with unitary weight. However, the optimal fusion rule requires to single out the common information in order to compute $\overline{p}^o$, and hence it cannot be implemented in the considered framework as we analyzed in the previous example. When this is not possible, the KLA provides an effective solution to the HBRS density fusion problem in view of Theorem 3.

C. Consensus hybrid Bernoulli filter

Consensus [22], [23] has emerged as a powerful tool for distributed computation (e.g., averaging, minimization, maximization, ...) over networks and has found widespread application in distributed parameter/state estimation [22], [28], [34]–[39]. In essence, consensus aims to perform a collective computation over a whole network by iterating, in each node j of the network, a sequence of regional computations of the same type involving the subnetwork $\mathcal{N}_j$ of its in-neighbors. In the context of this work, it is assumed that each fusion node $j \in \mathcal{N}$ is provided with an agent HBRS density $p_j(\mathcal{A}, x)$ of form (26) and attempts to compute, in a distributed and scalable way, the collective Kullback-Leibler fusion

$$\overline{p} = \bigoplus_{j\in\mathcal{N}} \left(\frac{1}{|\mathcal{N}|} \odot p_j\right) = \frac{1}{|\mathcal{N}|} \odot \left(\bigoplus_{j\in\mathcal{N}} p_j\right). \tag{44}$$

To this end, let $p_{j,0} = p_j$, then a consensus algorithm for the computation of (44) takes the iterative form

$$p_{j,k+1}(\mathcal{A}, x) = \bigoplus_{h\in\mathcal{N}_j} \left(\, \omega_{j,h} \odot \, p_{h,k}(\mathcal{A}, x)\right), \qquad \forall j \in \mathcal{N} \tag{45}$$

where the *consensus weights* must satisfy the conditions

$$\omega_{j,h} \geq 0 \quad \forall j, h \in \mathcal{N}; \qquad \sum_{h\in\mathcal{N}_j} \omega_{j,h} = 1 \quad \forall j \in \mathcal{N}. \tag{46}$$

In fact, thanks to the properties of the operators $\oplus$ and $\odot$ listed in [29, p. 513], it can be seen that

$$p_{j,k}(\mathcal{A}, x) = \bigoplus_{h\in\mathcal{N}} \left(\, \omega_{j,h}^{(k)} \odot \, p_h(\mathcal{A}, x)\right), \qquad \forall j \in \mathcal{N} \tag{47}$$

where $\omega_{j,h}^{(k)}$ is defined as the element (j, h) of the matrix Ω^k and Ω is the consensus matrix whose generic (j, h)-element coincides with the consensus weight $\omega_{j,h}$ (if $h \notin \mathcal{N}_j$ then $\omega_{j,h}$ is taken as 0). In this respect, it is well known that if Ω is primitive (i.e. there exists an integer m such that all entries of Ω^m are strictly positive) and doubly stochastic (i.e. all its rows and columns sum up to one), one has

$$\lim_{k\to+\infty} \omega_{j,h}^{(k)} = \frac{1}{|\mathcal{N}|}, \quad \forall j, h \in \mathcal{N}.$$

Hence, as the number of consensus steps increases, each local density "tends" to the collective KLA (44).

A necessary condition for the matrix Ω to be primitive is that the graph $\mathcal{G}$ associated with the sensor network be strongly connected [37]. In this case, a possible choice ensuring convergence to the collective average for undirected graphs is given by the so-called *Metropolis weights* [23], [37]

$$\begin{aligned}
\omega_{j,h} &= \frac{1}{\max\{|\mathcal{N}_j|, |\mathcal{N}_h|\}}, \quad j \in \mathcal{N}, h \in \mathcal{N}_j, j \neq h \\
\omega_{j,j} &= 1 - \sum_{h\in\mathcal{N}_j, h\neq j} \omega_{j,h}.
\end{aligned}$$

In Theorem 3 it has been proved that the Kullback-Leibler fusion guarantees immunity to double counting of information and that, further, the consensus approach always gives rise to densities which avoid double counting irrespectively of the number of consensus iterations being carried out.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

V. Distributed Bayesian filter for joint attack detection and state estimation

Exploiting Theorem 2 on the Kullback-Leibler fusion of HBRSs and the HBRS filtering algorithm of [11, Section III], it is possible to derive a distributed joint attack and secure state estimation algorithm to be described hereinafter. Let at time t each fusion node $j \in \mathcal{N}$ have a HBRS density $p^j_{t-1}(\mathcal{A}, x)$ summarizing the available information on $(\mathcal{A}_{t-1}, x_{t-1})$, obtained by processing the measurements $\mathcal{Z}^j_{1:t-1} \triangleq \cup_{s=1}^{t-1} \cup_{i \in \mathcal{S}_j} \mathcal{Z}^{i,j}_s$ as well as by fusing information with the neighbors (i.e. fusion nodes belonging to $\mathcal{N}_j$). Then, the local HBRS $p^j_t(\cdot,\cdot)$ can be updated by means of the following steps to be carried out at each time t in fusion node j.

Distributed HBRS (DHBRS) filter (in node j at time t)

1) **Prediction** - Obtain $p^j_{t|t-1}(\mathcal{A}, x)$ from $p^j_{t-1}(\mathcal{A}, x)$ exploiting the dynamic model according to the results of Theorem 3 in reference [11].
2) **Correction** - Obtain $p^j_{t|t}(\mathcal{A}, x)$ from $p^j_{t|t-1}(\mathcal{A}, x)$ exploiting the measurements $\mathcal{Z}^j_t \triangleq \cup_{i \in \mathcal{S}_j} \mathcal{Z}^{i,j}_t$ according to the results of either Theorem 1 (for packet substitution) or of Theorem 2 (for packet injection) in reference [11].
3) **Fusion** - Initialize consensus by setting $p_{j,0}(\mathcal{A}, x) = p^j_{t|t}(\mathcal{A}, x)$. Then perform L consensus iterations, i.e. (45) for $k = 0, \ldots, L-1$, to finally get $p_{j,L}(\mathcal{A}, x)$. Then, set $p^j_t(\mathcal{A}, x) = p_{j,L}(\mathcal{A}, x)$.
4) **Attack detection & state estimation** - Perform attack detection using r^j_t from the available current HBRS $p^j_t(\cdot,\cdot)$. Based on the MAP decision rule, given $\mathcal{Z}^j_t$, assign $\hat{\mathcal{A}}^j_t \neq \emptyset$ (the system is under signal attack) if and only if $\text{Prob}(\mathcal{A}_t \neq \emptyset | \mathcal{Z}^j_t) > \text{Prob}(\mathcal{A}_t = \emptyset | \mathcal{Z}^j_t)$, where $\text{Prob}(\mathcal{A}_t \neq \emptyset | \mathcal{Z}^j_t) = r^j_t$ and $\text{Prob}(\mathcal{A}_t = \emptyset | \mathcal{Z}^j_t) = 1 - r^j_t$, i.e. iff $r^j_t > 1/2$. Finally, perform secure state estimation by extracting the estimates $\hat{x}^{0j}_t$ from $p^{0j}_t(x)$ (if $\hat{\mathcal{A}}^j_t = \emptyset$) or $\hat{x}^{1j}_t$ and $\hat{a}^j_t$ from $p^{1j}_t(a, x)$ (if $\hat{\mathcal{A}}^j_t \neq \emptyset$) according to some criterion (e.g., MAP or MMSE). ■

Notice that the first two steps (i.e. prediction and correction) together make up a local HBRS filtering cycle that allows to update information in node j on the basis of local measurements from $\mathcal{S}_j$ and the system model, while the third (fusion) step allows to diffuse information throughout the network. Recall that the HBRS density $p^j_t(\mathcal{A}, x)$ is characterized by the triplet $\left(r^j_t, p^{0j}_t(x), p^{1j}_t(a, x)\right)$ where the attack existence probability r^j_t can be used in the fourth step, e.g. by a MAP detector, to ascertain whether the system is under attack or not. Based on this decision about attack existence, a secure state estimate can be finally extracted from either $p^{0j}_t(x)$ or $p^{1j}_t(a, x)$ according to some criterion (e.g., MAP or MMSE). Note that whenever $\hat{\mathcal{A}}^j_t \neq \emptyset$, an optimal estimate of the unknown attack input can also be obtained from $p^{1j}_t(a, x)$. This means that, differently from previous work on secure state estimation of CPSs where the corrupted information is usually discarded, here we seek to guarantee not only robustness against attacks, but also performance restoration after the adversary-induced degradation by means of signal attack reconstruction using all the available information. For practical implementation of the HBRS filter, the infinite-dimensional PDFs $p^{0j}_t(x)$ and $p^{1j}_t(a, x)$ need to be approximated with finite-dimensional parametrizations, e.g., as Gaussian-mixtures. In [40] it has been shown that Gaussian-mixture HBRSs are closed under prediction and correction although both steps imply a growth of the number of Gaussian components that needs to be contrasted by suitable merging and/or pruning procedures. As far as the fusion step is concerned, this does not preserve the Gaussian-mixture form; however, good Gaussian-mixture approximations of the KLA of Gaussian-mixtures can be obtained using the techniques presented in [29] and [41].

Details on the Gaussian-mixture implementation can be found in [40] for the HBRS filter prediction and correction steps and in [29] for the fusion of Gaussian-mixtures.

Before ending this section, the following remarks are in order.

Remark 2: The proposed approach to resilient distributed state estimation in the presence of malicious attacks relies on the assumption that the communication between fusion nodes is secure. In fact, the deployment in the network of trusted nodes with higher security has been proposed as a possible way of ensuring resilience to adversarial attacks in distributed computation (see [42]–[44] and the references therein). In cases wherein the employment of secured nodes is ruled out, an alternative approach consists of introducing a redundancy in the communication topology of the network (for instance by increasing the connectivity degree of each fusion node) and then applying some outlier detection technique in order to detect data falsification attacks [45]. In this respect, since the proposed distributed information fusion algorithm is consensus-based, it is well-suited to be modified so as to include some data falsification attack mitigation technique following, for example, the ideas of [46]–[48]. For instance, each node j can flag as *suspicious* a neighbor ℓ for which the distance $D_{KL}(p^j_t(\mathcal{A}, x) || p^\ell_t(\mathcal{A}, x))$ between its local density and the one received from the neighbor exceeds a certain threshold τ^j_t. Then, using the flags received from neighboring nodes, a majority rule can be used to convert the status of neighbor ℓ from *suspicious* to *attacker* [48], [49]. Such flags can be used to modify the consensus weights, which are reduced for suspicious nodes and set to zero for nodes flagged as *attacker*. Since a full treatment of this issue would go beyond the scope of this manuscript, we refer the reader to [46]–[48] for in-depth studies on how to modify the consensus weights and how to adapt the time-varying thresholds τ^j_t so that the attackers are eventually filtered out.

Remark 3: As pointed out in Section IV, the proposed distributed information fusion algorithm is inherently immune to the data incest phenomenon. Hence, in accordance with the setting of Section II-B, it is admitted that each sensor node sends its measurements to more than one fusion center (as it can happen in the case of broadcast communication), in that the fusion rule prevents double counting of such information.

VI. Numerical example: Wide-area monitoring system

In this section, the performance of the proposed distributed random-set approach for joint attack detection and secure

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

CPS state estimation is analyzed in the presence of both signal and extra packet injection attacks as well as uncertainty on measurement delivery. Let us consider the motivating example of a 4-area IEEE 14-bus system (Fig. 1) introduced in Section II-A, consisting of 5 synchronous generators and 11 load buses. The parameters relative to transmission lines, generators' inertia and damping, nominal power injections and demands are the same considered in the 14-bus case of [50]. As shown in Fig. 1, the IEEE 14-bus system is partitioned into $N_f = 4$ distinct clusters containing $b_1 = 3, b_2 = 4, b_3 = 4$, and $b_4 = 3$ buses, respectively. The dynamics of the system can be described by the linearized swing equation [51] derived through the Kron reduction [52] of the linear small-signal power network model. The DC state estimation model assumes 1 p.u. (per unit) voltage magnitudes in all buses and $j1$ p.u. branch impedance, with j denoting imaginary unit. The system dynamics is thus represented by the evolution of $n = 10$ states comprising both the rotor angles δ^i and the frequencies ψ^i of each generator i in the 4-cluster network. After discretization (with sampling interval $T = 0.01s$), the model of the considered wide-area monitoring system takes the form (1)–(2), where the whole state and the phase angles in all buses of each cluster are measured by a network $\mathcal{S}_j$ of local sensors. The inter-area communication network is shown in Fig. 2. The system is assumed to be corrupted by additive zero mean Gaussian white process and measurement noises with variances $\sigma_w^2 = 10^{-5}$ and $\sigma_{v^i}^2 = 10^{-2}$. At time step $t = 50$ a signal attack vector $a_t = [0.7, 0.2]^T$ 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 74% and, respectively, 68%. This type of attack, referred to as *load altering attack* in [53], 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 test case, the probabilities of attack-birth and attack-survival are fixed, respectively, at $p_b = 0.05$ and $p_s = 0.95$. In each cluster, the system-generated measurement vector is supposed to be delivered at the local fusion node with probability $p_d = 0.99$. The extra fake measurements injected into the local sensor channels are modeled as a Poisson RFS with average number $\xi = 30$ and probability density uniformly distributed over the interval $[-10, 5]$, suitably chosen to emulate system-originated observations. Fig. 3 shows the resulting number of fake measurements maliciously injected at each time step.

The Gaussian-mixture implementation of the proposed Kullback-Leibler fusion with HBRSs in (27) is realized by sequentially applying the pairwise fusion rule (28)–(30) for two fusing nodes $|\mathcal{N}| - 1$ times, where the ordering of the pairwise fusions is irrelevant. A similar approach has been widely used in distributed fusion with other RFS based filters [29], [54]. The parameters of the Gaussian-mixture implementation are chosen as follows: the pruning threshold is $\gamma_p = 10^{-5}$; the merging threshold is $\gamma_m = 4$; the maximum number of Gaussian components is $N_{max} = 15$.

In this paper the problem of *security* in CPSs is addressed by considering, in a unified and general framework, the two fundamental aspects of attack detection and secure state estimation. In this context, the performance of the proposed

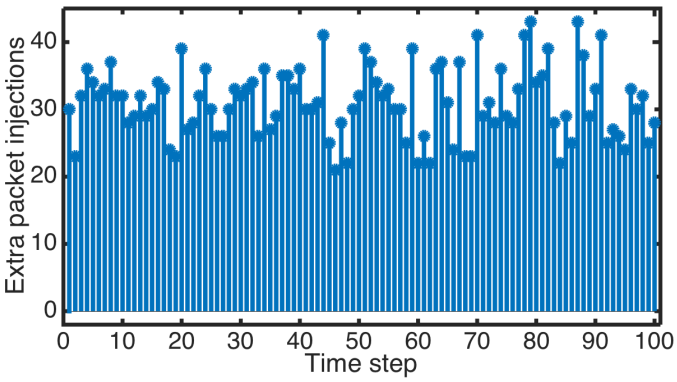


Fig. 3: Number of extra fake measurements injected vs. time.

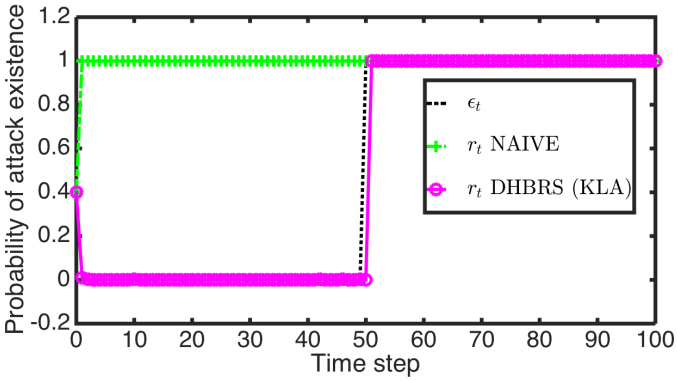


Fig. 4: Estimated probability of attack existence r_t compared to *true* attack existence ϵ_t. The signal attack gets into action at time step $t = 50$ ($\epsilon_t = 1$ for $t = 50, \ldots, 100$).

		Estimated attack set	
		$\hat{\mathcal{A}} \neq \emptyset$	$\hat{\mathcal{A}} = \emptyset$
True attack set	$\mathcal{A} \neq \emptyset$	98.04	1.96
	$\mathcal{A} = \emptyset$	0	100

TABLE I: Confusion matrix (in %) of the MAP detector. The false alarm and misdetection rates appear as off-diagonal entries.

DHBRS filter is evaluated in terms of both attack detection and secure state estimation. Figs. 4–9 show the performance of the distributed HBRS filter with $L = 10$ consensus steps. Specifically, Fig. 4 displays the performance in terms of attack detection by comparing the estimated probability of attack existence r_t (averaged over 100 independent Monte Carlo trials and all the fusion nodes) with the *true* binary random variable of attack existence ϵ_t. As it can be seen, while the proposed DHBRS filter demonstrates reliable detection of the signal attack, the filter implementing naive instead of KLA fusion (see Section IV-B) erroneously estimates the presence of a signal attack for the entire duration of the simulation, even when $\epsilon_t = 0$. The error matrix of the MAP detector, described in step 4) of the DHBRS filter, is reported in Table I to evaluate the attack detection accuracy. The percentage

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

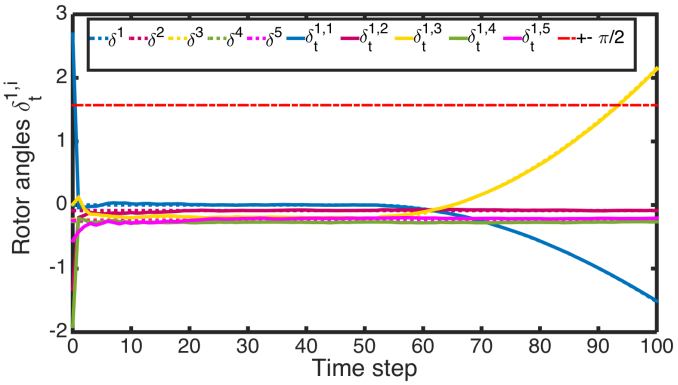


Fig. 5: *True* δ^i and estimated rotor angles $\delta_t^{j,i}$, for node $j = 1$ and generators $i = 1, \ldots, 5$.

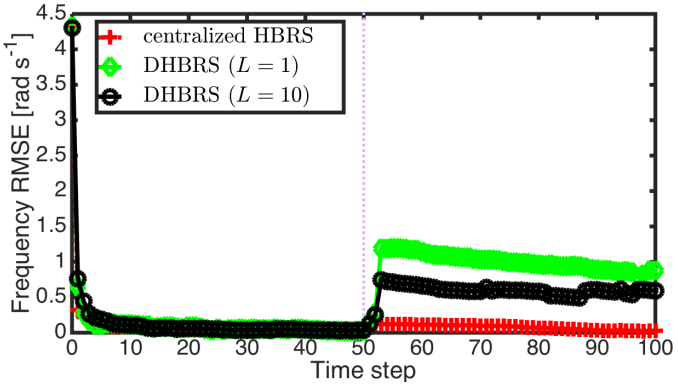


Fig. 6: Comparison of performance between centralized and distributed ($L = 1$ and $L = 10$) HBRS filters in terms of frequency RMSE.

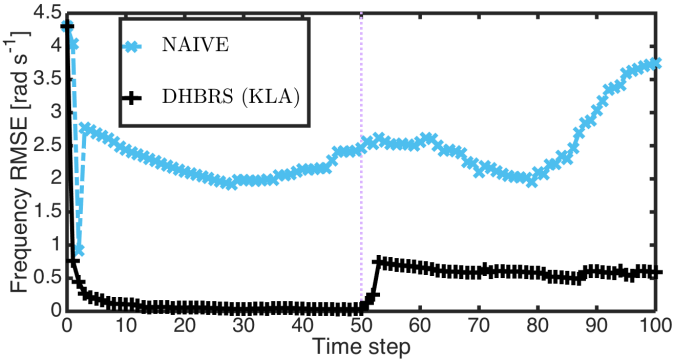


Fig. 7: Comparison of performance between distributed HBRS filters using KLA vs. naive fusion ($L = 10$) in terms of frequency RMSE.

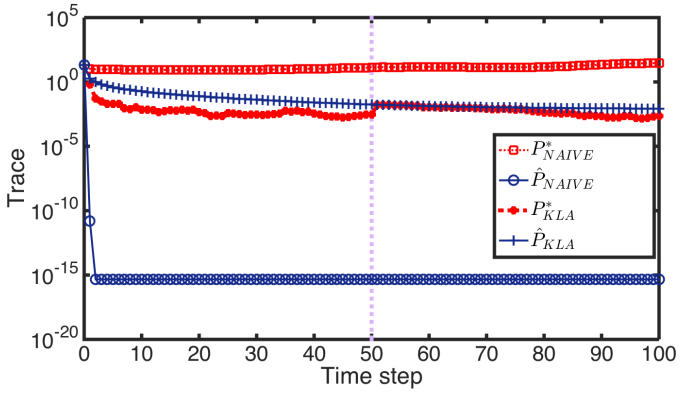


Fig. 8: Consistency check for the DHBRS filter implementing KLA vs. naive fusion step. The trace of the error covariance matrix $\widehat{P}$ is compared to the actual mean squared error P^* to check if the two distributed filters guarantee *consistent* secure state estimation.

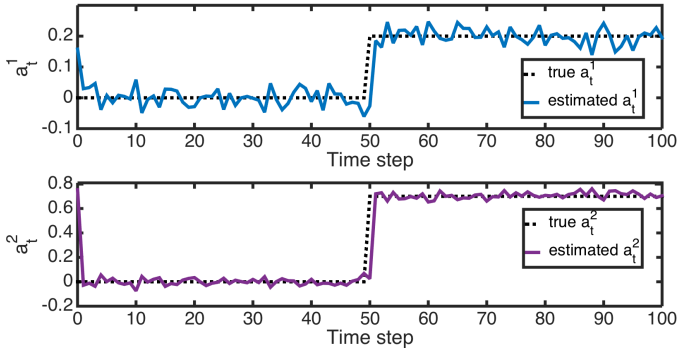


Fig. 9: *True* and estimated components of the attack vector.

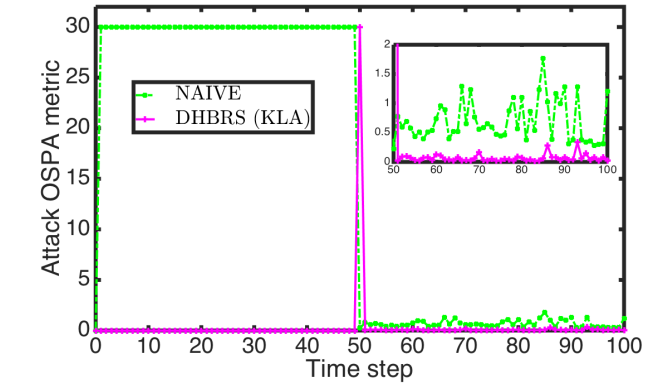


Fig. 10: Comparison between DHBRS (using KLA) and naive filter in terms of OSPA metric ($p = 1, c = 30$) of the attack random set $\mathcal{A}_t$.

errors, averaged over the number of Monte Carlo runs, show that the MAP decision rule is characterized by null false alarm rate and low misdetection rate (1.96%) due to a slight delay in attack detection. Figs. 5–9 evaluate the performance of the DHBRS filter in terms of secure state estimation. Fig. 5 provides a comparison between the true and the estimated values of rotor angles of each generator and clearly shows how δ^1 and δ^3 lose synchrony once the load altering attack gets into action. The centralized algorithm provides a performance benchmark for the proposed distributed strategy. Fig. 6 compares the Root Mean Square Error (RMSE), averaged over all fusion nodes and Monte Carlo runs, of the frequency estimates obtained with the DHBRS filter and, respectively, a centralized HBRS filter that processes measurements from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

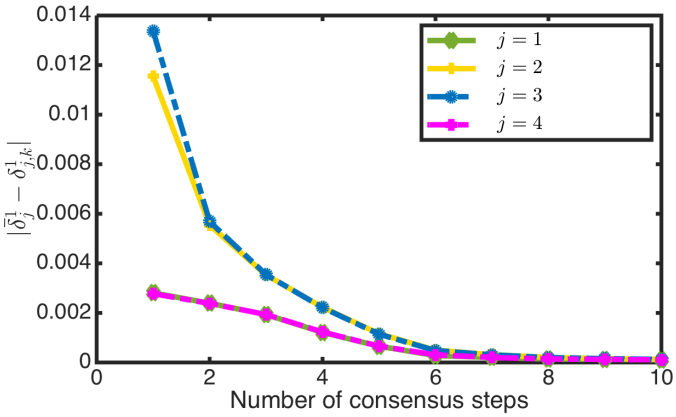


Fig. 11: Convergence rate of KLA fusion of HBRS densities. Each local estimate of rotor angle $\delta^1_{j,k}$ in node $j = 1, \ldots, 4$ tends to the collective KLA $\bar{\delta}^1$ for $k = 1, \ldots, L$ and $L = 10$.

all sensors. It can be observed that the performance of the DHBRS filters (performing $L = 1$ and, respectively, $L = 10$ consensus steps) is comparable to the one of the centralized algorithm which, in this linear-Gaussian case, is equivalent to what would be obtained using the optimal fusion (38). In particular, the error is almost identical for the two filters when the system is not under attack, while the gap in accuracy becomes more evident when the signal attack is active. Despite the short diameter of the considered communication network, the number of consensus steps affects the accuracy of state estimation when the system is under attack; though accuracy is already satisfactory with a single consensus step, it improves with $L = 10$. We also present the performance gain in terms of state estimation provided by KLA fusion with respect to the naive approach in Fig. 7, which clearly shows that the distributed naive filter provides unreliable estimates of the state. This is due to the fact that naive fusion combines the information from different nodes under the assumption that the local HBRS densities are independent when they are actually correlated via previous information flows propagated among neighbors. As shown in Fig. 8, this can lead to inconsistent estimates, i.e. estimates that do not satisfy the consistency condition [55] $\widehat{P} \geqslant P^*$, where $\widehat{P}$ is the error covariance matrix expressing the uncertainty associated with the estimate of the state vector and P^* is the actual mean squared error calculated by averaging the squared estimation error over the Monte Carlo trials. It can be noticed from Fig. 8 that the above consistency condition holds for the DHBRS filter performing KLA fusion, while inconsistent (i.e overconfident) estimates are obtained when the naive fusion rule is used. Fig. 9 evaluates the performance in terms of attack reconstruction by comparing the true and estimated components, extracted from $p^{11}_t(a, x)$, of the malicious signal attack for a single Monte Carlo realization. To provide an unique metric evaluating both attack detection and attack reconstruction performance, we also consider the well known OSPA (Optimal SubPattern Assignment) distance [56] for random sets, to measure the error between the true and estimated attack set also taking into account misdetections/false detections. A comparison between the DHBRS filter using KLA vs. naive fusion in terms of OSPA metric of order $p = 1$ and cut-off value $c = 30$ is shown in Fig. 10. The OSPA distance highlights how the use of naive fusion, which is not immune to data incest, leads to false detections before time step $t = 50$ and less accurate state estimation with respect to the proposed DHBRS filter when the system is under signal attack. Finally, the performance of the distributed HBRS filter is assessed in terms of convergence rate of the proposed consensus algorithm based on the Kullback-Leibler fusion of HBRSs (step 3 of the DHBRS filter described in Section V). Fig. 11 shows the behavior of the distance $|\bar{\delta}^1_j - \delta^1_{j,k}|$ for fusion nodes $j = 1, \ldots, 4$ concerning the estimated rotor angle of generator 1 as a function of the number of consensus steps. We can see how each local estimate $\delta^1_{j,k}$, averaged over all Monte Carlo trials and time steps, tends to the collective KLA $\bar{\delta}^1$ for $k = 1, \ldots, L$, i.e. as the number of consensus steps increases. Analogous results are achieved for the remaining generators of the power network.

VII. Conclusions

The paper has addressed a fundamental issue in the operation of networked cyber-physical systems (CPSs), i.e. how to detect incoming cyber-attacks and securely estimate the system state by means of distributed processing techniques. Different types of cyber-attacks (i.e. sensor/actuator data corruption, packet substitution and extra packet injection) and a cluster-based network configuration, wherein multiple cluster-heads receive data from remote sensors via non-secure links and exchange processed information neighborwise via secure links, have been considered. The joint attack-detection & state estimation problem has been cast in a Bayesian random set framework using *Hybrid Bernoulli Random Set* (HBRS) densities to summarize the available information on the signal attack and system state as well as appropriate filtering algorithms to update such densities. Then, distributed fusion of locally computed HBRSs has been exploited in order to spread information over the network thus deriving a novel distributed HBRS filter for secure monitoring of CPSs. A simulation case-study concerning a wide-area monitoring power system has been fully investigated in order to both motivate the proposed approach and demonstrate its practical effectiveness. Future work will concern: (i) extension to the case of non-secure links among cluster-heads; (ii) application to distributed detection & localization of malicious sources.

References

[1] "The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT)." `https://ics-cert.us-cert.gov/`.

[2] 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.

[3] 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.

[4] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *Proc. IEEE 54th Conference on Decision and Control*, pp. 5820–5826, Osaka, Japan, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSIPN.2017.2760804, IEEE Transactions on Signal and Information Processing over Networks

[5] 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.

[6] 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.

[7] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *Proc. 54th IEEE Conference on Decision and Control*, pp. 5827–5832, Osaka, Japan, 2015.

[8] 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, IL, USA, 2015.

[9] 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.

[10] 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.

[11] 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, NV, USA, 2016.

[12] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "Secure state estimation of cyber-physical systems under switching attacks," in *20th IFAC World Congress*, Toulouse, France, 2017.

[13] 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.

[14] R. P. S. Mahler, *Statistical Multisource Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.

[15] 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.

[16] Q. Gu, P. Liu, S. Zhu, and C.-H. Chu, "Defending against packet injection attacks in unreliable ad hoc networks," in *IEEE Global Telecommunications Conference*, vol. 3, pp. 1837–1841, St. Louis, MO, USA, 2005.

[17] 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`.

[18] C.-Y. Chong, S. Mori, and K.-C. Chang, *Distributed Multitarget Multisensor Tracking*. Y. Bar-Shalom Ed., Artech House: Multitarget-Multisensor Tracking: Advanced Applications, 1990.

[19] R. Mahler, "Optimal/robust distributed data fusion: a unified approach," in *Proc. of the SPIE Defense and Security Symposium*, vol. 4052, Orlando, FL, USA, 2000.

[20] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. American Control Conference*, vol. 4, pp. 2369–2373, Albuquerque, NM, USA, 1997.

[21] M. Uney, D. Clark, and S. Julier, "Distributed fusion of PHD filters via exponential mixture densities," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 521–531, 2013.

[22] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. 46th IEEE Conference on Decision and Control*, pp. 5492–5498, New Orleans, LA, USA, 2007.

[23] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symposium on Information Processing in Sensor Networks*, pp. 63–70, Los Angeles, CA, USA, 2005.

[24] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. M. Begovic, and A. Phadke, "Wide-area monitoring, protection, and control of future electric power networks," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 80–93, 2011.

[25] L. Xie, D. H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1154–1169, 2012.

[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] H. Van Trees, *Detection, Estimation, and Modulation Theory*. Wiley, 2004.

[28] G. Battistelli and L. Chisci, "Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.

[29] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and A. Graziano, "Consensus CPHD filter for distributed multitarget tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 508–520, 2013.

[30] L. Campbell, "Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities," *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 1011–1015, 1970.

[31] H. Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle*, pp. 199–213. New York, NY: Springer New York, 1998.

[32] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and R. Mahler, "Distributed fusion of multitarget densities and consensus PHD/CPHD filters," in *Proc. SPIE Defense, Security and Sensing*, vol. 9474, Baltimore, MD, USA, 2015.

[33] S. Mori, K. C. Chang, and C. Y. Chong, "Comparison of track fusion rules and track association metrics," in *Proc. 15th International Conference on Information Fusion*, pp. 1996–2003, Singapore, 2012.

[34] M. Kamgarpour and C. Tomlin, "Convergence properties of a decentralized Kalman filter," in *Proc. 47th IEEE Conference on Decision and Control*, pp. 3205–3210, Cancun, Mexico, 2008.

[35] F. S. Cattivelli and A. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, pp. 2069–2084, 2010.

[36] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, 2008.

[37] G. C. Calafiore and F. Abrate, "Distributed linear estimation over sensor networks," *International Journal of Control*, vol. 82, no. 5, pp. 868–882, 2009.

[38] S. S. Stankovic, M. Stankovic, and D. Stipanovic, "Consensus based overlapping decentralized estimation with missing observations and communication faults," *Automatica*, vol. 45, no. 6, pp. 1397–1406, 2009.

[39] M. Farina, G. Ferrari-Trecate, and R. Scattolini, "Distributed moving horizon estimation for linear constrained systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 11, pp. 2462–2475, 2010.

[40] 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: `https://arxiv.org/abs/1612.08478v1`.

[41] M. Günay, U. Orguner, and M. Demirekler, "Chernoff fusion of Gaussian mixtures based on sigma-point approximation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2732–2746, 2016.

[42] S. Zheng, T. Jiang, and J. S. Baras, "Robust state estimation under false data injection in distributed sensor networks," in *Proc. IEEE Global Telecommunications Conference*, pp. 1–5, Miami, FL, USA, 2010.

[43] M. Yampolskiy, Y. Vorobeychik, X. D. Koutsoukos, P. Horvath, H. J. LeBlanc, and J. Sztipanovits, "Resilient distributed consensus for tree topology," in *Proc. 3rd International Conference on High Confidence Networked Systems*, pp. 41–48, Berlin, Germany, 2014.

[44] W. Abbas, Y. Vorobeychik, and X. Koutsoukos, "Resilient consensus protocol in the presence of trusted nodes," in *Proc. 7th International Symposium on Resilient Control Systems*, pp. 1–7, Denver, CO, USA, 2014.

[45] V. P. Illiano and E. C. Lupu, "Detecting malicious data injections in wireless sensor networks: A survey," *ACM Computing Surveys*, vol. 48, no. 2, pp. 1–33, 2015.

[46] S. Liu, H. Zhu, S. Li, X. Li, C. Chen, and X. Guan, "Detecting integrity attacks on control systems using a moving target approach," in *Proc. IEEE Global Communications Conference*, pp. 603–608, Anaheim, CA, USA, 2012.

[47] S. Mi, H. Han, C. Chen, J. Yan, and X. Guan, "A secure scheme for distributed consensus estimation against data falsification in heterogeneous wireless sensor networks," *Sensors*, vol. 16, no. 2, 2016.

[48] M. Toulouse, H. Le, C. V. Phung, and D. Hock, "Robust consensus-based network intrusion detection in presence of Byzantine attacks," in *Proc. 7th International Symposium on Information and Communication Technology*, pp. 278–285, Ho Chi Minh, Vietnam, 2016.

[49] Q. Yan, M. Li, T. Jiang, W. Lou, and Y. T. Hou, "Vulnerability and protection for distributed consensus-based spectrum sensing in cognitive radio networks," in *Proc. 31st Annual IEEE International Conference on Computer Communications*, pp. 900–908, Orlando, FL, USA, 2012.

[50] 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.

[51] P. Kundur, N. Balu, and M. Lauby, *Power System Stability and Control*. McGraw-Hill, 1994.
[52] F. Pasqualetti, A. Bicchi, and F. Bullo, "A graph-theoretical characterization of power network vulnerabilities," in *Proc. American Control Conference*, pp. 3918–3923, San Francisco, CA, USA, 2011.
[53] 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, Washington, DC, USA, 2015.
[54] B. Wang, W. Yi, R. Hoseinnezhad, S. Li, L. Kong, and X. Yang, "Distributed fusion with multi-Bernoulli filter based on generalized Covariance Intersection," *IEEE Transactions on Signal Processing*, vol. 65, no. 1, pp. 242–255, 2017.
[55] M. Liggins, D. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice*. Taylor & Francis, 2008.
[56] D. Schuhmacher, B. T. Vo, and B. N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.

Nicola Forti received the B.S. degree in mechanical engineering, the M.S. degree in automation engineering and the Ph.D. degree in information engineering from the University of Florence, Florence, Italy, in 2009, 2013 and 2016. He is currently a Postdoctoral Researcher in the Department of Information Engineering, University of Florence. He is also affiliated with the Department of Electrical and Computer Engineering at Carnegie Mellon University in Pittsburgh, PA, USA, where he is working on security of cyber-physical systems.

Dr. Forti's main research interests include state estimation and control theory, information fusion, sensor networks and statistical machine learning, with special emphasis on inference, learning, and security over large-scale networked systems, multi-sensor multi-target tracking, and monitoring/control of distributed parameter systems.

Giorgio Battistelli received the Laurea degree in electronic engineering and the Ph.D. degree in robotics from the University of Genoa, Genoa, Italy, in 2000 and 2004, respectively. From 2004 to 2006, he was a Research Associate with the Dipartimento di Informatica, Sistemistica e Telematica, University of Genoa. Since 2006 he has been with the University of Florence, Florence, Italy, where he is currently an Associate Professor of automatic control with the Dipartimento di Ingegneria dell'Informazione. His current research interests include adaptive and learning systems, real-time control reconfiguration, linear and nonlinear estimation, hybrid systems, sensor networks, and data fusion.

Dr. Battistelli was a member of the editorial boards of the IFAC Journal Engineering Applications of Artificial Intelligence and of the IEEE Transactions on Neural Networks and Learning Systems. He is currently an Associate Editor of the IFAC Journal Nonlinear Analysis: Hybrid Systems, and a member of the conference editorial boards of IEEE Control Systems Society and the European Control Association.

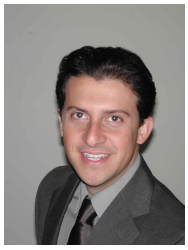
Luigi Chisci (SM'13) was born in Florence, Italy, in 1959. He received the M.S. degree in electrical engineering from the University of Florence, Florence, Italy, in 1984 and the Ph.D. in systems engineering from the University of Bologna, Bologna, Italy, in 1989. Currently he is a Full Professor of control engineering at the University of Florence since December 2004. His educational and research career have been in the area of control and systems engineering. His research interests have spanned over adaptive control and signal processing, algorithms and architectures for real-time control and signal processing, recursive identification, filtering and estimation, predictive control. He has co-authored over 170 papers of which over 60 on international journals. His current interests concern networked estimation, multi-target multi-sensor tracking, multi-agent systems and sensor data fusion.

Dr. Chisci served on the Conference Editorial Board of the IEEE Control Systems Society as an Associate Editor, from 2000 to 2008. He is currently Associate Editor of the Wiley's International Journal of Adaptive Control and Signal Processing.

Suqi Li was born in 1990. She received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, in 2011. Since September 2011, she has been pursuing the Ph.D. degree at the School of Electronic Engineering, University of Electronic Science and Technology of China. Currently, she is a visiting student with the Department of Information Engineering, University of Florence, Italy. Her research interests include random finite sets, multi-target tracking, nonlinear filtering.

Bailu Wang received his B.S. degree from the University of Electronic Science and Technology of China (UESTC) in 2011. He is now working toward his Ph.D. degree on signal and information processing at UESTC.

From August 2016, he has been a visiting student at University of Florence, Florence, Italy. His current research interests include radar and statistical signal processing, and multi-sensor multi-target fusion.



Bruno Sinopoli received the Dr. Eng. degree from the University of Padova in 1998 and his M.S. and Ph.D. in Electrical Engineering from the University of California at Berkeley, in 2003 and 2005 respectively. After a postdoctoral position at Stanford University, Dr. Sinopoli joined the faculty at Carnegie Mellon University where he is professor in the Department of Electrical and Computer Engineering with courtesy appointments in Mechanical Engineering and in the Robotics Institute and co-director of the Smart Infrastructure Institute, a research center aimed at advancing innovation in the modeling analysis and design of smart infrastructure. His research interests include the modeling, analysis and design of Secure by Design Cyber-Physical Systems with applications to Energy Systems, Interdependent Infrastructures and Internet of Things.

Dr. Sinopoli was awarded the 2006 Eli Jury Award for outstanding research achievement in the areas of systems, communications, control and signal processing at U.C. Berkeley, the 2010 George Tallman Ladd Research Award from Carnegie Mellon University and the NSF Career award in 2010.