Differentially-Private Distributed Fault Diagnosis for Large-Scale Nonlinear Uncertain Systems ⋆

Vahab Rostampour * **Riccardo Ferrari** *
André M.H. Teixeira ** **Tamás Keviczky** *

* *Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands. (e-mail: {v.rostampour, r.ferrari, t.keviczky}@tudelft.nl)*
** *Department of Engineering Sciences, Uppsala University, PO Box 534, SE-75121, Uppsala, Sweden. (e-mail: andre.teixeira@angstrom.uu.se)*

Abstract: Distributed fault diagnosis has been proposed as an effective technique for monitoring large scale, nonlinear and uncertain systems. It is based on the decomposition of the large scale system into a number of interconnected subsystems, each one monitored by a dedicated Local Fault Detector (LFD). Neighboring LFDs, in order to successfully account for subsystems interconnection, are thus required to communicate with each other some of the measurements from their subsystems. Anyway, such communication may expose private information of a given subsystem, such as its local input. To avoid this problem, we propose here to use differential privacy to pre-process data before transmission.

Keywords: Privacy Preserving, Differential Privacy, Distributed Fault Diagnosis, Uncertain Network of Nonlinear Systems.

1. INTRODUCTION

The problem of fault diagnosis and security for large scale nonlinear systems such as critical infrastructures or interconnected Cyber Physical Systems (CPS) have received increasing attention in the recent years (Kyriakides and Polycarpou (2014)). Indeed, one way to increase the resiliency of such systems to faults or deliberate cyber attacks is to endow them with architectures capable of monitoring, detecting, isolating and counteracting such anomalies and threats. Such systems being large scale, centralized monitoring and diagnosing architectures are rarely feasible, thus favoring distributed or decentralized ones. While decentralized solutions do not require communication between diagnosis nodes, they are not able to account for interconnection effects between different parts, or subsystems, of the large scale system being monitored. As this may lead to unacceptable performances, distributed methods, which instead do require communication, are thus preferable (Ferrari et al. (2012); Zhang and Zhang (2013); Zhang et al. (2013); Ge and Han (2014); Riverso et al. (2016); Noursadeghi and Raptis (2017)). One unexplored issue about the implementation of such distributed schemes, regards indeed the necessity of communication between neighbouring nodes. In the case where such nodes may be operated by different, possibly competing entities, mutual communication may be opposed as it may lead to leaking privacy-sensitive information. We may consider as an example a smart grid where neighbouring diagnosis nodes are each monitoring different subgrids with distributed energy sources and each is managed by its own grid operator. The two grid operators must exchange data about nodes on their respective boundaries in order to allow for grid balancing, but they would rather keep private the way that they are allocating energy supply to their different energy sources and satisfying their energy demand (Han et al. (2014b); Sankar et al. (2011)). A powerful and mathematically rigorous concept for dealing with privacy problems is *differential privacy*. This concept emerged in the Computer Science community (Dwork et al. (2006, 2014)), but recently found applications in Control Systems as well (see for instance Han et al. (2014b,a, 2017); Le Ny and Pappas (2014); Mo and Murray (2017)). It assumes that each piece of *user* data whose privacy must be protected is contained in a separate record in a database. A trusted party, called *curator*, maintains such database and answers queries posed by possibly adversarial, external parties. Differential privacy aims at modifying the query output to guarantee that no adversarial can guess whether a single record is present or has been altered, either by combining the results from several queries, or using side-channel information. In the previous example, the role of user data is taken by the local input applied to a subgrid, while the query corresponds to the communication of a subsystem boundary values to adversarial neighbours, such values being dependent on the subsystem local input

⋆ This research was supported by the Uncertainty Reduction in Smart Energy Systems (URSES) research program funded by the Dutch organization for scientific research (NWO) and Shell under the project Aquifer Thermal Energy Storage Smart Grids (ATES-SG) with grant number 408-13-030, and by the European Union H2020 program under the project "SURE: Safe Unmanned Robotic Ensembles" with grant number 707546.

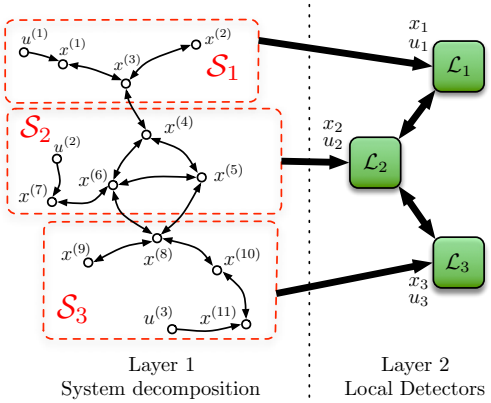

Fig. 1. The proposed distributed fault detection architecture. On the left side, the decomposition of the original system $\mathcal{S}_I$ is shown, where $I = 1, 2, 3$: thin black lines represent causal dependency between variables. On the right, the communication and the acquisition of measurements by the agents $\mathcal{L}_I$ is depicted, where $I = 1, 2, 3$.

sequence and the subsystem dynamics. The original and novel contribution of the present paper is the application of a differential privacy mechanism to the distributed fault diagnosis approach of Ferrari et al. (2012). In particular, Theorem 1 will provide a connection between the privacy level of the aforementioned subsystem boundary values and the privacy of its local inputs. The distributed diagnosis problem formulation, based on Local Fault Detectors (LFD) will be presented in Section 2, where we will extend existing results by considering a probabilistic detection threshold. In Section 3 we will introduce a privacy preserving mechanism to be applied to boundary data that neighbouring LFDs need to exchange. The paper will be completed by a numerical study in Section 4, showing the effectiveness of the proposed approach in the case of a multi-tank network simulated example, and some final remarks in Section 5.

2. PROBLEM STATEMENT

In this paper we will consider the case of a large-scale dynamical system $\mathcal{S}$, originating from the interconnection of N smaller subsystems $\mathcal{S}_I$, $I = 1, \dots, N$. Following Ferrari et al. (2012), we will allow each subsystem to be monitored by a dedicated agent $\mathcal{L}_I$, called *Local Fault Detector* (LFD), having access to locally available information, coming from measurements on its subsystem, and information from neighboring agents (see Fig. 1).

2.1 Large-scale System Dynamics

We will assume $\mathcal{S}$ to be described by the following nonlinear uncertain discrete time system

$$\begin{cases} x_{k+1} &= g(x_k, u_k, w_k, f_k) \\ y_k &= x_k + v_k \end{cases}, \tag{1}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$ and $y_k \in \mathbb{R}^n$ are the state, the input and the output of $\mathcal{S}$ at discrete time index $k \in \mathbb{N}$, respectively, while $g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^n$ models the state dynamics. The variable $w_k \in \mathbb{R}^p$, instead, represents unavoidable modeling uncertainties affecting eq. (1), while $f_k \in \mathcal{F} \subseteq \mathbb{R}^q$ represents a parametrization of the whole class of faults that can affect $\mathcal{S}$. Such formulation is purposely as general as possible, and comprises the cases where w_k and f_k affects the dynamics g as additive or multiplicative terms, or where they affect one or more parameters that appear in the definition of g: we conventionally assume anyway that for $w_k = 0$ and $f_k = 0$ the nominal and healthy behavior of $\mathcal{S}$, that is in the absence of uncertainties and faults, is obtained. Furthermore, g will be assumed to be differentiable and Lipschitz with respect to u, as detailed in Ass. 3. Finally, for the sake of simplicity here the full state is assumed to be available, up to a measurement uncertainty $v_k \in \mathbb{R}^n$: the extension to general input-output systems could be addressed similarly to Ferrari et al. (2008).

Assumption 1. No faults act on the system, that is $f_k = 0$, for $0 \leq k < k_f$, with k_f being the anomaly occurrence time. Moreover, the variables x_k and u_k remain bounded before and after the occurrence of an anomaly, i.e., there exist some stability regions $\mathcal{S} = \mathcal{S}^x \times \mathcal{S}^u \subset \mathbb{R}^n \times \mathbb{R}^m$, such that $(x_k, u_k) \in \mathcal{S}$, for all k. □

Assumption 2. w_k and v_k are random variables defined on some probability spaces $(\mathcal{W}, \mathfrak{B}(\mathcal{W}), \mathbb{P}_\mathcal{W})$, and $(\mathcal{V}, \mathfrak{B}(\mathcal{V}), \mathbb{P}_\mathcal{V})$, respectively, where $\mathcal{W} \subseteq \mathbb{R}^p$, $\mathcal{V} \subseteq \mathbb{R}^n$, $\mathfrak{B}(\cdot)$ denotes a Borel σ-algebra, and $\mathbb{P}_\mathcal{W}$, $\mathbb{P}_\mathcal{V}$ are probability measures defined over $\mathcal{W}$, $\mathcal{V}$, respectively. Furthermore, w_k and v_k are not correlated and are independent from x_k, u_k and f_k, $\forall\, k$. □

2.2 Sub-systems Dynamics

We assume that $\mathcal{S}$ can be described through a *non-overlapping decomposition* $\mathcal{D}$ into N subsystems $\mathcal{S}_I$, with $I \in \{1, \dots, N\}$, each defined via an *extraction index* n_I-tuple $\mathcal{I}_I$ (see Ferrari et al. (2012)). It is then possible to define a local state $x_{I,k} \in \mathbb{R}^{n_I}$, where $x_{I,k} \coloneqq \mathrm{col}(x_k^{(i)} : i = \mathcal{I}_I^{(j)}, j = 1, \dots, n_I)$, and similarly a local output $y_{I,k}$ and a local measurement uncertainty $v_{I,k}$. The local input $u_{I,k}$ is instead built with all the components of u_k that *structurally affect* at least one component of $x_{I,k+1}$, and similarly for building the local $w_{I,k}$ and $f_{I,k}$.

Definition 1. A variable c *structurally affects* a variable $a = b(c, d)$ through a multi-input function b, and is written $c \xrightarrow{b} a$, if there exists at least a pair of distinct values $\bar{c}$ and $\bar{c}'$ and a value $\bar{d}$ such that $\bar{a} = b(\bar{c}, \bar{d})$ is distinct from $\bar{a}' = b(\bar{c}', \bar{d})$.

Remark 1. It is important to stress that here we are only assuming that we have a structural knowledge of the effect of w_k and f_k on each component of g. This does not preclude the capability for our problem formulation to capture the case where the uncertainty, or the anomaly, are non parametric and arbitrary. For instance, we could assume in this case the dynamics to be decomposable as $g(x_k, u_k, w_k, f_k) = g^\star(x_k, u_k) + w_k + f_k$, where $g^\star$ represents the nominal dynamics, and w_k and f_k are arbitrarily varying signals, but respecting Assumptions 1 and 2.

We can proceed further and describe the dynamics of the generic subsystem $\mathcal{S}_I$ as

$$\begin{cases} x_{I,k+1} &= g_I(x_{I,k}, u_{I,k}, x_{\mathcal{N}_I,k}, w_{I,k}, f_{I,k}) \\ y_{I,k} &= x_{I,k} + v_{I,k} \end{cases}, \quad (2)$$

where the *local dynamics function* $g_I : \mathbb{R}^{n_I} \times \mathbb{R}^{m_I} \times \mathbb{R}^{n_{\mathcal{N}_I}} \times \mathbb{R}^{p_I} \times \mathbb{R}^{q_I} \mapsto \mathbb{R}^{n_I}$ can be simply obtained by taking in the right order the components of g that are contained in the index tuple $\mathcal{I}_I$. In general we cannot assume that all the resulting subsystems $\mathcal{S}_I$ are decentralized, i.e. their dynamics depend only on the local state x_I, therefore we introduced the *interconnection variable* $x_{\mathcal{N}_I,k}$ as in Ferrari et al. (2012)

Definition 2. The *interconnection variable* $x_{\mathcal{N}_I,k} \in \mathbb{R}^{n_{\mathcal{N}_I}}$ of the subsystem $\mathcal{S}_I$ is the vector $x_{\mathcal{N}_I,k} \coloneqq \mathrm{col}(x_k^{(j)} : x_k^{(j)} \xrightarrow{g} x_{I,k+1}^{(i)},\ i \in \{1,\dots,n_I\},\ j \in \{1,\dots,n\})$.

The role of $x_{\mathcal{N}_I,k}$ is to describe the functional dependence of the local dynamics g_I on state components from other subsystems, which we will call *neighboring subsystems* or simply *neighbors*. The set of all the neighbors of $\mathcal{S}_I$ will be denoted by $\mathcal{N}_I$.

Remark 2. As Assumption 1 holds for the original system $\mathcal{S}$, then it will continue to do so for every subsystem and we can introduce a stability region $\mathcal{S}_I$ for each one, where the local state x_I and input u_I are assumed to always belong. Similarly, we can easily build the domains $\mathcal{V}_I$, $\mathcal{W}_I$, $\mathcal{F}_I$ and $\mathcal{V}_{\mathcal{N}_I}$ of, respectively: the local measurement and modeling uncertainties, the local fault parameters, and the measurement uncertainties of the interconnection variable.

2.3 Residual Generator

For fault detection purpose each LFD $\mathcal{L}_I$ shall compute a residual $r_{I,k} \coloneqq y_{I,k} - \hat{y}_{I,k}$ and compare it to a dynamic detection threshold. In this subsection the residual will be addressed. As a direct extension of Rostampour et al. (2017), it shall be obtained as the output estimation error of the following nonlinear estimator

$$\begin{cases} \hat{x}_{I,k+1} &= g_I(y_{I,k}, u_{I,k}, y_{\mathcal{N}_I,k}, 0, 0) + \Lambda(\hat{y}_{I,k} - y_{I,k}) \\ \hat{y}_{I,k} &= \hat{x}_{I,k} \end{cases}, \quad (3)$$

where $\hat{x}_I$, $\hat{y}_I \in \mathbb{R}^{n_I}$ are, respectively, the local state and output estimates, $y_{\mathcal{N}_I,k} \in \mathbb{R}^{n_{\mathcal{N}_I}}$ are the measurements of the interconnection variables $x_{\mathcal{N}_I,k}$, $\Lambda \triangleq \mathrm{diag}(\lambda^i,\ i = 1 \dots n_I)$ is a diagonal matrix, and $\lambda^i \in (0,1)$ denotes some filtering parameters chosen to guarantee the stability of the estimator.

By using eqs. (1) and (3), we can then write the residual dynamics as

$$r_{I,k+1} = \Lambda\, r_{I,k} + \delta_{I,k}\ , \quad (4)$$

where the *total uncertainty* $\delta_{I,k}$ is a stochastic process representing the uncertain part of the residual dynamics:

$$\begin{aligned} \delta_{I,k} \coloneqq\ & g_I(x_{I,k}, u_{I,k}, x_{\mathcal{N}_I,k}, w_{I,k}, f_{I,k}) \\ & - g(y_{I,k}, u_{I,k}, y_{\mathcal{N}_I,k}, 0, 0) + v_{I,k+1} \\ =\ & g_I(y_{I,k} - v_{I,k}, u_{I,k}, y_{\mathcal{N}_I,k} - v_{\mathcal{N}_I,k}, w_{I,k}, f_{I,k}) \\ & - g(y_{I,k}, u_{I,k}, y_{\mathcal{N}_I,k}, 0, 0) + v_{I,k+1}\ . \end{aligned} \quad (5)$$

Thanks to Ass. 1, 2 and eq. (5), it follows that, given $y_{I,k}$, $u_{I,k}$, and $y_{\mathcal{N}_I,k}$, $\delta_{I,k}$ is a conditioned random variable on a probability space $(\Delta_{I,k}, \mathfrak{B}(\Delta_{I,k}), \mathbb{P}_{\Delta_{I,k}})$, where $\Delta_{I,k}$ is a time varying set defined as follows.

Definition 3. The time varying *total uncertainty set* $\Delta_{I,k} \subset \mathbb{R}^{n_I}$ at time index k is defined as

$$\begin{aligned} \Delta_{I,k} \coloneqq \{\delta_{I,k} \,|\, & y_{I,k},\ y_{\mathcal{N}_I,k},\ u_{I,k},\ w_{I,k} \in \mathcal{W}_I,\ f_{I,k} \in \mathcal{F}_I, \\ & v_{I,k} \in \mathcal{V}_I, v_{I,k+1} \in \mathcal{V}_I,\ v_{\mathcal{N}_I,k} \in \mathcal{V}_{\mathcal{N}_I}\}\,, \end{aligned}$$

with $\delta_{I,k}$ being computed according to (5).

As a special case of Definition 3, we introduce the uncertainty set corresponding to a healthy plant as follows.

Definition 4. The time varying *healthy total uncertainty set* $\Delta^0_{I,k} \subset \mathbb{R}^{n_I}$ at time index k is defined as

$$\begin{aligned} \Delta^0_{I,k} \coloneqq \{\delta_{I,k} \,|\, & w_{I,k} \in \mathcal{W}_I,\ f_{I,k} \in \{0\},\ v_{I,k} \in \mathcal{V}_I, \\ & v_{I,k+1} \in \mathcal{V}_I,\ v_{\mathcal{N}_I,k} \in \mathcal{V}_{\mathcal{N}_I}\}\,, \end{aligned}$$

where $\delta_{I,k}$ is computed according to (5).

Remark 3. The role of $\Delta_{I,k}$ and $\Delta^0_{I,k}$ is to quantify the range of possible values that $\delta_{I,k}$ can take, respectively, corresponding to situations when a fault *may* be present, and in a healthy situations where a fault is absent. Apart from simple cases, no closed form is available for computing such sets, and numerical approximations techniques such as those in Dabbene et al. (2015) may be used.

We can now introduce a compact notation for the *residual generator* described by eqs. (3),(4),(5), through a mapping function $\Sigma_I : \mathbb{R}^{n_I} \times \mathbb{R}^{n_I} \mapsto \mathbb{R}^{n_I}$ defined as

$$r_{I,k+1} = \Sigma_I(r_{I,k}, \delta_{I,k}) \coloneqq \Lambda\, r_{I,k} + \delta_{I,k}\,. \quad (6)$$

Remark 4. While at time index k the residual $r_{I,k}$ can be computed from y_I and $\hat{y}_k$ and is thus a deterministic quantity, from (4), (6) it follows that the next value $r_{I,k+1}$ is a random variable on the same probability space as $\delta_{I,k}$.

Given these preliminaries, it is now possible to write the following two fundamental definitions (see Fig. 2).

Definition 5. The time varying *residual set* $\mathcal{R}_{I,k+1}$ at time index $k+1$ is defined as the image of the set $\Delta_{I,k}$ through Σ_I, that is

$$\begin{aligned} \mathcal{R}_{I,k+1} \coloneqq & \Sigma_I(r_{I,k}, \Delta_{I,k}) \\ = & \{r_{I,k+1} \,|\, r_{I,k+1} = \Sigma_I(r_{I,k}, \delta_I),\ \delta_I \in \Delta_{I,k}\}. \end{aligned}$$

Definition 6. The time varying *healthy residual set* $\mathcal{R}^0_{I,k+1}$ at time index $k+1$ is defined as the image of the set $\Delta^0_{I,k}$ through Σ_I, that is

$$\begin{aligned} \mathcal{R}^0_{I,k+1} \coloneqq & \Sigma_I(r_{I,k}, \Delta^0_{I,k}) \\ = & \{r_{I,k+1} \,|\, r_{I,k+1} = \Sigma_I(r_{I,k}, \delta_I),\ \delta_I \in \Delta^0_{I,k}\}. \end{aligned}$$

For ease of notation, when there is no ambiguity, in the rest of the paper we will drop the index I to denote that a quantity refers to the generic subsystem $\mathcal{S}_I$ or the generic LD agent $\mathcal{L}_I$. The index $\mathcal{N}$ will be retained to indicate the neighbor set of the generic subsystem or agent.

2.4 Fault Detection Threshold Design Problem

In order to reduce the detrimental effects on fault detectability of deterministic thresholds, which in practice can be overly conservative, in this paper we will seek a probabilistically robust threshold instead (see Rostampour et al. (2017)). In particular, by extending Boem et al.

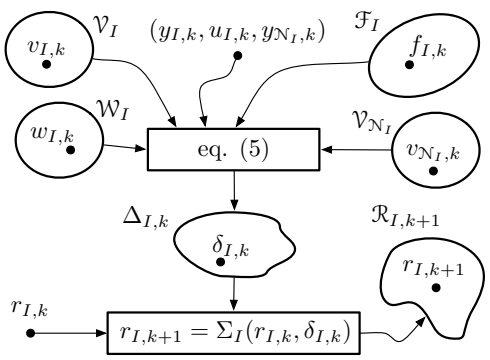

Fig. 2. The residual set $\mathcal{R}_{I,k+1}$ can be thought of as the image obtained by computing the output Σ_I by letting $\delta_{I,k}$ vary over its domain $\Delta_{I,k}$ and fixing the residual $r_{I,k}$ to its actual value. The domain $\Delta_{I,k}$ in turn is computed through eq. (5) by letting $v_{I,k}$, $w_{I,k}$, $f_{I,k}$ and $v_{\mathcal{N}_I,k}$ vary over their respective domains, and fixing the local output and input $y_{I,k}$ and $u_{I,k}$, as well as the interconnection variables measurement $y_{\mathcal{N}_I,k}$, to their actual values. The healthy residual set $\mathcal{R}^0_{I,k+1}$ can be obtained similarly, but by fixing the value $f_{I,k} \in \{0\}$.

(2015) and Ferrari et al. (2017), we propose the following residual evaluation logic and threshold for fault detection

$$d_M(r_{I,k+1}) \leq \bar{d}_M \triangleq \frac{n_I}{\alpha} \Rightarrow \mathcal{S}_I \text{ is healthy} \tag{7}$$

based on the Mahalanobis distance of the residual

$$d_M(r_I) \triangleq \sqrt{(r_I - \mu_{r_I})' C_{r_I}^{-1} (r_I - \mu_{r_I})} \tag{8}$$

where $\mu_{r_I} \triangleq \mathbb{E}[r_I] \in \mathbb{R}^{n_I}$ and $C_{r_I} \triangleq \mathrm{Cov}[r_I] \in \mathbb{R}^{n_I \times n_I}$ are the expected value and covariance matrix of the random variable r_I. Indeed, thanks to the Multivariate Chebyshev Inequality (see Chen (2007)), we can bound the probability of false positives during healthy conditions as

$$\mathbb{P}\left[d_M(r_I) > \bar{d}_M\right] < 1 - \alpha\,, \tag{9}$$

where $\alpha \in (0,\, 1]$ is a user defined constant representing the desired probabilistic robustness of the threshold $\bar{d}_M$ and $\mathbb{P}$ denotes probability.

Remark 5. While the detection logic (7) employs a static threshold, it must be noted that it is equivalent to testing whether $r_{I,k+1}$ belongs to a time varying ellipsoid centered in $\mu_{r_{I,k+1}}$ and whose orientation and size are described by $C_{r_{I,k+1}}$. Indeed the moments of $r_{I,k+1}$ depend via the mapping Σ_I on those of $\delta_{I,k}$, which are not assumed to be time invariant unless g_I is a linear function and v and w are stationary processes. In such case and also when g_I is bilinear the techniques proposed in Ferrari et al. (2017) can be employed.

Remark 6. For the general nonlinear case we will assume that the moments can be approximated by their corresponding sample moments by generating a number N of samples $r^j_{I,k+1}$, with $j = 1 \dots N$, as in Boem et al. (2015) and Rostampour et al. (2017).

3. DIFFERENTIALLY PRIVATE FRAMEWORK

The distributed fault detection scheme outlined in the previous section requires every agent $\mathcal{L}_I$ to have the following quantities communicated by its neighbours: their output components appearing in $y_{\mathcal{N}}$, which are needed to update the nonlinear estimator (3) and thus generate the residual r; and N samples of the measurement uncertainties appearing in $v_{\mathcal{N}}$, such that the sample moments of r can be computed by repeated evaluation of eqs. (4) and (5), and used to evaluate the residual according to (7).

The goal of this section is indeed to show how, relying on concepts from *Differential Privacy* (DP), an agent $\mathcal{L}_I$ and its neighbours may communicate with each other without exposing private information on their local input. The next subsection will introduce the basics of DP.

3.1 The Concept of Differential Privacy

Differential privacy "addresses the paradox of learning nothing about an individual while learning useful information about a population" (Dwork et al. (2014)). The initial concern that drove its development is in fact protecting the privacy of human individuals, for instance when personal health data is collected and used in medical studies.

As a preliminary notion, we need to introduce the concepts of *database* and of *query*.

Definition 7. A *database* D of length n is a set $D = \{d_1, d_2, \cdots, d_n\}$ taking values in $\mathcal{D}$, where $\mathcal{D}$ is the *universe* of all possible databases.

Definition 8. A *query* q is a mapping $q : \mathcal{D} \to \mathbb{R}^{n_q}$, where n_q is the size of the result provided by the query.

In DP it is assumed that data contained in a database D can be accessed only through the results of queries, which are answered by the subject holding D, called *curator*. Protecting the privacy of an element d_i in D can thus be obtained by making the results of any query run on D insensitive enough to the single d_i. This can also be expressed by ensuring that two adjacent databases are nearly indistinguishable from the answers to a query.

Definition 9. (Han et al. (2017)) Two databases $D = \{d_1, \cdots, d_n\}$ and $D' = \{d'_1, \cdots, d'_n\}$ are said to be *adjacent*, and it is written as $\mathrm{adj}(D, D')$, if there exists $i \in \{1, \cdots, n\}$ such that $d_j = d'_j$ for all $j \neq i$.

This is enforced by introducing so-called *mechanisms*, which are randomized mappings from the universe $\mathcal{D}$ to some subset in $\mathbb{R}^{n_q}$, and letting the curator use the mechanism in lieu of the query. A mechanism that acts on a database is said to be *differentially private* if it complies with the following definition from Dwork et al. (2006).

Definition 10. Given $\epsilon \geq 0$ as the desired level of privacy, a mechanism M preserves $\epsilon-$differential privacy if for all $\mathcal{R} \subset \mathrm{range}(M)$ and all adjacent databases D and D' in $\mathcal{D}$, it holds that

$$\mathbb{P}\left[M(D) \in \mathcal{R}\right] \leq e^{\epsilon}\, \mathbb{P}\left[M(D') \in \mathcal{R}\right]\ . \tag{10}$$

Remark 7. A smaller ϵ implies higher level of privacy. By using differential privacy, one can hide information at the individual level, no matter what side information others may have. Definition 10 shows that DP is based on randomization, but is independent on the contents of databases, as long as they belong to $\mathcal{D}$ and are adjacent.

A popular mechanism in the DP literature is the so-called Laplace mechanism, that introduces a Laplacian additive noise dependent on the query ℓ_p-sensitivity

Definition 11. (Han et al., 2017, Definition 10) For any query $q : \mathcal{D} \to \mathbb{R}^{n_q}$, the ℓ_p-sensitivity of q under the adjacency relation, adj, is defined as

$$\sigma := \max\{\|q(D) - q(D')\|_p : D, D' \in \mathcal{D} \text{ s.t. } \text{adj}(D, D')\}\,.$$

It is worth mentioning that ℓ_p-sensitivity of q does not depend on a specific database D. We now recall the following results from (Han et al., 2014a, Theorem 9).

Proposition 1. Consider a query $q : \mathcal{D} \to \mathbb{R}^{n_q}$ whose ℓ_2-sensitivity is σ. Define the mechanism M as $M(D) = q(D) + \nu$, where $\nu \in \mathbb{R}^{n_q}$ is a random vector whose probability density function is given by $p_\nu(\nu) \propto \exp(-\epsilon\|\nu\|/\sigma)$. Then the mechanism M preserves $\epsilon-$differential privacy.

3.2 Privacy-Preserving Mechanism

The proposed privacy-preserving framework for distributed fault detection will be now presented. To simplify the notation and formulation, we will assume without loss of generality the case of a given agent $\mathcal{L}$ having a single neighbor $\mathcal{L}_\mathcal{N}$, connected through an interconnection variable $x_\mathcal{N} \in \mathbb{R}^{n_\mathcal{N}}$. We will also drop the time indexes to simplify our notation whenever possible. As said previously, $\mathcal{L}_\mathcal{N}$ should send to $\mathcal{L}$ at each time indexes its last interconnection variable measurement $y_\mathcal{N}$. From the point of view of the DP formulation, agent $\mathcal{L}_\mathcal{N}$ is the curator of a database that contains the local input $u_{\mathcal{N},k-1}$, and that at time k is answering a query from $\mathcal{L}$ by providing the measurement $y_\mathcal{N}$, which depends on the previous local state of $\mathcal{L}_\mathcal{N}$ and on $u_{\mathcal{N},k-1}$ via its dynamics (2). In general it does not hold that $u_\mathcal{N}$ can be reconstructed from values of $y_\mathcal{N}$. Anyway, in the DP setting a privacy breach does not require the capability of fully reconstructing a piece of information, but only the capability of determining whether it will cause the query result to belong or not to a given set (Def. 10). This, in turns, depends on the query sensitivity (Def. 11). For these reasons, $\mathcal{L}_\mathcal{N}$ does want to replace such answer with a mechanism that guarantees the privacy of $u_\mathcal{N}$.

Before proceeding further, we need an extended definition of adjacency.

Definition 12. Two control actions $u_\mathcal{N}$, $u'_\mathcal{N} \in \mathcal{U} \subset \mathbb{R}^{m_\mathcal{N}}$ are two adjacent control inputs at time step $k-1$ if and only if $\|u_\mathcal{N} - u'_\mathcal{N}\|_0 \leq 1$, and it is written $\text{adj}(u_\mathcal{N}, u'_\mathcal{N})$. Such a distance between databases is referred to as the Hamming distance, i.e., the number of rows on which they differ. The set $\mathcal{U}$ is a compact set over which the input sequence $\{u_{\mathcal{N},k}\}_{k=0}^{\infty}$ can take values.

Remark 8. Following Defin. 12, we can say that two adjacent control inputs belong to a bounded set $\mathcal{U}$ such that:

$$\max_{i\in\{1,\dots,m_\mathcal{N}\}} |(u_\mathcal{N})^{(i)} - (u'_\mathcal{N})^{(i)}| \leq 2\zeta\,,$$

where $\zeta \geq 0$ is a positive constant number which depends on the set $\mathcal{U}$.

Since the query $q(\cdot)$ answered by $\mathcal{L}_\mathcal{N}$ is actually the output of the generic subsystem $\mathcal{S}_\mathcal{N}$, the constant σ that appears in Definition 11 can be computed as

$$\sigma_{\mathcal{N}_u} = \max_{\substack{u_\mathcal{N}, u'_\mathcal{N}\in\mathcal{U} \\ \text{adj}(u_\mathcal{N}, u'_\mathcal{N}) \\ \psi_\mathcal{N}\in\Psi}} \|g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N}) - g_\mathcal{N}(\psi_\mathcal{N}, u'_\mathcal{N})\|_p\,, \quad (11)$$

where $g_\mathcal{N}(\psi_{\mathcal{N},k-1}, u_{\mathcal{N},k-1}) := y_{\mathcal{N},k}$ represents a compact notation for $\mathcal{S}_\mathcal{N}$ dynamics in (2). The new quantity $\psi_\mathcal{N} \in \Psi$ represents the other variables, apart from the input $u_\mathcal{N}$, which influence $\mathcal{S}_\mathcal{N}$, and is defined as $\psi_\mathcal{N} := \text{col}(x, x_\mathcal{N}, w, f)$, with $\Psi := \mathcal{S}^x \times \mathcal{S}^{x_\mathcal{N}} \times \mathcal{W} \times \mathcal{F}$. The bound $\sigma_{\mathcal{N}_u}$ can be seen as a bound on the global ℓ_p-sensitivity of the mapping function $g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N})$ with respect to the control input $u_\mathcal{N}$ at each time step k for all $p \geq 1$. The following assumption is needed to compute $\sigma_{\mathcal{N}_u}$.

Assumption 3. The nonlinear dynamics function $g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N})$ of the generic subsystem $\mathcal{S}_\mathcal{N}$ is measurable and differentiable in $u_\mathcal{N}$ such that at each sampling time k

$$\frac{\partial g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N})}{\partial u_\mathcal{N}} \neq 0\ , \forall u_\mathcal{N} \in \mathcal{U}\ , \psi_\mathcal{N} \in \Psi\ ,$$

and there exists a constant L for all time step k, $u_\mathcal{N}$, $u'_\mathcal{N} \in \mathcal{U}$ and $\psi_\mathcal{N} \in \Psi$ such that:

$$\begin{aligned}\|g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N}) - g_\mathcal{N}(\psi_\mathcal{N}, u'_\mathcal{N})\| &\leq L\|\varphi_\mathcal{N} - \varphi'_\mathcal{N}\| \\ &= L\|u_\mathcal{N} - u'_\mathcal{N}\|\,,\end{aligned} \quad (12)$$

where $\varphi_\mathcal{N}$ and $\varphi'_\mathcal{N}$ are two vectors obtained by concatenating $\psi_\mathcal{N}$ with $u_\mathcal{N}$ and $u'_\mathcal{N}$, respectively. We refer to L as the Lipschitz constant of the nonlinear function $g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N})$ of the generic subsystem $\mathcal{S}_\mathcal{N}$.

Remark 9. An essential factor is the differentiability of $g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N})$ in order to derive the sensitivity of the output signal with respect to small variations (adjacent relations) of input control signals. The key assumption is the Lipschitz condition (12). An approximation of the Lipschitz constant L at time step k can be calculated from eq. (2) using the available values of $\psi_\mathcal{N} \in \Psi$ and drawing a sufficiently high number of samples of the uncertainties $v_\mathcal{N}$ and $w_\mathcal{N}$, following a Monte Carlo approach.

Proposition 2. The global ℓ_2-sensitivity of the output of the generic subsystem $\mathcal{S}_\mathcal{N}$ is bounded by $\sigma_{\mathcal{N}_u} \leq 2\zeta L$.

Proof. Following Defin. 12 and Rem. 8 together with Ass. 3, the proof is straightforward by making use of eqs. (11) and (12), from which we can derive the inequality

$$\begin{aligned}\sigma_{\mathcal{N}_u} &\leq \max_{\substack{u_\mathcal{N}, u'_\mathcal{N}\in\mathcal{U} \\ \text{adj}(u_\mathcal{N}, u'_\mathcal{N})}} L\|u_\mathcal{N} - u'_\mathcal{N}\| \\ &= \max_{\substack{u_\mathcal{N}, u'_\mathcal{N}\in\mathcal{U} \\ \text{adj}(u_\mathcal{N}, u'_\mathcal{N})}} L \max_{i\in\{1,\dots,m_\mathcal{N}\}} |(u_\mathcal{N})^{(i)} - (u'_\mathcal{N})^{(i)}| \leq 2\zeta L\,.\end{aligned}$$

The proof is completed. □

We are now ready to state the problem that we are going to address in the present section.

Problem 1. Find a randomized mechanism M_u such that it preserves ϵ_u-differential privacy for the neighboring agent $\mathcal{L}_\mathcal{N}$ under the adjacency relation described in Definition 12.

Proposition 3. The mechanism $M_u(u_\mathcal{N}) = g_\mathcal{N}(\psi_\mathcal{N}, u_\mathcal{N}) + \nu_{u_\mathcal{N}}$, where $u_\mathcal{N}$ is the control input signal and $\nu_{u_\mathcal{N}} \in \mathbb{R}^{n_\mathcal{N}}$ is a noisy vector drawn from a probability density function that is proportional to $\exp(-\epsilon_u\|\nu_{u_\mathcal{N}}\|/2\zeta L)$, is ϵ_u-differentially private.

Proof. The proof is the direct result of combining Proposition 2 with Proposition 1. □

Output Signal as Database By looking at mechanism M_u in Proposition 3 it can be seen that it is equivalent to a mechanism M_y acting on a database containing $y_{\mathcal{N}}$, where the query is an identity. Indeed this equivalence can be easily shown. We first introduce the following

Definition 13. Two output signals $y_{\mathcal{N}}$, $y'_{\mathcal{N}} \in \mathcal{Y} \subset \mathbb{R}^{n_{\mathcal{N}}}$ are two adjacent output signals if and only if $\|y_{\mathcal{N}} - y'_{\mathcal{N}}\|_0 \leq 1$, and it is written as $\text{adj}(y_{\mathcal{N}}, y'_{\mathcal{N}})$. The set $\mathcal{Y}$ is a compact set over which the output sequence $\{y_{\mathcal{N},k}\}_{k=0}^{\infty}$ can take values, and since two output signals belong to $\mathcal{Y}$, we can have:

$$\max_{i \in \{1,\ldots,n_{\mathcal{N}}\}} |(y_{\mathcal{N}})^{(i)} - (y'_{\mathcal{N}})^{(i)}| \leq 2\xi\,,$$

where $\xi \geq 0$ is a positive constant number which depends on the set $\mathcal{Y}$.

Since the query is an identity mapping, a bound $\sigma_{\mathcal{N}_y}$ on the global ℓ_2- sensitivity of such a query can be obtained from:

$$\sigma_{\mathcal{N}_y} = \max_{\substack{y_{\mathcal{N}}, y'_{\mathcal{N}} \in \mathcal{Y} \\ \text{adj}(y_{\mathcal{N}}, y'_{\mathcal{N}})}} \|y_{\mathcal{N}} - y'_{\mathcal{N}}\| \leq 2\xi\,. \tag{13}$$

The following proposition provides a randomized mechanism M_y such that it preserves ϵ_y-differential privacy for the agent $\mathcal{L}_{\mathcal{N}}$ under the adjacency relation of Definition 13.

Proposition 4. The mechanism $M_y(y_{\mathcal{N}}) = y_{\mathcal{N}} + \nu_{y_{\mathcal{N}}}$, where $y_{\mathcal{N}}$ is the output signal and $\nu_{y_{\mathcal{N}}} \in \mathbb{R}^{n_{\mathcal{N}}}$ is a noisy vector drawn from a probability density function that is proportional to $\exp(-\epsilon_y \|\nu_{y_{\mathcal{N}}}\|/2\xi)$, is ϵ_y-differentially private.

Proof 1. It directly results from combining Proposition 2 with Proposition 1.

We next provide a theoretical connection between $M_u(u_{\mathcal{N}})$ and $M_y(y_{\mathcal{N}})$.

Theorem 1. Let $M_u(u_{\mathcal{N}})$ and $M_y(y_{\mathcal{N}})$ be the two randomized mechanisms introduced in Propositions 3 and Proposition 4 for a generic nonlinear system dynamics $\mathcal{S}_{\mathcal{N}}$ such that they preserve ϵ_u and ϵ_y level of differential privacy with $\epsilon_y = \epsilon_u \frac{\xi}{\zeta L}$, respectively. Given ζ in Remark 8 and ξ in Definition 13 with L in Assumption 3, if $\xi \leq \zeta L$, then,

$$\frac{\mathbb{P}[M_u(u_{\mathcal{N}}) \in \mathcal{R}_u]}{\mathbb{P}[M_u(u'_{\mathcal{N}}) \in \mathcal{R}_u]} = \frac{\mathbb{P}[M_y(y_{\mathcal{N}}) \in \mathcal{R}_y]}{\mathbb{P}[M_y(y'_{\mathcal{N}}) \in \mathcal{R}_y]} \leq e^{\epsilon_y} \leq e^{\epsilon_u}\,.$$

Proof. Following Propos. 3 together with Propos. 4, let $p_{u_{\mathcal{N}}}$ and $p_{u'_{\mathcal{N}}}$ denote the probability density function of $M_u(u_{\mathcal{N}})$ and $M_u(u'_{\mathcal{N}})$, respectively, and let $p_{y_{\mathcal{N}}}$ and $p_{y'_{\mathcal{N}}}$ denote the probability density function of $M_y(y_{\mathcal{N}})$ and $M_y(y'_{\mathcal{N}})$, respectively. We now compare $p_{u_{\mathcal{N}}}$ and $p_{u'_{\mathcal{N}}}$ at some arbitrary point $z \in \mathbb{R}^{n_{\mathcal{N}}}$ in order to show the first inequality in the above assertion as follows:

$$\begin{aligned}
\frac{p_{u_{\mathcal{N}}}(z)}{p_{u'_{\mathcal{N}}}(z)} &= \frac{\exp\left(\frac{-\epsilon_u \|g_{\mathcal{N}}(\psi_{\mathcal{N}}, u_{\mathcal{N}}) - z\|}{2\zeta L}\right)}{\exp\left(\frac{-\epsilon_u \|g_{\mathcal{N}}(\psi_{\mathcal{N}}, u'_{\mathcal{N}}) - z\|}{2\zeta L}\right)} \\
&= \frac{\exp\left(\frac{-\epsilon_u \|y_{\mathcal{N}} - z\|}{2\zeta L}\right)}{\exp\left(\frac{-\epsilon_u \|y'_{\mathcal{N}} - z\|}{2\zeta L}\right)} \\
&= \frac{\exp\left(\frac{-\epsilon_y \|y_{\mathcal{N}} - z\|}{2\xi}\right)}{\exp\left(\frac{-\epsilon_y \|y'_{\mathcal{N}} - z\|}{2\xi}\right)} = \frac{p_{y_{\mathcal{N}}}(z)}{p_{y'_{\mathcal{N}}}(z)}
\end{aligned}$$

where the second equality follows from choosing $\epsilon_y = \frac{\xi \epsilon_u}{\zeta L}$. Observe that $\epsilon_y \leq \epsilon_u$ holds for $\xi \leq \zeta L$. The rest of the proof follows the same steps as in (Dwork et al., 2014, Theorem 3.6):

$$\begin{aligned}
\frac{p_{y_{\mathcal{N}}}(z)}{p_{y'_{\mathcal{N}}}(z)} &= \frac{\exp\left(\frac{-\epsilon_y \|y_{\mathcal{N}} - z\|}{2\xi}\right)}{\exp\left(\frac{-\epsilon_y \|y'_{\mathcal{N}} - z\|}{2\xi}\right)} \\
&= \exp\left(\frac{-\epsilon_y \left(\|y_{\mathcal{N}} - z\| - \|y'_{\mathcal{N}} - z\|\right)}{2\xi}\right) \\
&\leq \exp\left(\frac{\epsilon_y \left(\|y'_{\mathcal{N}} - y_{\mathcal{N}}\|\right)}{2\xi}\right) \\
&\leq \exp\left(\epsilon_y\right) \\
&\leq \exp\left(\epsilon_u\right)\,,
\end{aligned}$$

where the first inequality follows from the inverse triangle inequality, the second follows from the definition of sensitivity and the last is due to $\xi \leq \zeta L$.

It is important to highlight that Theorem 1 is the first result, to the best of our knowledge, toward privatizing a desired database, e.g. the control input actions, using another database, e.g. the output signals of a generic nonlinear system dynamics $\mathcal{S}_{\mathcal{N}}$. Theorem 1 provides a theoretical link between two randomized mechanisms $M_u(u_{\mathcal{N}})$ in Proposition 3 and $M_y(y_{\mathcal{N}})$ in Proposition 4. Strictly speaking, one can consider the output signals of a generic dynamical system $\mathcal{S}_{\mathcal{N}}$ as a database to develop a randomized mechanism $M_y(y_{\mathcal{N}})$ such that it preserves ϵ_y-differential privacy together with achieving the ϵ_u-differential privacy of the input control signals as the main desired privacy goal by considering that $\epsilon_y = \frac{\xi \epsilon_u}{\zeta L}$ and $\xi \leq \zeta L$.

4. NUMERICAL STUDY

In this section we are going to present the results of a numerical study, in order to illustrate the effectiveness of the proposed approach. The system under study will be a multi-tank system (see Ferrari et al. (2012) for details on modeling such a system), whose structural graph contains 22 nodes, each representing a state variable corresponding to the level of a tank, while edges represent pipes interconnecting such tanks (Fig. 3). The graph has been obtained by application of the Barabási-Albert model Albert and Barabási (2002), which as known leads to scale-free networks. After labeling the nodes according to their degree, in descending order, two subsystems have been obtained by defining the extraction index tuples $\mathcal{I}_1 = [1, 2, 6, 8, 9, 13, 16, 17, 19, 21, 22]$ and $\mathcal{I}_2 = [3, 4, 5, 7, 10, 11, 12, 14, 15, 18, 20]$ of 11 elements each. Finally, in order to make the interconnection between

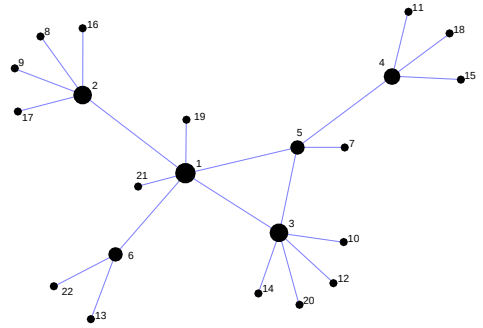

Fig. 3. The structural graph of the 22-tanks system chosen for the numerical study. It will be decomposed into two subsysystems: one comprising node 1 and all the nodes to its left; the other comprising the remaining nodes on its right. The interconnection between the two susbsytems is represented by the two edges $(1,3)$ and $(1,5)$, corresponding to two pipes.

the two resulting subsystems asymmetric and thus more interesting, an edge between nodes 1 and 3 has been added, on top of the edges produced by the Barabási-Albert algorithm. The actual tanks' cross section have been chosen equal to $1\,\mathrm{m}^2$, while pipes' cross sections equal to $0.2\,\mathrm{m}^2$. Drains with the same section as interconnecting pipes have been assumed to be connected to terminal nodes (i.e. nodes with unitary degree). A single source pump, with a sinusoidal time profile with a frequency of $0.1\,\mathrm{Hz}$, has been connected to tank no. 1. All tank levels are assumed to be measured, with a gaussian measurement uncertainty with zero mean and a standard deviation equal to $0.15\,\mathrm{m}$. When building the LFD estimators, a gaussian parametric uncertainty is introduced, having zero mean and a variance equal to 5% and 1.5%, respectively, of the tanks and pipes cross sections. The privacy mechanism M_u will be used, with the value $\zeta = 0.01$; a number $N = N_s = 512$ of samples is used by each LFD to compute the moments μ_r and C_r appearing in the Mahalanobis distance definition (8) and for generating the set $\mathcal{X}_\mathcal{N}$ that is communicated to neighbouring LFDs.

The fault that is presented in the current study represents a clogging in the pipe between tanks 1 and 3, reducing its flow to 50% of its nominal value. The reason we have chosen this kind of fault is that it affects exactly each subsystem interconnection variable, and as such may be hidden, that is made undetectable, by the introduction of the privacy mechanism. The following figures present the results obtained by simulating such fault occurring at time $T_f = 125\,\mathrm{s}$. In particular, the effect on detection performance of varying privacy levels, that is of a varying parameter ϵ, were analyzed.

Fig. 4 shows the time behaviour of tank no. 3, which is allocated to LFD no. 2 and is an interconnection variable for LFD no. 1. It can be seen how, for the smallest value of ϵ considered in this study ($\epsilon = 0.04$), the privatized version of its level can be dramatically different from the real one. Nevertheless, its estimated value by LFD no. 2 can still be relatively close to the real one, and clearly show its sensitivity to the fault happening at 125 s. In particular, Fig. 5 shows how the residual computed by LFD no. 2 is still sensitive to the fault and able to cross

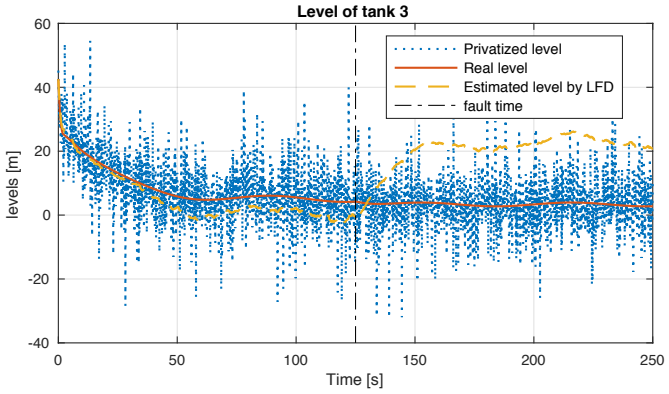

Fig. 4. Real level of tank no.3, along with its privatized version ($\epsilon = 0.04$) and its estimation by LFD no. 2.

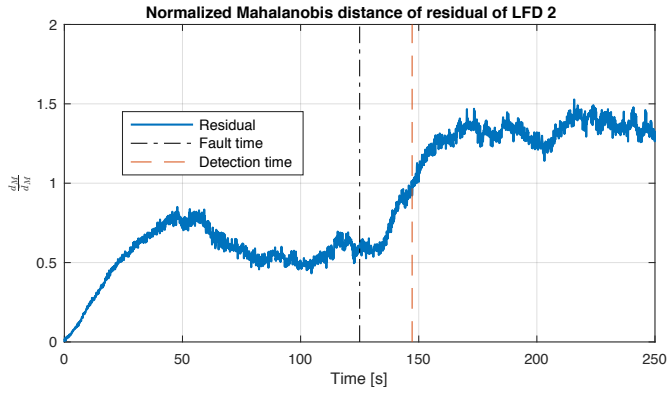

Fig. 5. Normalized Mahalanobis distance $\frac{d_M(r_2)}{\bar{d}_M}$ of LFD2 residual for $\epsilon = 0.04$.

the threshold at about 148 s. In this figure the normalized Mahalanobis distance of the residual, that is $\frac{d_M(r_2)}{\bar{d}_M}$, is plotted: a successful detection occurs when this quantity gets larger than 1. Finally, Fig. 6 presents a boxplot analysis of the effect of varying ϵ on the detection time of LFD no. 2. In order to produce such plot, a Monte Carlo approach has been used. For each value of ϵ 64 rounds were simulated, where every source of uncertainty in the system (that is v and w) and in the mechanism M_u have been implemented thanks to the default Matlab/Simulink [1] random numbers generators. The seeds of the generators have been independently initialized before each round according to the system clock time, to avoid undesired repetitions and correlations. In particular, a formula of the type seed $=$ round$(s(\text{clock}) * \text{mult} + \text{off})$ has been used, where s is an arbitrary monotone function, and mult and off are real numbers that are unique for each source of uncertainty. The end results, albeit a bit unexpected, shows how, for the chosen range of values [2] of ϵ, the median of the detection time is not significantly affected by the addition of the privacy mechanism. This first result implies that indeed the proposed privacy-preserving mechanism is feasible and will not hamper fault detection performances.

[1] Matlab/Simulink R2016a on Mac Os X 10.11.6

[2] No successful detection was obtained for values smaller than 0.04

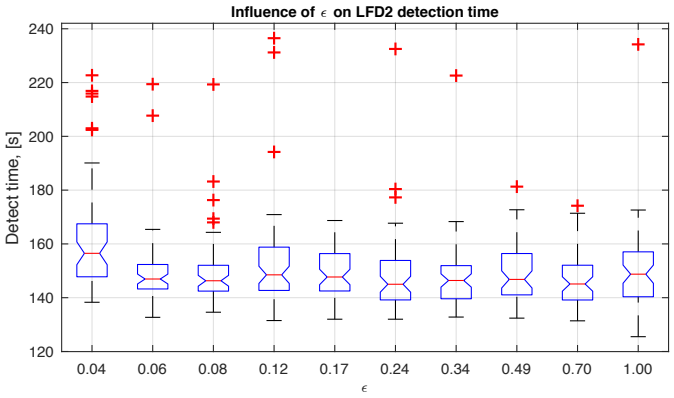

Fig. 6. Boxplot analysis of influence of privacy on detection time. Lower ϵ corresponds to higher privacy. For each value of ϵ, 64 Monte Carlo rounds were simulated.

5. CONCLUSIONS

This paper presented for the first time, to the best of the authors' knowledge, a differentially private approach to distributed fault diagnosis of large scale, nonlinear and uncertain systems. The data whose privacy must be preserved was considered to be the local input of each subsystem, which corresponds to protecting the privacy of each subsystem control algorithms and policies. A novel theoretical result linking the privacy level of the local input, to the privacy level corresponding to a given privacy mechanism applied to the subsystem output, was presented in Theorem 1. Simulation results were included, where a Monte Carlo analysis was used to show the negligible influence of the privacy mechanism on the detectability properties of the original diagnosis scheme.

REFERENCES

Albert, R. and Barabási, A.L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.

Boem, F., Ferrari, R.M.G., Parisini, T., and Polycarpou, M.M. (2015). Optimal topology for distributed fault detection of large-scale systems. *Safeprocess*, 48(21), 60–65.

Chen, X. (2007). A new generalization of Chebyshev inequality for random vectors. *arXiv preprint arXiv:0707.0805*.

Dabbene, F., Henrion, D., Lagoa, C., and Shcherbakov, P. (2015). Randomized approximations of the image set of nonlinear mappings with applications to filtering. *IFAC Symposium on Robust Control Design*, 48(14), 37–42.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, 265–284. Springer.

Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4), 211–407.

Ferrari, R.M., Baldi, S., and Dibowski, H. (2017). A message passing algorithm for automatic synthesis of probabilistic fault detectors from building automation ontologies. In *Procs. of 20th IFAC World Congress, Toulouse (France) July 9 - 14, 2017*. IFAC.

Ferrari, R.M., Parisini, T., and Polycarpou, M.M. (2008). A robust fault detection and isolation scheme for a class of uncertain input-output discrete-time nonlinear systems. In *American Control Conference, 2008*, 2804–2809.

Ferrari, R.M., Parisini, T., and Polycarpou, M.M. (2012). Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach. *IEEE Trans. Autom. Contr.*, 57(2), 275–290.

Ge, X. and Han, Q.L. (2014). Distributed fault detection over sensor networks with markovian switching topologies. *International Journal of General Systems*, 43(3-4), 305–318.

Han, S., Topcu, U., and Pappas, G.J. (2014a). Differentially private convex optimization with piecewise affine objectives. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, 2160–2166. IEEE.

Han, S., Topcu, U., and Pappas, G.J. (2014b). Differentially private distributed protocol for electric vehicle charging. In *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, 242–249. IEEE.

Han, S., Topcu, U., and Pappas, G.J. (2017). Differentially private distributed constrained optimization. *IEEE Transactions on Automatic Control*, 62(1), 50–64.

Kyriakides, E. and Polycarpou, M. (2014). *Intelligent Monitoring, Control, and Security of Critical Infrastructure Systems*, volume 565. Springer.

Le Ny, J. and Pappas, G.J. (2014). Differentially private filtering. *IEEE Transactions on Automatic Control*, 59(2), 341–354.

Mo, Y. and Murray, R.M. (2017). Privacy preserving average consensus. *IEEE Transactions on Automatic Control*, 62(2), 753–765.

Noursadeghi, E. and Raptis, I. (2017). Reduced-order distributed fault diagnosis for large-scale nonlinear stochastic systems. *Journal of Dynamic Systems, Measurement, and Control*.

Riverso, S., Boem, F., Ferrari-Trecate, G., and Parisini, T. (2016). Plug-and-play fault detection and control-reconfiguration for a class of nonlinear large-scale constrained systems. *IEEE Transactions on Automatic Control*, 61(12), 3963–3978.

Rostampour, V., Ferrari, R., and Keviczky, T. (2017). A set based probabilistic approach to threshold design for optimal fault detection. In *2017 American Control Conference (ACC)*, 5422–5429.

Sankar, L., Kar, S., Tandon, R., and Poor, H.V. (2011). Competitive privacy in the smart grid: An information-theoretic approach. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, 220–225. IEEE.

Zhang, D., Wang, Q.G., Yu, L., and Song, H. (2013). Fuzzy-model-based fault detection for a class of nonlinear systems with networked measurements. *IEEE Transactions on Instrumentation and Measurement*, 62(12), 3148–3159.

Zhang, Q. and Zhang, X. (2013). Distributed sensor fault diagnosis in a class of interconnected nonlinear uncertain systems. *Annual Reviews in Control*, 37(1), 170–179.