Privatized distributed anomaly detection for large-scale nonlinear uncertain systems

Rostampour, Vahab; Ferrari, Riccardo M.G.; Teixeira, Andre M.H.; Keviczky, Tamas

DOI
10.1109/TAC.2020.3040251

Publication date
2020

Document Version
Final published version

Published in
IEEE Transactions on Automatic Control

Privatized Distributed Anomaly Detection for Large-Scale Nonlinear Uncertain Systems

Vahab Rostampour, *Member, IEEE*, Riccardo M.G. Ferrari, *Member, IEEE*, André M.H. Teixeira, *Member, IEEE*, and Tamás Keviczky, *Senior Member, IEEE*

***Abstract*—In this article two limitations in current distributed model based approaches for anomaly detection in large-scale uncertain nonlinear systems are addressed. The first limitation regards the high conservativeness of deterministic detection thresholds, against which a novel family of set-based thresholds is proposed. Such set-based thresholds are defined in a way to guarantee robustness in a user-defined probabilistic sense, rather than a deterministic sense. They are obtained by solving a chance-constrained optimization problem, thanks to a randomization technique based on the Scenario Approach. The second limitation regards the requirement, in distributed anomaly detection architectures, for different parties to regularly communicate local measurements. In settings where these parties want to preserve their privacy, communication may be undesirable. In order to preserve privacy and still allow for distributed detection to be implemented, a novel privacy-preserving mechanism is proposed and a so-called *privatized communication* protocol is introduced. Theoretical guarantees on the achievable level of privacy, along with a characterization of the robustness properties of the proposed distributed threshold set design, taking into account the privatized communication scheme, are provided. Finally, simulation studies are included to illustrate our theoretical developments.**

***Index Terms*—Fault detection, large scale systems, privacy.**

Manuscript received June 16, 2020; revised September 17, 2020; accepted November 11, 2020. Date of publication November 24, 2020; date of current version November 4, 2021. 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) under Grant 408-13-030, by the European Union H2020 program under the project SURE: Safe Unmanned Robotic Ensembles under Grant 707546, by the Swedish Research Council under Grant 2018-04396, and by the Swedish Foundation for Strategic Research. Recommended by Associate Editor G. Gu. *(Corresponding author: Riccardo M.G. Ferrari.)*

Vahab Rostampour is with the Engineering and Technology Institute Groningen, Faculty of Science and Engineering, University of Groningen, Groningen, CP 9712, The Netherlands (e-mail: v.rostampour@rug.nl).

André M.H. Teixeira is with the Department of Electrical Engineering, Division of Signals and Systems, Uppsala University, Box 65, Uppsala SE-751 03, Sweden (e-mail: andre.teixeira@angstrom.uu.se).

Riccardo M.G. Ferrari and Tamás Keviczky are with the Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, Delft, CD 2628, The Netherlands (e-mail: r.ferrari@tudelft.nl; t.keviczky@tudelft.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2020.3040251.

Digital Object Identifier 10.1109/TAC.2020.3040251

I. Introduction

FAULT diagnosis and security for large-scale nonlinear systems, such as critical infrastructures or interconnected cyber physical systems have received increasing attention in recent years [1]. One way to increase the resiliency of such systems to faults or cyber attacks is to endow them with the capabilities of detecting, isolating and mitigating those threats. In particular, *model-based* (MB) fault diagnosis methods have emerged in some sectors, such as aerospace, as powerful tools for guaranteeing high operational readiness levels and reducing maintenance costs [2]. In MB approaches a mathematical model of the system under monitoring is used to produce a time-varying residual, which is then compared to a static or dynamic threshold for detection. Anyway, widespread industrial adoption has been slow due to at least the following limitations:

a) The scarcity of robust design methods for diagnosis of nonlinear systems, leading to easy-to-tune performance levels in terms of the so-called *false alarm ratio* (FAR) and *missed detection ratio* (MDR). Ideally, a threshold should be *robust* with respect to model and measurement uncertainties, thus having a zero or low FAR. At the same time, it should have good *detection* properties, which translates into a negligible MDR. The problem of minimizing both FAR and MDR has been solved for linear systems [3] and a class of nonlinear systems [4], where by using geometric tools it is possible to design a detection residual that is insensitive to uncertainties or unknown inputs, and sensitive to a single class of faults. For general nonlinear systems and/or unstructured uncertainties and faults, to the best of authors' knowledge, it has not. In this case it is customary to assume the existence of a known, static or dynamic deterministic upper bound on the uncertainties' magnitude, thus allowing to obtain a zero FAR by design [2], [5]. Unfortunately, such a powerful property often comes at the cost of conservative thresholds, which lead to high MDR.

b) The lack of privacy-preserving distributed anomaly detection implementations for large-scale systems. For such systems, distributed anomaly detection approaches were shown to possess favourable properties. They are based on several *local detectors* (LD), each one monitoring only a limited subsystem and communicating with neighboring LDs [6]–[12]. An unexplored issue in this setting indeed arises from the need for communication. In the case of a very large infrastructure, where such LDs 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 neighboring LDs 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 [13], [14].

The first contribution of this article is to ease the first limitation by introducing a class of novel, adaptive, parametrized threshold *sets* for distributed fault detection. We aim to reduce the conservatism of existing threshold designs [2], [6], [7], [10] by relaxing the deterministic robust zero-FAR condition, in favor of a more flexible, *probabilistic* one. Through a set-based approach, the probability of false alarms will be defined as a user-tunable design parameter, and the detection with respect to a given class of faults will be simultaneously maximized. The use of probabilistic thresholds in MB fault diagnosis has been investigated previously in the literature (see [2] and the references cited therein), and recently the important case of nonlinear uncertain systems has been considered [15], [16]. The use of sets in fault diagnosis has been inspired by the corpus of works on set-membership system identification [17], which initially addressed the inverse problem of finding, at each time step, the set of system parameters that could be able to explain current measurements, and compare it to a nominal one [18], [19]. Other works considered instead the direct problem of describing the admissible values of the residual in healthy condition using a set [20], with [21] being a notable example in the field of active fault diagnosis.

The main contribution of this article is to address the second limitation, by designing a novel privacy-preserving mechanism based on a so-called *privatized communication scheme*. In particular, we will show how the proposed mechanism satisfies a relaxed notion of *differential privacy* (DP). The concept of DP emerged in the computer science community [22], [23], but found application in several problems in the fields of consensus, distributed estimation and control [13], [24]–[27], as well as distributed monitoring and fault [28] and attack detection [29].

Differently than existing works on DP, including the preliminary results published by the authors in [28], the proposed privatized communication scheme will not rely on the classic additive Laplacian noise mechanism, nor will require the computation of the query sensitivity. Instead, it will be based on LDs communicating the parametrization of randomized sets guaranteed to satisfy the DP condition with a given confidence level. The advantages of this approach will be twofold. On one side, it will reduce the quantity of data that neighboring LDs need to communicate at each sampling time. On the other side, it will allow to connect theoretically the performances of the distributed anomaly detection scheme to the level of privacy desired.

The structure of this article is as follows. Section II describes a large scale system as the interconnection of nonlinear uncertain systems, and introduces a distributed set-based probabilistic threshold design. Section III presents a privacy preserving mechanism and the so-called privatized communication scheme for local detectors. Two different simulation studies are provided in Section IV to illustrate the effectiveness of our proposed approach. Some final remarks and future work will be given in Section V.

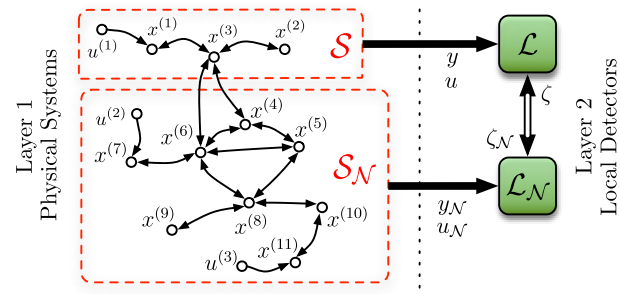


Fig. 1. Proposed distributed anomaly detection architecture, based on a simplification of [6] for the case of two interconnected systems only. On the left side, the structure of a hypothetical large-scale system is represented as a directed graph, and is already decomposed as the interconnection of systems $\mathcal{S}$ and $\mathcal{S}_\mathcal{N}$. Nodes represent state variables and a directed edge represents a causal dependence. On the right side, the local detectors $\mathcal{L}$ and $\mathcal{L}_\mathcal{N}$ are depicted. Thick black lines represent the acquisition of local measurements by the LDs, while thick white lines represent the communication between neighboring LDs. Such communication include the measured values ζ and $\zeta_\mathcal{N}$ of the interconnection variables $z = [x^{(4)}\ x^{(6)}]^\top$ and $z_\mathcal{N} = x^{(3)}$.

II. Problem Statement

In this article we will consider interconnected, uncertain nonlinear dynamical systems. We will assume that for anomaly detection purposes each system $\mathcal{S}$ is monitored by a dedicated *local detector* (LD) $\mathcal{L}$. This will lead to a distributed anomaly detection architecture (see Fig. 1) similar to the one proposed in [6], [11]. For the sake of simplicity and ease of notation we will assume that only two systems are making up the interconnection. They will be referred to as the *ego* system $\mathcal{S}$ and the *neighbour* $\mathcal{S}_\mathcal{N}$. The index $\mathcal{N}$ will be used throughout the article to indicate when a given quantity refers to $\mathcal{S}_\mathcal{N}$. There shall be no loss of generality, as the extension to an arbitrary number of interconnected (sub)systems can be easily obtained using the framework introduced in [6].

A. System Dynamics

Let the dynamics of $\mathcal{S}$ and $\mathcal{S}_\mathcal{N}$ have the following form

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

$$\mathcal{S}_\mathcal{N}: \begin{cases} x_{\mathcal{N},k+1} &= g(x_{\mathcal{N},k},\ u_{\mathcal{N},k},\ z_{\mathcal{N},k},\ w_{\mathcal{N},k},\ f_{\mathcal{N},k}) \\ y_{\mathcal{N},k} &= x_{\mathcal{N},k} + v_{\mathcal{N},k} \end{cases} \tag{1}$$

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 the ego system $\mathcal{S}$ at discrete time index k, respectively. Similarly, the variables with the index $\mathcal{N}$ represent the same quantities for the neighbour system $\mathcal{S}_\mathcal{N}$.

In the following, for the sake of brevity, variables, equations, assumptions and theoretical results will be presented for the ego

system $\mathcal{S}$ only but will be intended to hold as well, *mutatis mutandis*, for $\mathcal{S}_\mathcal{N}$. Still, we will not assume that $\mathcal{S}_\mathcal{N}$ and $\mathcal{S}$ have the same dimension or the same dynamics.

The term $z \in \mathbb{R}^s$ is called *interconnection variable* [6] and is comprised of all the components of the neighbour state $x_\mathcal{N}$ which influence the dynamics of the ego system $\mathcal{S}$. The full state x is assumed to be measured, albeit corrupted by a measurement uncertainty $v_k \in \mathbb{R}^n$. The vector $\zeta_k = z_k + \xi_k$ will be used to denote the measurements of the interconnection variables z_k, with $\xi_k \in \Xi \subset \mathbb{R}^s$ collecting the components of the neighbour measurement noise $v_{\mathcal{N},k}$ that affect ζ_k.

The variable $w_k \in \mathbb{R}^p$, instead, represents unavoidable modeling uncertainties affecting (1), while $f_k \in \mathcal{F} \subseteq \mathbb{R}^q$ represents a parametrization of the dynamic influence of anomalies. Such formulation comprises the cases, where either w_k and f_k affect the nonlinear dynamics function $g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^n$ as additive/multiplicative terms, or where these affect one or more parameters that appear in the definition of g. For instance, if $\mathcal{S}$ models an electrical circuit, w_k could correspond to parametric uncertainties in the electrical resistance of individual conductors or components, f_k could describe eventual open circuit or ground faults, and v_k represents the uncertainty in the measurements provided by a number of voltmeters connected to the circuit. In this respect, the functional dependence of g on f_k, together with its domain set $\mathcal{F}$, describe the class of the possible dynamic anomalies occurring in $\mathcal{S}$. The only structural assumption required on g is that $w_k = 0$ and $f_k = 0$ corresponds to the nominal and normal behavior of $\mathcal{S}$, that is in the absence of uncertainties and anomalies. The following technical assumptions are needed for the upcoming analysis.

Assumption 1: w_k and v_k are random variables defined on some compact 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 a probability measure 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, for all k.

Assumption 2: No anomaly acts on the system 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, on and after k_f, 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 3: The vector field $g(\cdot)$ will be assumed to be differentiable and globally Lipschitz with respect to all its arguments.

Remark 1: Assumption 1 mentions that uncertainties are from independent compact spaces: it is worth to highlight that we do not require the sample spaces $\mathcal{W}$, $\mathcal{V}$ and the probability measures $\mathbb{P}_\mathcal{W}$, $\mathbb{P}_\mathcal{V}$ to be known explicitly, as it will be explained in Section II-C. Assumption 2 is required for well posedness when designing the detection thresholds described later in this Section. It is important to note that Assumption 1 is in general needed for Assumption 2 to hold. Finally, Assumption 3 is not restrictive for the proposed framework, and indeed, it can be easily relaxed to a more easily enforceable local Lipschitz condition using Assumption 2 and [30, Assumption 2].

B. Distributed Residual Generator

In the present article we will generalize the approach followed in [31] and define the residual computed by $\mathcal{L}$ as $r_k := y_k - \hat{y}_k$. It can be obtained as the output estimation error of the following nonlinear observer:

$$\begin{cases} \hat{x}_{k+1} &= g(y_k, u_k, \zeta_k, 0, 0) + \Lambda(\hat{y}_k - y_k) \\ \hat{y}_k &= \hat{x}_k \end{cases} \tag{2}$$

where $\hat{x}$ and $\hat{y} \in \mathbb{R}^n$ are, respectively, the local state and output estimates, $\Lambda \triangleq \mathrm{diag}(\lambda^i,\ i = 1 \ldots n)$ is a diagonal matrix, whose elements denote filtering parameters which are chosen such that $|\lambda^i| < 1$.

It can be seen that the estimator in (2) is defined using only quantities that are supposed to be available at run-time to the agent $\mathcal{L}$, namely: the nominal local dynamics g, the local outputs y_k and inputs u_k, and the measurements ζ_k of the interconnection variables of the neighboring agent $\mathcal{L}_\mathcal{N}$. The local use of ζ_k requires some form of regular communication between $\mathcal{L}_\mathcal{N}$ and $\mathcal{L}$: the resulting privacy implications will be addressed in Section III.

By using (1) and (2), we can write the residual dynamics as

$$r_{k+1} = \Lambda\, r_k + \delta_k \tag{3}$$

where we introduced the *total uncertainty* δ_k

$$\begin{aligned} \delta_k &:= g(x_k, u_k, z_k, w_k, f_k) - g(y_k, u_k, \zeta_k, 0, 0) + v_{k+1} \\ &= g(y_k - v_k, u_k, \zeta_k - \xi_k, w_k, f_k) \\ &\quad - g(y_k, u_k, \zeta_k, 0, 0) + v_{k+1} \end{aligned} \tag{4}$$

which is a stochastic process representing the uncertain part of the residual dynamics. Owing to Assumption 1 and 2, it follows that δ_k is a random variable on a probability space $(\Delta_k, \mathfrak{B}(\Delta_k), \mathbb{P}_{\Delta_k})$, where Δ_k is a time-varying set defined as follows.

Definition 1: The time-varying *total uncertainty set* $\Delta_k \subset \mathbb{R}^n$ at time index k is defined as

$$\Delta_k := \{\delta_k \,|\, w_k \in \mathcal{W},\ f_k \in \mathcal{F},\ v_k \in \mathcal{V},\ v_{k+1} \in \mathcal{V},\ \xi_k \in \Xi\}$$

where δ_k is computed according to (4).

The following lemma is now provided to present the stability of the proposed residual generator.

Lemma 1: Given Assumptions 1, 2, and 3 the following statements hold:

a) δ_k is bounded and the set Δ_k is compact;
b) the residual r_k is bounded; and
c) the state estimation error $e_k := x_k - \hat{x}_k$ is bounded.

Proof: In order to prove a), we recall that under Assumption 1 the uncertainties w_k and v_k belong to the compact sets $\mathcal{W}$ and $\mathcal{V}$, respectively. Similarly, Assumption 2 introduced bounded stability regions $\mathcal{S}^x$ and $\mathcal{S}^u$ for, respectively, the state x_k and control input u_k. Based on the Lipschitz condition on g given in Assumption 3, this directly leads to δ_k being bounded and so the set Δ_k being compact, proving a). To prove b), we notice that (3) represents the dynamics of an asymptotically stable discrete time linear system driven by the input δ_k, as by construction Λ is

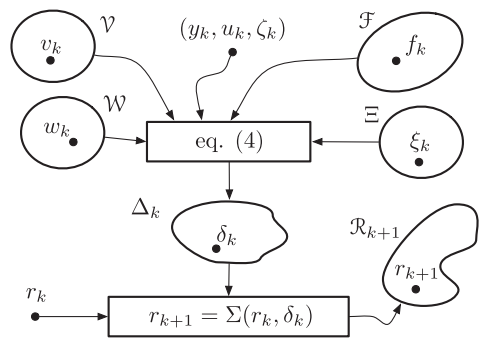


Fig. 2. Residual set $\mathcal{R}_{k+1}$ can be thought of as the set obtained by computing the output Σ while letting δ_k vary over its domain Δ_k and fixing the residual r_k to its actual value. The domain Δ_k in turn is computed through (4) by letting v_k, w_k, f_k and ξ_k vary over their respective domains, and fixing the local output and input y_k and u_k, as well as the interconnection variables measurement ζ_k, to their actual values. The normal residual set $\mathcal{R}^0_{k+1}$ can be obtained similarly, but by fixing the value $f_k = 0$.

a Schur matrix. By making use of the concept of bounded input-bounded output stability, which for linear systems is implied by asymptotic stability (cf., the classical stability results for discrete time systems [32, Section 3.2]), the proof is straightforward. Finally, recalling that $r_k = y_k - \hat{y}_k = e_k + v_k$, the statement c) directly follows from b) and Assumption 1, which, respectively, state that r_k and v_k are bounded. □

We will introduce also the following definition, as a special case of Definition 1.

Definition 2: The time-varying *normal total uncertainty set* $\Delta^0_k \subset \mathbb{R}^n$ at time index k is defined as

$$\Delta^0_k := \{\delta_k \,|\, w_k \in \mathcal{W},\ f_k \in \{0\},\ v_k \in \mathcal{V},\ v_{k+1} \in \mathcal{V},\ \xi_k \in \Xi\}$$

where δ_k is computed according to (4).

The role of Δ_k and Δ^0_k is to quantify the range of possible values that the total uncertainty δ_k can take, respectively, in an arbitrary condition during which an anomaly *may* be present, and a normal condition where an anomaly is not.

It is important to highlight that Δ^0_k has a central role in deriving a probabilistically robust detection threshold, whereas, Δ_k is instrumental in improving detectability.

We can now introduce a compact notation for the *residual generator* described by (2)–(4), through a mapping function $\Sigma : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n$ defined as

$$r_{k+1} := \Sigma(r_k, \delta_k)\,. \tag{5}$$

The mapping from the uncertain variable $\delta_k \in \Delta_k$ to the residual variables r_{k+1} is measurable, so that the residual signal r_{k+1} can be viewed as a random variable on the same probability space as δ_k.

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

Definition 3: The time-varying *residual set* $\mathcal{R}_{k+1}$ at time index $k+1$ is defined as the image of the set Δ_k through Σ, that is

$$\mathcal{R}_{k+1} := \Sigma(r_k, \Delta_k) = \{r_{k+1} \,|\, r_{k+1} = \Sigma(r_k, \delta),\ \delta \in \Delta_k\}.$$

Similarly, for a particular class of anomalies such that f_k belongs to a given set $\mathcal{F}'$, the notation $\mathcal{R}^{\mathcal{F}'}_{k+1}$ will be used.

Definition 4: The time-varying *normal residual set* $\mathcal{R}^0_{k+1}$ at time index $k+1$ is defined as the image of the set Δ^0_k through Σ, that is

$$\mathcal{R}^0_{k+1} := \Sigma(r_k, \Delta^0_k) = \{r_{k+1} \,|\, r_{k+1} = \Sigma(r_k, \delta),\ \delta \in \Delta^0_k\}.$$

Remark 2: The observer-based residual generator introduced in (2) is inspired by the observer used in [5] and related literature. Anyway, it is important to note that the set-based threshold design methodology described in the following Subsection is not dependent on the specific residual generator, or the way its parameters are chosen. Instead, the threshold robustness to uncertainty and detectability properties will only depend on the possibility to generate samples of normal and abnormal residuals and use them to solve the chance-constrained optimization problems (10a) and (10b). Furthermore, differently from classical structured residual and threshold generation methods or from hypothesis testing techniques (see e.g., [4] or [33]) the present article does not require or make use of specific structures of the system dynamics or of the uncertainty, nor is based on a specific probability distribution of the latter.

C. Distributed Set-Based Probabilistic Threshold

Given a residual generator Σ as defined in (5), we now design a threshold for anomaly detection with suitable robustness and detection performance guarantees by leveraging the probabilistic set-based approach introduced in [31]. We will extend it to a distributed setting, leading to the need for agents to communicate to each neighbour for implementing the required computations. This highlights the necessity for developing a privatized communication protocol as our next main contribution. We now introduce the following detection logic.

Definition 5: An anomaly is *detected* at time index k if $r_k \notin \mathcal{T}_k$, where $\mathcal{T}_k \subseteq \mathbb{R}^n$ is an *adaptive threshold set*.

While this definition explains how the residual is evaluated by $\mathcal{L}$ (or $\mathcal{L}_{\mathcal{N}}$), it does not specify how the threshold set $\mathcal{T}_k$ shall be computed at each time index k. In the fault diagnosis literature, the concepts of *False Alarm Ratio* (FAR) and *Missed Detection Rate* (MDR) have been introduced to characterize the performance of detection algorithms (see for instance [2], [34]). The design approach in [31] allows to define $\mathcal{T}_k$ such that a prescribed FAR can be obtained in a probabilistic sense, while minimizing the MDR.

In particular, if the user-desired FAR level is equal to $1-\alpha$, with $\alpha \in [0,\,1]$, the following threshold can guarantee it in a probabilistic sense.

Definition 6: Given the residual generator function Σ in (5) and a fixed $\alpha \in [0, 1]$, an adaptive threshold set $\mathcal{T}_{k+1}$ is said to be *probabilistically α–robust*, if

$$\mathcal{V}(\mathcal{T}_{k+1}) := \mathbb{P}\left[r_{k+1} \notin \mathcal{T}_{k+1} \,\middle|\, \delta_k \in \Delta^0_k\right] < 1-\alpha \tag{6}$$

where $\mathcal{V}(\mathcal{T}_{k+1})$ is the *violation probability (FAR)* of $\mathcal{T}_{k+1}$.

In order to find a threshold set fulfilling Definition 6 and minimizing the MDR, a chance-constrained optimization problem

such as the following shall be solved

$$\begin{cases} \min\limits_{\theta_k} & \mathrm{vol}\,(\mathcal{T}_k) \\ \text{s.t.} & \mathcal{V}(\mathcal{T}_k) < 1-\alpha \end{cases} \tag{7}$$

where $\mathrm{vol}(\mathcal{T}_k) := \int_{\mathcal{T}_k} \mathrm{d}r$ is the volume or Lebesgue measure of $\mathcal{T}_k$, and θ_k is a time-varying parameters vector that characterizes the set $\mathcal{T}_k$.

The rationale for seeking the minimum volume set is indeed to minimize the MDR, as explained in [31]. To solve (7), we characterize the adaptive threshold set $\mathcal{T}_k$ as the *c–superlevel set* [35] of a *generalized indicator function* $\mathbf{1}_{\mathcal{T}}(r,\theta_k) : \mathbb{R}^n \times \mathbb{R}^t \mapsto \mathbb{R}$, which in turn is parameterized by $\theta_k \in \mathbb{R}^t$

$$\mathcal{T}_k := \{r \in \mathbb{R}^n \,|\, \mathbf{1}_{\mathcal{T}}(r,\theta_k) \geq c\}\,. \tag{8}$$

In particular, as in [31] $\mathbf{1}_{\mathcal{T}}(r,\theta_k)$ will be restricted to be a Sum-of-Squares (SoS) polynomial function of given degree d [36], [37], with θ_k containing the polynomial coefficients in a given order.

Denoting by $\pi_\xi(r)$ a vector of monomials of degree up to $\xi := \lceil d/2 \rceil$,[1] we can conveniently define $\mathbf{1}_{\mathcal{T}}(r,\theta_k) := \pi_\xi(r)^\top G(\theta_k)\,\pi_\xi(r)$, where $G(\theta_k)$ is a symmetric Gram matrix depending on θ_k. This choice allows to bound the objective function as

$$\mathrm{vol}\,(\mathcal{T}_k) = \int_{\mathcal{T}_k} \mathrm{d}r \leq \frac{1}{c}\int_{\mathcal{B}} \mathbf{1}_{\mathcal{T}}(r,\theta_k)\mathrm{d}r = \frac{1}{c}\mathrm{trace}(G(\theta_k)M)$$

where $\mathcal{B} \in \mathbb{R}^n$ is an arbitrary compact set so that $\mathcal{R}_k^0 \subset \mathcal{B}$ and $M := \int_{\mathcal{B}} \pi_\xi(r)\pi_\xi(r)^\top \mathrm{d}r$ denotes the matrix of moments of the Lebesgue measure on $\mathcal{B}$ in basis $\pi_\xi(r)$.

We are now in a position to propose the following cascade of two chance-constrained optimization problems for designing a probabilistic threshold set for time $k+1$:

$$\begin{cases} \min\limits_{\theta,\gamma} & \gamma \\ \text{s.t.} & G(\theta) \succ 0\,, \quad \mathrm{trace}(G(\theta)M) \leq \gamma\,, \\ & \mathbb{P}\left[\,\mathbf{1}_{\mathcal{T}}(r^0,\theta) \geq c\,\right] \geq \alpha \end{cases} \tag{9a}$$

$$\begin{cases} \max\limits_{\theta} & \|G(\theta) - G(\psi^*)\|_\infty \\ \text{s.t.} & G(\theta) \succ 0\,, \quad \mathrm{trace}(G(\theta)M) \leq \gamma^*\,, \\ & \mathbb{P}\left[\,\mathbf{1}_{\mathcal{T}}(r^0,\theta) \geq c\,\right] \geq \alpha \end{cases} \tag{9b}$$

where the quantity γ^* is the optimal cost obtained by solving the first stage (9a), while (9) has to be solved sequentially in a *lexicographic* (multiobjective) sense [38].

The first problem (9a) aims at determining the minimum volume threshold set $\mathcal{T}_{k+1}$ subject to the probabilistic α–robust constraint, but in doing so is ignoring any information on the abnormal residual set $\mathcal{R}_{k+1}^{\mathcal{F}'}$. This could possibly lead to unsatisfactory detection properties due to a large intersection $\mathcal{T}_{k+1} \cap \mathcal{R}_{k+1}^{\mathcal{F}'}$. The goal of the second stage problem (9b) is then to find a new parameter θ_{k+1}, leading to a new threshold set $\mathcal{T}_{k+1}$ with the same robustness guarantee and a volume which is not worse than the one resulting from the solution of problem (9a), but which is as distant as possible from the set $\mathcal{R}_{k+1}^{\mathcal{F}'}$. To achieve this goal, we formulate the objective function to aim at maximizing the Chebyshev distance, or polynomial height [39], between $\mathbf{1}_{\mathcal{T}_k}$ and $\mathbf{1}_{\mathcal{R}_k^{\mathcal{F}'}}(r,\psi^*) := \pi_\xi(r)^\top G(\psi^*)\,\pi_\xi(r)$. The parameter ψ^* is such that $\mathbf{1}_{\mathcal{R}_k^{\mathcal{F}'}}(r,\psi^*) \geq c$, $\forall r \in \mathcal{R}^{\mathcal{F}'}$. Indeed, by assuming that both $\mathbf{1}_{\mathcal{T}_k}$ and $\mathbf{1}_{\mathcal{R}_k^{\mathcal{F}'}}$ share the same monomial basis vector $\pi_\xi(r)$, this leads to the maximization of the distance $\|G(\theta) - G(\psi^*)\|_\infty$ between their Gram matrices [39]. We refer the interested reader to [31] for a more complete explanation on the second stage problem formulation.

The proposed optimization problem (9) is however, nonconvex and hard to solve due to chance constraints being in general difficult to enforce. To overcome this difficulty, we provide a computationally tractable approach, thanks to a randomization technique based on the *scenario approach* [40], leading to the following tractable problem formulation:

$$\begin{cases} \min\limits_{\theta,\gamma} & \gamma \\ \text{s.t.} & G(\theta) \succ 0\,, \quad \mathrm{trace}(G(\theta)M) \leq \gamma\,, \\ & \mathbf{1}_{\mathcal{T}}(r^{0,i},\theta) \geq c\,, \quad i = 1,\ldots,N_s \end{cases} \tag{10a}$$

$$\begin{cases} \max\limits_{\theta} & \|G(\theta) - G(\psi^*)\|_\infty \\ \text{s.t.} & \theta \succ 0\,, \quad \mathrm{trace}(G(\theta)M) \leq \gamma^*\,, \\ & \mathbf{1}_{\mathcal{T}}(r^{0,i},\theta) \geq c\,, \quad i = 1,\ldots,N_s \end{cases} \tag{10b}$$

where the chance constraint has been replaced by N_s hard constraints corresponding to samples from the uncertainty realizations. In particular, the samples $r^{0,i}$ are evaluated for time $k+1$ following $r_{k+1}^{0,i} = \Sigma(r_k, \delta_k^i)$, and $\delta_k^i \in \Delta_k^0$ are samples of the random variable δ_k. We assume to be able to generate N_s samples based on the knowledge of $\Sigma(\cdot)$, and availability of the uncertainty samples from Δ_k^0, following Definition 2.

The following Theorem provides a link between the chance-constrained problem (9), its randomized counterpart (10) and the number of samples N_s.

Theorem 1: Given d the degree of polynomial function of $\mathbf{1}_{\mathcal{T}}(r,\theta_k)$, consider $\upsilon := [\theta,\gamma]^\top \in \mathbb{R}^\ell$ to be the augmented vector of all the decision variables of (10). Let $\beta \in [0,1]$ and $N_s \geq \mathsf{N}_s(\alpha,\beta,\ell)$, where

$$\mathsf{N}_s(\alpha,\beta,\ell) = \min\left\{N_s \in \mathbb{N} \middle| d\sum_{i=0}^{\ell-1}\binom{N_s}{i}(1-\alpha)^i\alpha^{N_s-i} \leq \beta\right\},$$

Then, the optimizer $\upsilon^* := [\theta_b^*, \gamma^*]^\top$ of the randomized convex program (10) is a feasible solution of the chance-constrained optimization problem (9) with confidence level $(1-\beta)$, in the average.

Following the proposed optimization problem in (10a), generating samples of the normal residual r^0 requires the availability of samples of δ. According to Definition 2 together with (4), it emerges that for computing samples δ^i the agent $\mathcal{L}$ must know the current measured value ζ_k of the interconnection variables, as well as samples of candidate values for the true interconnection variables that are compatible with the measurement uncertainties. Such samples will be denoted as $x_{\mathrm{N},k}^i = \zeta_k - \xi_k^i$, with i

[1] $\lceil\cdot\rceil$ is the ceiling operator which returns the smallest integer greater than or equal to its argument.

indicating the i–th sample, and their generation necessitates in turn the ability to generate samples of the uncertainties ξ_k. This requires that the neighboring agent $\mathcal{L}_\mathcal{N}$ computes such samples and communicates them to $\mathcal{L}$ along with the measurement of the interconnection variable z_k. Such communication may expose private information of $\mathcal{S}_\mathcal{N}$, such as its local input or a possibly high number of samples of its measurement uncertainty, which could be used to estimate the probability density function of such uncertainty. This highlights the necessity for developing a privatized communication protocol as our next main contribution in the following section.

III. Privatized Distributed Anomaly Detection

We now develop a new communication scheme for each agent to preserve its privacy. We then present a theoretical analysis to show that such a communication is indeed a novel differentially private [23] mechanism with high confidence level, which yields a new framework that we refer to as differentially private distributed anomaly detection. Using such a privatized communication scheme, we finally provide a theoretical guarantee to accommodate the privatized communication scheme for the proposed probabilistic threshold design technique in Section II.

We will briefly recall here that as a foundation in differential privacy (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 [25] are nearly indistinguishable from the answers to a query.

A. Inter-Agent Communication Scheme

For the sake of simplicity, we will drop the time indices to ease our notation whenever possible.

Following the scheme developed in Section II, the neighbor $\mathcal{L}_\mathcal{N}$ should send to $\mathcal{L}$ at each time index its last interconnection variable measurement ζ, along with a set of samples $\xi = \{\xi^1, \ldots, \xi^{N_s}\}$ of its measurement uncertainty. With such data $\mathcal{L}$ can build the following set of candidate values for the variable z as follows:

$$\mathcal{Z} = \{z^1, \ldots, z^{N_s}\} := \{\zeta\} \ominus \Xi \tag{11}$$

where $\ominus$ is defined as $\mathcal{A} \ominus \mathcal{B} := \{a - b \,|\, a \in \mathcal{A}, b \in \mathcal{B}\}$.

In this line of thought, agent $\mathcal{L}_\mathcal{N}$ is the curator of a database that contains the last local input $u_{\mathcal{N},k-1}$, and at time k is answering a query from $\mathcal{L}$ by providing the following set

$$D_\mathcal{N} := \{\zeta\} \cup \mathcal{Z}\,. \tag{12}$$

A desired goal of $\mathcal{L}_\mathcal{N}$ is to replace such an answer with a mechanism that guarantees the privacy of $u_\mathcal{N}$. It is important to mention that all elements of $D_\mathcal{N}$ are related to ζ through (11), and therefore $D_\mathcal{N}$ contains no more useful information than ζ itself.

Before proceeding further, we need to provide the following definition, which is an extension of adjacency relation as the basic concept of differential privacy (DP) [23].

Definition 7: 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 $\mathrm{adj}(u_\mathcal{N}, u'_\mathcal{N})$. 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 3: Throughout this section, when referring to *privacy concerns*, an implicit adversary model is considered, where the adversary aims at making a correct decision about the status of confidential properties of the system, based on observed data and accurate models of the system. In particular, such an adversary may be cast as a decision problem, where the adversary must decide between two mutually-exclusive hypothesis about confidential system configurations: $\mathcal{H}_1 : u_\mathcal{N} = u^*_\mathcal{N}$ and $\mathcal{H}_0 : u_\mathcal{N} = u^{*\prime}_\mathcal{N}$, where $u^*_\mathcal{N}$ and $u^{*\prime}_\mathcal{N}$ are adjacent control inputs.

Following typical approaches in change detection theory, the adversary could construct an hypothesis test by filtering the observed data based on the known model dynamics, similar to the framework introduced in Section II to detect anomalies.

To preserve the privacy of the input of the system monitored by agent $\mathcal{L}_\mathcal{N}$, consider its system output to be $y_{\mathcal{N},k} := g_\mathcal{N}(\psi_{\mathcal{N},k-1}, u_{\mathcal{N},k-1})$, where $g_\mathcal{N}(\psi_{\mathcal{N},k-1}, u_{\mathcal{N},k-1})$ represents a compact notation for $\mathcal{S}_\mathcal{N}$ dynamics in (1), and 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 $\psi_\mathcal{N} := \mathrm{col}(x_\mathcal{N}, z_\mathcal{N}, w, f)$, with $\Psi := \mathcal{S}^{x_\mathcal{N}} \times \mathcal{S}^x \times \mathcal{W} \times \mathcal{F}$.

For our proposed probabilistic threshold set design $\mathcal{T}_k$, agent $\mathcal{L}$ requests from neighboring agents the complete inter-agent data, element by element. It is important to mention that the number of required samples of z, N_s, is chosen according to Theorem 1 in order to have a given probabilistic robustness for $\mathcal{T}_k$.

To address the privacy concern of the agent $\mathcal{L}_\mathcal{N}$, we propose an alternative scheme, where instead $\mathcal{L}_\mathcal{N}$ sends a suitable parametrization of a set that contains all the possible values of its data with a desired level of probability $\tilde{\alpha}_\mathcal{N}$. By considering a simple family of sets, such as for instance boxes in $\mathbb{R}^{n_\mathcal{N}}$, communication cost can be also kept at reasonable levels. We refer to this scheme as a *privatized communication protocol* between agents.

To this end, the neighboring agent $\mathcal{L}_\mathcal{N}$ has to prepare two set of samples using its adjacent control inputs presented in 7 as follows:

$$\begin{cases} \tilde{D}_\mathcal{N} := \{\zeta\} \cup \tilde{\mathcal{Z}} \\ \tilde{D}'_\mathcal{N} := \{\zeta'\} \cup \tilde{\mathcal{Z}}' \end{cases} \tag{13}$$

where ζ and ζ' are built with the relevant components of the two outputs $y_\mathcal{N}$ and $y'_\mathcal{N}$, respectively.

They are generated using the two adjacent control inputs $u_{\mathcal{N},k-1}$ and $u'_{\mathcal{N},k-1}$, respectively, such that $y_{\mathcal{N},k} := g_\mathcal{N}(\psi_{\mathcal{N},k-1}, u_{\mathcal{N},k-1})$ and $y'_{\mathcal{N},k} := g_\mathcal{N}(\psi_{\mathcal{N},k-1}, u'_{\mathcal{N},k-1})$. In particular, for generating ζ' the agent $\mathcal{L}_\mathcal{N}$ may use a model of the dynamics of the system $\mathcal{S}_\mathcal{N}$, as outlined in Fig. 4. The two sets

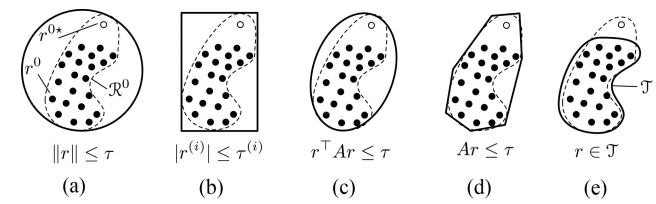


Fig. 3. Pictorial, intuitive comparison of different robust threshold and residual evaluation approaches. Representative normal values r^0 of the residual are drawn as filled black circles, while rare ones $r^{0\star}$ are drawn as empty circles. For convenience, in all cases the evaluation condition is represented as membership in a set drawn with a tick line. (a) Norm based [5], (b) Limit checking [6], (c) Ellipsoid [28], (d) Polytope [18] (e) The proposed, probabilistic set-based approach.

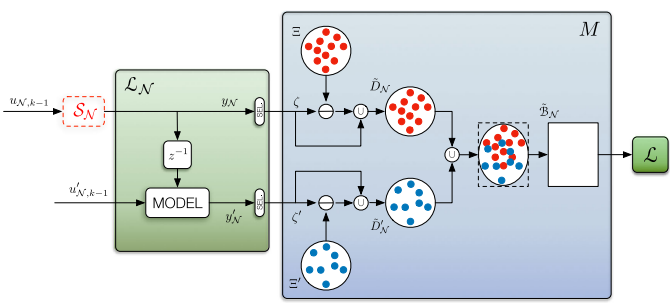


Fig. 4. Depiction of the proposed privatized, set-based communication scheme. By using a model of $\mathcal{S}_\mathcal{N}$ dynamics, the agent $\mathcal{L}_\mathcal{N}$ can generate the value ζ' of the interconnection variable measurements under the hypothesis that $\mathcal{S}_\mathcal{N}$ is driven by the adjacent input $u'_\mathcal{N}$. By subtracting to both ζ' and to the true ζ two sets of samples of the measuring uncertainties, the sets $\tilde{D}'_\mathcal{N}$ and $\tilde{D}_\mathcal{N}$ are obtained. The hyper-box $\tilde{\mathcal{B}}_\mathcal{N}$ is then computed as the smallest one containing both and its bounds are communicated to the agent $\mathcal{L}$.

$\tilde{\mathcal{Z}}$ and $\tilde{\mathcal{Z}}'$ can be also built via

$$\begin{cases} \tilde{\mathcal{Z}} := \{\zeta\} \ominus \Xi = \{z^1, \ldots, z^{\tilde{N}_s}\} \\ \tilde{\mathcal{Z}}' := \{\zeta'\} \ominus \tilde{\Xi}' = \{z'^1, \ldots, z'^{\tilde{N}'_s}_\mathcal{N}\} \end{cases} \tag{14}$$

where $\tilde{\Xi} = \{\Xi^1, \ldots, \Xi^{\tilde{N}_s}\}$ and $\tilde{\Xi}'_N = \{\Xi^1, \ldots, \Xi^{\tilde{N}'_s}\}$ such that the number of samples $\tilde{N}_s$ is greater than or equal to $\tilde{N}'_s$.

It is important to notice that, in the privatized communication protocol, both the number $\tilde{N}_s$ and $\tilde{N}'_s$ of samples generated by $\mathcal{L}_\mathcal{N}$ may be different from that needed by $\mathcal{L}$, which is N_s, as will be explained later. For sake of simplicity, we denote $\tilde{d}_i$ and $\tilde{d}'_i$ as an element of the database $\tilde{D}_\mathcal{N}$ and $\tilde{D}'_\mathcal{N}$.

Let us then introduce $\mathcal{B}_\mathcal{N} \subset \mathbb{R}^{n_\mathcal{N}}$ as a bounded set containing all the elements of $\tilde{D}_\mathcal{N}$ and $\tilde{D}'_\mathcal{N}$. We assume for simplicity that $\mathcal{B}_\mathcal{N}$ is an axis-aligned hyper-rectangular set. This is not a restrictive assumption and any convex set could have been chosen instead as in [41]. We can so define $\mathcal{B}_\mathcal{N}$ as the Cartesian product of $n_\mathcal{N}$ intervals of the type $[-b^{(i)}_\mathcal{N}, b^{(i)}_\mathcal{N}]$, where $i = 1, \ldots, n_\mathcal{N}$ and the vector $b_\mathcal{N} \in \mathbb{R}^{n_\mathcal{N}}$ defines the hyper-rectangle bounds. For convenience, we will introduce the shorthand notation $\mathcal{B}_\mathcal{N} = [-b_\mathcal{N}, b_\mathcal{N}]$.

Consider now the following optimization problem that aims to determine the set $\mathcal{B}_\mathcal{N}$ with minimal volume:

$$(\mathcal{P}^\mathcal{C}_\mathcal{N}) \begin{cases} \min\limits_{b_\mathcal{N}} & \|b_\mathcal{N}\|_1 \\ \text{s.t.} & \tilde{d}_i \in [-b_\mathcal{N}, b_\mathcal{N}], \forall \tilde{d}_i \in \tilde{D}_\mathcal{N}, i = 1, \ldots, \tilde{N}_s+1 \cdot \\ & \tilde{d}'_i \in [-b_\mathcal{N}, b_\mathcal{N}], \forall \tilde{d}'_i \in \tilde{D}'_\mathcal{N}, i = 1, \ldots, \tilde{N}'_s+1 \end{cases}$$

If we denote by $\tilde{\mathcal{B}}_\mathcal{N} = [-\tilde{b}_\mathcal{N}, \tilde{b}_\mathcal{N}]$ the optimal solution of $(\mathcal{P}^\mathcal{C}_\mathcal{N})$ computed by the neighbor agent $\mathcal{L}_\mathcal{N}$, then for implementing the privatized communication protocol the latter needs to communicate to agent $\mathcal{L}$ only the vector $\tilde{b}_\mathcal{N}$ along with the probability of violation (*the level of reliability*) $\tilde{\alpha}_\mathcal{N}$. The level of reliability $\tilde{\alpha}_\mathcal{N}$ can be determined as a direct application of the scenario approach theory in [42], leading to the following result.

Theorem 2: Fix $\tilde{\beta}_\mathcal{N} \in (0, 1)$ and let

$$\tilde{\alpha}_\mathcal{N} = \sqrt[\tilde{N}_s+1-n_\mathcal{N}]{\frac{\tilde{\beta}_\mathcal{N}}{\binom{\tilde{N}_s+1}{n_\mathcal{N}}}}. \tag{15}$$

We then have

$$\mathbb{P}^{\tilde{N}_s+1}\left\{\{\tilde{d}_1, \ldots, \tilde{d}_{\tilde{N}_s+1}\} \in \tilde{D}_\mathcal{N}^{\tilde{N}_s+1} : \mathbb{P}\left\{\tilde{d} \in \tilde{D}_\mathcal{N} : \tilde{d} \notin [-\tilde{b}_\mathcal{N}, \tilde{b}_\mathcal{N}]\right\} \leq 1 - \tilde{\alpha}_\mathcal{N}\right\} \leq \tilde{\beta}_\mathcal{N}. \tag{16}$$

Proof: (15) is a direct result of the scenario approach theory in [42], if $\tilde{\beta}_\mathcal{N}$ is chosen such that

$$\binom{\tilde{N}_s+1}{n_\mathcal{N}} \tilde{\alpha}_\mathcal{N}^{\tilde{N}_s+1-n_\mathcal{N}} \leq \tilde{\beta}_\mathcal{N}.$$

Considering the worst-case equality in the above relation and some algebraic manipulations, one can obtain the above assertion. □

Theorem 2 implies that given a hypothetical new privatized sample $\tilde{d}$, we have a confidence of at least $1 - \tilde{\beta}_\mathcal{N}$ that the probability of it belonging to $\tilde{\mathcal{B}}_\mathcal{N} = [-\tilde{b}_\mathcal{N}, \tilde{b}_\mathcal{N}]$ is at least $\tilde{\alpha}_\mathcal{N}$. In other words, the optimal set $\tilde{\mathcal{B}}_\mathcal{N}$ is an $\tilde{\alpha}_\mathcal{N}$*−probabilistic approximation of the set* $\tilde{D}_\mathcal{N}$. Therefore, one can *rely on* $\tilde{\mathcal{B}}_\mathcal{N}$ *up to* $\tilde{\alpha}_\mathcal{N}$ *probability*.

Remark 4: The number of samples $\tilde{N}_s$ in the proposed formulation $(\mathcal{P}^\mathcal{C}_\mathcal{N})$ is a design parameter chosen by the neighboring agent $\mathcal{L}_\mathcal{N}$. We however remark that one can also set a given $\tilde{\alpha}_\mathcal{N}$ as the desired level of reliability and obtain from (15) the required number of samples $\tilde{N}_s$.

Another important property of the optimal set $\tilde{\mathcal{B}}_\mathcal{N}$ is presented in the following corollary.

Corollary 1: Given $\tilde{\beta}_\mathcal{N} \in (0, 1)$ and let

$$\tilde{\alpha}'_\mathcal{N} = \sqrt[\tilde{N}'_s+1-n_\mathcal{N}]{\frac{\tilde{\beta}_\mathcal{N}}{\binom{\tilde{N}'_s+1}{n_\mathcal{N}}}}. \tag{17}$$

We then have

$$\mathbb{P}^{\tilde{N}'_s+1}\left\{\{\tilde{d}'_1, \ldots, \tilde{d}_{\tilde{N}'_s+1}\} \in \tilde{D}'^{\tilde{N}'_s+1}_\mathcal{N} : \mathbb{P}\left\{\tilde{d}' \in \tilde{D}'_\mathcal{N} : \tilde{d}' \notin [-\tilde{b}_\mathcal{N}, \tilde{b}_\mathcal{N}]\right\} \leq 1 - \tilde{\alpha}'_\mathcal{N}\right\} \leq \tilde{\beta}_\mathcal{N}. \tag{18}$$

Proof: The proof is similar to the proof of Theorem 2 and therefore for sake of clarity is omitted here. □

The interpretation of the above corollary together with Theorem 2 is as follows. The optimal set $\tilde{\mathcal{B}}_{\mathcal{N}}$ contains both distributions, $\mathbb{P}\{\tilde{d} \in \tilde{D}_{\mathcal{N}} : \tilde{d} \in [-\tilde{b}_{\mathcal{N}}, \tilde{b}_{\mathcal{N}}]\}$ and $\mathbb{P}\{\tilde{d}' \in \tilde{D}'_{\mathcal{N}} : \tilde{d}' \in [-\tilde{b}_{\mathcal{N}}, \tilde{b}_{\mathcal{N}}]\}$, with desired levels, $\tilde{\alpha}_{\mathcal{N}}$ and $\tilde{\alpha}'_{\mathcal{N}}$, and high confidence level, $1 - \tilde{\beta}_{\mathcal{N}}$.

Based on this property of optimal set $\tilde{\mathcal{B}}_{\mathcal{N}}$ in the following section, we present that the proposed communication scheme is a novel differential privacy mechanism.

Remark 5: The condition to have $\tilde{N}_s$ greater than or equal to $\tilde{N}'_s$, which leads to have $\tilde{\alpha}_{\mathcal{N}} \geq \tilde{\alpha}'_{\mathcal{N}}$, is not restrictive for the proposed communication scheme, and can be easily removed by considering $\tilde{\alpha}'_{\mathcal{N}}$ as the level of reliability for the optimal set $\tilde{\mathcal{B}}_{\mathcal{N}}$ when $\tilde{N}_s$ is less than or equal to $\tilde{N}'_s$. In case $\tilde{N}_s$ is equal to $\tilde{N}'_s$, then $\tilde{\alpha}_{\mathcal{N}}$ would also be equal to $\tilde{\alpha}'_{\mathcal{N}}$. In Corollary 1 we consider to have the same $\tilde{\beta}_{\mathcal{N}} \in (0, 1)$ as in Theorem 2 for sake of simplicity. One can also consider a different parameter, e.g., $\tilde{\beta}'_{\mathcal{N}} \in (0, 1)$.

B. Differentially Private Communication Scheme

DP 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 [22].

Definition 8: Given $\epsilon \geq 0$ as the desired level of privacy, a randomized mechanism M preserves $\epsilon-$differential privacy if for all $\mathcal{R} \subset \text{range}(M)$ and all adjacent databases D and D' in $\mathcal{D}$, it holds that

$$\mathbb{P}\left[M(D) \in \mathcal{R}\right] \leq e^{\epsilon}\, \mathbb{P}\left[M(D') \in \mathcal{R}\right]. \tag{19}$$

Remark 6: 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 8 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.

We are now in a position to show that the proposed inter-agent communication scheme, which is depicted in Fig. 4, is indeed a differentially private mechanism in the following theorem.

Theorem 3: Given $\tilde{\alpha}_{\mathcal{N}} \geq \tilde{\alpha}'_{\mathcal{N}}$, then the optimal set $\tilde{\mathcal{B}}_{\mathcal{N}}$ obtained via the optimization problem $(\mathcal{P}^{\mathcal{C}}_{\mathcal{N}})$ has the following property with high confidence level, $1 - \tilde{\beta}_{\mathcal{N}}$:

$$\mathbb{P}\left[\tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\right] \leq e^{\epsilon}\, \mathbb{P}\left[\tilde{d}' \in \tilde{\mathcal{B}}_{\mathcal{N}}\right] \tag{20}$$

where ϵ is the desired level of privacy and obtained using

$$\epsilon = \ln(\tilde{\alpha}_{\mathcal{N}}) - \ln(\tilde{\alpha}'_{\mathcal{N}}). \tag{21}$$

Proof: The proof is straightforward by substituting ϵ into the assertion of theorem and some algebraic manipulations as follows:

$$\mathbb{P}\left[\tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\right] \leq e^{\epsilon}\, \mathbb{P}\left[\tilde{d}' \in \tilde{\mathcal{B}}_{\mathcal{N}}\right]$$

$$\frac{1}{e^{\ln(\tilde{\alpha}_{\mathcal{N}}) - \ln(\tilde{\alpha}'_{\mathcal{N}})}} \leq \frac{\mathbb{P}\left[\tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\right]}{\mathbb{P}\left[\tilde{d}' \in \tilde{\mathcal{B}}_{\mathcal{N}}\right]}$$

$$\frac{\tilde{\alpha}'_{\mathcal{N}}}{\tilde{\alpha}_{\mathcal{N}}} \leq \frac{\mathbb{P}\left[\tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\right]}{\mathbb{P}\left[\tilde{d}' \in \tilde{\mathcal{B}}_{\mathcal{N}}\right]}$$

where the last line of the above equations is the ratio between the results obtained in Theorem 2 and Corollary 1 with high confidence level $1 - \tilde{\beta}_{\mathcal{N}}$. The proof is completed by noting that the optimal set $\tilde{\mathcal{B}}_{\mathcal{N}}$ is a random set as it is obtained via the optimization problem $(\mathcal{P}^{\mathcal{C}}_{\mathcal{N}})$ which depends on data sets $\tilde{D}_{\mathcal{N}}$ and $\tilde{D}'_{\mathcal{N}}$. □

The following corollary is a direct result of Theorem 3.

Corollary 2: The proposed privatized communication scheme is immune to postprocessing.

Proof: The proof is similar to the proof of [23, Proposition 2.1] and therefore for sake of clarity is omitted here. □

Remark 7: In the current literature on DP the so-called Laplacian mechanism is almost always used, which is based on additive noise whose magnitude is dependent on the query *sensitivity* [23]. The sensitivity depends on the worst case effect on the query value caused by using an adjacent input. Instead, the mechanism proposed in the present article does not require the computation of the sensitivity, nor to consider the worst case. Indeed, the current approach again requires only the capability of computing samples of the result of the query for the case of the adjacent input. This leads to the DP property (20) holding with a given confidence level which depends on the number of samples generated for the two hypotheses.

After receiving the parametrization of $\tilde{\mathcal{B}}_{\mathcal{N}}$ and the level of reliability $\tilde{\alpha}_{\mathcal{N}}$, agent $\mathcal{L}$ can then obtain the samples needed for computing its threshold by locally generating $N_s + 1$ samples, drawing them uniformly from inside $\tilde{\mathcal{B}}_{\mathcal{N}}$. Then, it should designate, using an arbitrary policy, one of them as the value $\tilde{y}_{\mathcal{N}}$ to use for the interconnection variable measurement $y_{\mathcal{N}}$. The remaining N_s ones would be used as values of the samples for $x_{\mathcal{N}}$.

In this way, we decoupled the sample generation of $\mathcal{L}_{\mathcal{N}}$ from the one of $\mathcal{L}$, preserving also the privacy of the former. We however note that the proposed privatized communication protocol introduces some level of stochasticity on the probabilistic threshold design of agent $\mathcal{L}$, due to the fact that the neighboring information is *probabilistically reliable*. In the following section, we characterize the threshold set probabilistic robustness as in Definition 6, and provide a new level of probability for the threshold design in order to accommodate this new situation.

C. Privatized Distributed Probabilistic Threshold Set

When an agent $\mathcal{L}$ and its neighbor $\mathcal{L}_{\mathcal{N}}$ adopt the privatized communication scheme we proposed in the previous section, there is an important effect on the local probabilistic threshold set $\mathcal{T}$ computed by agent $\mathcal{L}$. Such a scheme introduces an additional level of stochasticity, as the set $\tilde{\mathcal{B}}_{\mathcal{N}}$ which is a probabilistic approximation, is communicated instead of the N_s samples that would have been sent in the hard communication scheme.

This will affect the local threshold set probabilistic robustness guarantee, as explained in the following theorem. To accommodate the level of reliability of neighboring information, we need to marginalize the joint cumulative distribution function probability of the residual value at time step $k+1$ and the generic sample $\tilde{d}$ appearing in Theorem 2.

Theorem 4: Given $\tilde{\alpha}_{\mathcal{N}} \in (0,1]$ and a fixed $\alpha \in (0,1]$, then following Definition 6, the adaptive threshold set $\mathcal{T}_k$ is probabilistically $\bar{\alpha}$–robust with respect to the random total uncertainty $\delta_k \in \Delta_k^0$, i.e.,

$$\mathbb{P}\left[r_{k+1} \in \mathcal{T}_{k+1}\right] \geq \bar{\alpha} \tag{22}$$

where $\bar{\alpha} = 1 - \frac{1-\alpha}{\tilde{\alpha}_{\mathcal{N}}}$, and for all $r_{k+1} \in \mathcal{R}_{k+1}^0$.

Proof: The proof is provided in the Appendix. □

Theorem 4 provides a new level of robustness for the threshold set $\mathcal{T}_k$ computed by agent $\mathcal{L}$. It is straightforward to observe that if $\tilde{\alpha}_{\mathcal{N}} \to 1$ then $\bar{\alpha} \to \alpha$. This means that if *the level of reliability* of the neighboring information is one, $\mathbb{P}[\,\tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\,] = 1$, then, the designed threshold set will have the same level of probabilistic robustness as the hard communication scheme, $\mathbb{P}[\,r_{k+1} \in \mathcal{T}_{k+1}\,] \geq \alpha$.

It is important to note that the proposed steps for the probabilistic threshold set design that we presented in Section II are directly applicable to the results that we obtained in Theorem 4. This is due to the fact that the proposed scheme in Section II is independent from the privatized communication scheme is used between neighboring agents.

D. Summary of Main Contributions

We will now summarize the main contributions of this article by presenting them as a list of steps needed to successfully design a privacy-preserving distributed anomaly scheme. This would provide a support in following the numerical studies presented in the next section.

(C1) *Adaptive parametrized probabilistic threshold set design:* As our first contribution presented in Section II, we devise a set-based probabilistic scheme where the probability of false alarms is defined as a user-tunable design parameter, and the detection rate with respect to a given class of faults is simultaneously maximized. Through this scheme, the conservativeness of existing threshold designs is reduced by relaxing the deterministic robust zero-FAR condition, in favor of a more flexible, *probabilistic* one. The scheme is summarized as follows. As a first step, each local agent $\mathcal{L}$ should design its own threshold set, which requires the completion of the following sub-steps:

- (C1-1) Choose a desired probabilistic level α for the threshold robustness and a desired confidence level $(1-\beta)$. Then, make use of the novel Theorem 1 to obtain the number N_s of required samples of the healthy residual.
- (C1-2) Generate N_s samples of healthy residuals from the set $\mathcal{R}_{k+1}^0$, by using eqs. (4) and (5).
- (C1-3) Solve the novel randomized problem (10) and determine the parameters of the optimal threshold set $\mathcal{T}_{k+1}^*$. With confidence level $(1-\beta)$ this set has a probability of robustness equal to α, and is the one that maximizes detectability in these conditions.

In Step (C1-2) agent $\mathcal{L}$ needs to know the current measured value ζ_k of the interconnection variables, as well as samples of the neighbour measurement uncertainties. This requires that the neighboring agent $\mathcal{L}_{\mathcal{N}}$ computes such samples and communicates them to $\mathcal{L}$ along with ζ_k. As this would raise privacy concerns for $\mathcal{L}_{\mathcal{N}}$, the following privacy-preserving communication scheme is used instead.

(C2) *Privatized probabilistic set-based communication scheme*

As our second contribution presented in Section III, we propose a privatized communication scheme that neither relies on the classic additive Laplacian noise mechanism, nor requires the computation of the query sensitivity. Instead, it is based on LDs communicating the parametrization of randomized sets in a way that is guaranteed to satisfy the differential privacy condition with a given confidence level. Through this scheme, the quantity of data that neighboring LDs need to communicate at each sampling time is reduced, and the results also provide a theoretical connection between the performances of the distributed anomaly detection scheme and the desired level of privacy. The scheme is summarized as follows.

The agent $\mathcal{L}_{\mathcal{N}}$ should perform locally the following steps, before sending the results to agent $\mathcal{L}$

- (C2-1) Choose a suitable level of reliability $\tilde{\alpha}_{\mathcal{N}}$ and determine the number of samples $\tilde{N}_s$ using the results of the novel Theorem 2.
- (C2-2) Fix the desired level of privacy ϵ and determine $\tilde{\alpha}'_{\mathcal{N}}$ using the results of the novel Theorem 3, i.e., $\tilde{\alpha}'_{\mathcal{N}} = \tilde{\alpha}_{\mathcal{N}}\, e^{-\epsilon}$.
- (C2-3) Determine the number of samples $\tilde{N}'_s$ using $\tilde{\alpha}'_{\mathcal{N}}$ and the results of Corollary 1.
- (C2-4) Construct two sets of the measuring uncertainties, corresponding to two adjacent local inputs: $\tilde{D}_{\mathcal{N}}$ and $\tilde{D}'_{\mathcal{N}}$. Such sets will consist of, respectively, $\tilde{N}'_s$ and $\tilde{N}'_s$ samples, following (12).
- (C2-5) Obtain the optimal hyper-box $\tilde{\mathcal{B}}_{\mathcal{N}}^*$ by solving the minimum-volume problem $\mathcal{P}_{\mathcal{N}}^{\mathcal{C}}$.
- (C2-6) Communicate the parametrization of $\tilde{\mathcal{B}}_{\mathcal{N}}^*$ and the level of reliability $\tilde{\alpha}_{\mathcal{N}}$ together with the level of privacy ϵ to agent $\mathcal{L}$.

Once the agent $\mathcal{L}$ receives such data from $\mathcal{L}_{\mathcal{N}}$, it can draw from $\tilde{\mathcal{B}}_{\mathcal{N}}^*$ the necessary samples for carrying out the step (C1-2) and compute its local threshold $\mathcal{T}_{k+1}^*$. By using the novel Theorem 4, the agent $\mathcal{L}$ can check if the new probability of robustness $\bar{\alpha}$ for the threshold $\mathcal{T}_{k+1}^*$ is still satisfying its requirements.

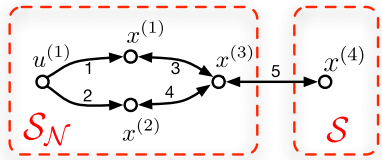


Fig. 5. Structural graph of the 4-tank system chosen for this study, which is decomposed into two systems. Levels are represented by state variables $x^{(i)}$, while pipes are named according to the structural graph edges labels. One source, not depicted, can deliver water to either tank 1 or 2 via pipes 1 or 2, according to the position of a valve controlled by the input $u^{(1)}$. One point of delivery is connected to tank 4.

IV. Numerical Study

In this section, we present two numerical studies to illustrate the effectiveness of our proposed approach. The first one shows that in the absence of any privacy mechanism, an LD may be able to detect changes in the input of other subsystems. We also verify that this can be prevented by a mechanism based on 3. The second one presents a full implementation of the proposed privacy-preserving distributed anomaly detection scheme.

A. Privacy Preservation

Let us consider the water distribution network depicted in 5 and decomposed into two subsystems: $\mathcal{S}_\mathcal{N}$, which can be thought of as a water resupply subnetwork, and $\mathcal{S}$, which is a customer of $\mathcal{S}_\mathcal{N}$ and acts as a water distribution subnetwork serving end customers connected to tank 4. The operator of $\mathcal{S}_\mathcal{N}$ can switch a valve commanded by $u^{(1)}$ in order to provide water to $\mathcal{S}_2$ either through the route $1 \to 3 \to 4$, or $2 \to 3 \to 4$. The two routes are supposed to lead to different operating costs and hence to different pricing policies that $\mathcal{S}_\mathcal{N}$ charges to $\mathcal{S}$: which route $\mathcal{S}_\mathcal{N}$ is operating at a given moment is considered a private information. Anomalies are not considered and subsystems dynamics can be described via

$$\mathcal{S}_\mathcal{N} : \begin{cases} x_+^{(1)} = g_\mathcal{N}^{(1)}(x_\mathcal{N}, u_\mathcal{N}, z_\mathcal{N}) \\ \quad = x^{(1)} + \frac{T}{A_1}\left[(1-u_1)\phi_{s,1} - \phi_{1,3}\right] \\ x_+^{(2)} = g_\mathcal{N}^{(2)}(x_\mathcal{N}, u_\mathcal{N}, z_\mathcal{N}) = x^{(2)} + \frac{T}{A_2}\left[u_1\phi_{s,2} - \phi_{2,3}\right] \\ x_+^{(3)} = g_\mathcal{N}^{(3)}(x_\mathcal{N}, u_\mathcal{N}, z_\mathcal{N}) \\ \quad = x^{(3)} + \frac{T}{A_3}\left[\phi_{1,3} + \phi_{2,3} - \phi_{3,4}\right] \end{cases}$$

$$\mathcal{S} : \left\{ x_+^{(4)} = g_2^{(1)}(x_2, x_{\mathcal{N}_2}, w_2) = x^{(4)} + \frac{T}{A_4}\left[\phi_{3,4} - w_2\right] \right.$$

where the index "+" is a shorthand to refer to the next time steps, A_i denotes the cross-section of the i–th tank, T the sampling interval. The input $u_\mathcal{N} = u^{(1)} \in \{0,\ 1\}$ denotes the position of a valve that can either connect tank 1, when $u^{(1)} = 0$, or tank 2, when $u^{(1)} = 1$, to a constant pressure water source serving $\mathcal{S}_\mathcal{N}$, which is equivalent to an infinite tank at a constant level x_s. The symbol $\phi_{a,b}$ denotes the flow from tank a to tank b along the pipe connecting them defined as [43]

$$\phi_{s,1} = \max\left(0, \operatorname{sign}(x_s - x^{(1)})\sqrt{2g|x_s - x^{(1)}|}c_1\right)$$

$$\phi_{s,2} = \max\left(0, \operatorname{sign}(x_s - x^{(2)})\sqrt{2g|x_s - x^{(2)}|}c_2\right)$$

$$\phi_{1,3} = \operatorname{sign}(x^{(1)} - x^{(3)})\sqrt{2g|x^{(1)} - x^{(3)}|}c_3$$

$$\phi_{2,3} = \operatorname{sign}(x^{(2)} - x^{(3)})\sqrt{2g|x^{(2)} - x^{(3)}|}c_4$$

$$\phi_{3,4} = \operatorname{sign}(x^{(3)} - x^{(4)})\sqrt{2g|x^{(3)} - x^{(4)}|}c_5$$

where c_j is the cross section of the j–th pipe and g the gravitational acceleration. Finally, $w_2(t)$ is an unknown external signal representing the time-varying demand of end users. $\mathcal{L}_\mathcal{N}$ shall measure the local state $x_\mathcal{N} = [x^{(1)}\, x^{(2)}\, x^{(3)}]^\top$ and send to $\mathcal{L}$ the privatized set $\tilde{\mathcal{B}}_\mathcal{N}$ from which $\mathcal{L}$ can select the value $\tilde{\zeta}$ for the interconnection variable $z = [x^{(3)}]$, as explained in Section III-B.

Using the adversary model presented in Remark 3, which aims at compromising the privacy of the agents' local inputs, we now propose an approach through which $\mathcal{L}$ can breach the privacy of $\mathcal{S}_\mathcal{N}$ by reconstructing the current value of its local input $u^{(1)}$, and show how our proposed mechanism can be used to prevent this.

First of all $\mathcal{L}$ needs to know a model of $\mathcal{S}_\mathcal{N}$ dynamics, which in general can be affected by uncertainty in the knowledge of $\mathcal{S}_\mathcal{N}$ parameters and/or structure. Here only parametric uncertainty is assumed. $\mathcal{L}$ can therefore breach $\mathcal{S}_\mathcal{N}$ privacy by using such model and ζ measurements to test the hypothesis "$\mathcal{H} :\ u^{(1)}$ is equal to 0" against the hypothesis "$\mathcal{H}' :\ u^{(1)}$ is equal to 1".

To this purpose, $\mathcal{L}$ will implement the following estimators:

$$\hat{\mathcal{S}}_\mathcal{N} : \begin{cases} \hat{x}_{\mathcal{N},+} = \hat{g}_\mathcal{N}(\hat{x}_\mathcal{N}, 0, z_\mathcal{N}) + \Lambda(\hat{y}_\mathcal{N} - \tilde{y}_\mathcal{N}) \\ \hat{y}_\mathcal{N} = \hat{x}_\mathcal{N}^{(3)} \end{cases} \tag{23}$$

$$\hat{\mathcal{S}}'_\mathcal{N} : \begin{cases} \hat{x}'_{\mathcal{N},+} = \hat{g}_\mathcal{N}(\hat{x}'_\mathcal{N}, 1, z_\mathcal{N}) + \Lambda(\hat{y}'_\mathcal{N} - \tilde{y}_\mathcal{N}) \\ \hat{y}'_\mathcal{N} = \hat{x}'^{(3)}_\mathcal{N} \end{cases} \tag{24}$$

where $\hat{g}_\mathcal{N}$ represents the uncertain $\mathcal{S}_\mathcal{N}$ dynamics model employed by $\mathcal{L}$. The variables $\hat{x}_\mathcal{N}$ and $\hat{x}'_\mathcal{N}$ and, respectively, $\hat{y}_\mathcal{N}$ and $\hat{y}_{\mathcal{N}'}$ are estimates of $\mathcal{S}_\mathcal{N}$ states and of the interconnection variable computed by $\mathcal{L}$ under the two hypotheses. By comparing the absolute value of the scalar residuals $r := \tilde{y}_\mathcal{N} - \hat{y}_\mathcal{N}$ and $r' := \tilde{y}_\mathcal{N} - \hat{y}'_\mathcal{N}$ to a fixed scalar threshold τ, the two hypotheses can be tested.

We can cast this hypothesis testing problem by defining the adjacent databases $D := \{0\}$ and $D' := \{1\}$ such that u_1 will belong at any time to only one of them. By applying the postprocessing property in Corollary 2, we can consider the query results to be the residual $r = M(D)$, and $r' = M(D')$, which indeed depends on $u^{(1)}$ belonging to D or D' through the composition of randomized mappings.

The DP condition in Definition 8 can then be checked by defining the test set $\mathcal{R} := [0\ \tau]$ and evaluating numerically the probabilities $\mathbb{P}[M(D) \in \mathcal{R}]$ and $\mathbb{P}[M(D') \in \mathcal{R}]$ and whether condition (19) holds.

In the following, we will compare the case where no privacy mechanism is applied, that is when z is directly communicated to $\mathcal{L}$, to the case when the proposed mechanism is applied (see

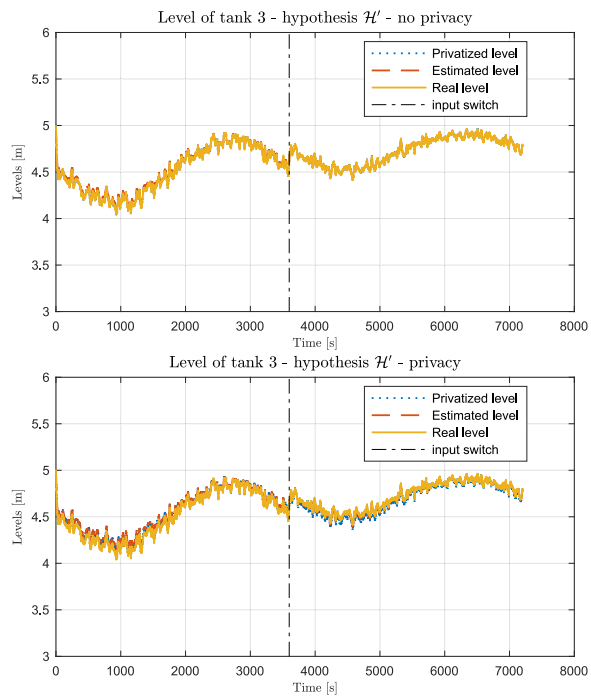


Fig. 6. Third tank estimated level and residual computed by $\mathcal{L}$ for testing hypothesis $\mathcal{H}'$, without and with privacy mechanism applied to communication from $\mathcal{L}_\mathcal{N}$ to $\mathcal{L}$.

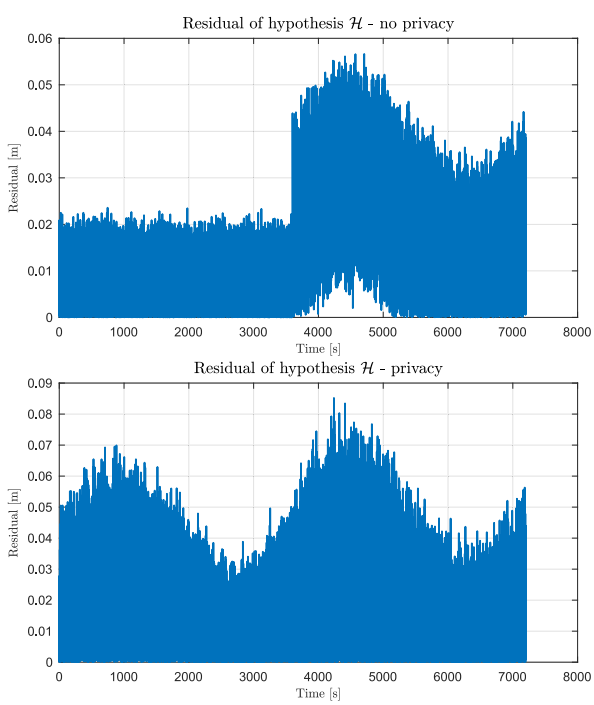


Fig. 7. Third tank residual computed by $\mathcal{L}$ for testing hypothesis $\mathcal{H}$, without and with privacy mechanism applied to communication from $\mathcal{L}_\mathcal{N}$ to $\mathcal{L}$.

C2 in Section III-D). We will test in simulation the effect of varying the value of the threshold τ and varying the ratio $\frac{\tilde{N}'_s}{\tilde{N}_s}$ between the number of samples for the alternate and for the true hypothesis that are used by $\mathcal{L}$ when computing the set $\tilde{\mathcal{B}}_\mathcal{N}$.

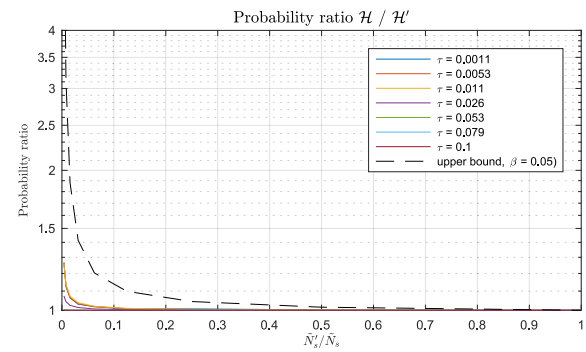


Fig. 8. Analysis of the ratio between $\mathbb{P}[M(D) \in \mathcal{R}]$ and $\mathbb{P}[M(D') \in \mathcal{R}]$. The theoretical upper bound for a given confidence level is plotted in dashed black.

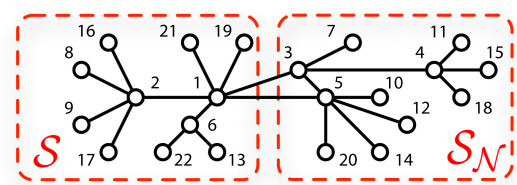


Fig. 9. Structural graph of the 22-tank system chosen for the numerical study, decomposed into two subsystems. The nodes are labelled by an index and the interconnection between the two is represented by the two edges (1, 3) and (1, 5).

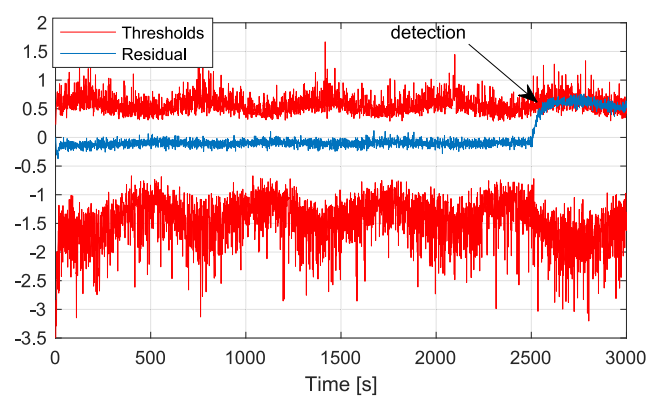


Fig. 10. Projection of the residual and the threshold set of agent $\mathcal{L}$ on component number 1, as a function of time.

The tank cross sections will be set to 1, 1, 5, and 2 m^2, and the pipe cross sections to 0.25, 0.25, 0.2, 0.5 and 0.2 m^2. It can thus be seen that the two water supply routes on which $\mathcal{S}_1$ can operate differ only in the cross section of the pipe feeding the third tank.

The unknown user demand w_2 is assumed to follow the expression $0.6 + 0.25 * \sin(2\pi/T_d t) + w_d$, where $T_d = 2$ hours is the demand periodicity and w_d is a random number obtained by sampling every 15 s from a normal distribution with zero mean and variance equal to 0.25.

The sampling time is $T = 0.1$ s and $\mathcal{S}_1$ water supply policy is to use the first route during the first half of every period, and the second in the second half. Finally, the model used by $\mathcal{L}_2$ is

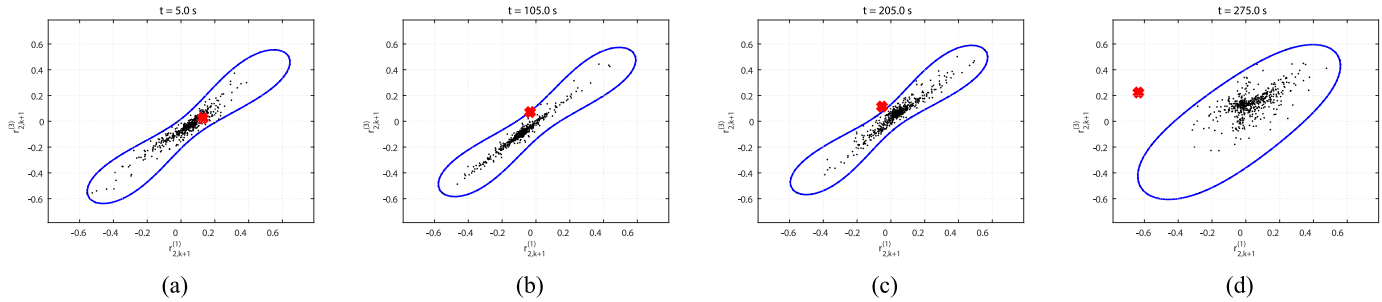


Fig. 11. Projection of the residual and the threshold set of agent $\mathcal{L}_{\mathcal{N}}$ on component number 3 and 5, at various instants in time. At each sampling period, the healthy residuals and the threshold set are shown with black dots and blue line, respectively, and the actual residual is presented via red cross symbol.

affected by a random parametric uncertainty with a maximum magnitude equal to 1% of the nominal values.

Fig. 6 shows the privatized level $\tilde{y}_{\mathcal{N}}$, the estimated level $\hat{y}'_{\mathcal{N}}$ under hypothesis $\mathcal{H}'$ and the real level for tank 3 when no privacy mechanism is in place and when it is. While the presence of the mechanism is inducing a difference in the estimated level, it is indeed negligible which is a good indication of the low performance deterioration in the diagnosis task that would be induced by the privacy.

Similarly, Fig. 6 shows the residual computed by $\mathcal{L}_2$ in the same cases, but for hypothesis $\mathcal{H}$. As it can be noticed, any static threshold τ in the range 0.025 to 0.05 will lead to a successful testing of this hypothesis when a privacy mechanism is not present. When it is, instead, the choice of a suitable τ is severely limited. A more rigorous verification of this assertion can be attained by looking at Fig. 8. Here we computed numerically the ratio between $\mathbb{P}[M(D) \in \mathcal{R}]$ and $\mathbb{P}[M(D') \in \mathcal{R}]$ during the simulation period for which $u^{(1)} = 1$, for several values of τ and as a function of the ratio $\frac{\tilde{N}'_s}{\tilde{N}_s}$. The plots of these ratios are compared to the term e^{ϵ} in order to check whether (19) is satisfied, with ϵ computed for a given confidence level $\tilde{\beta}_{\mathcal{N}}$ as stated in Theorem 3, Corollary 1, and Theorem 4. As it can be seen, the theoretical upper bound plotted in Fig. 8 is always satisfied, for all the considered values of τ and of the samples ratio. It is interesting to note how the probability ratio tends to unity when the samples ratio does the same. This means that when the two hypotheses are equally represented when $\mathcal{L}_1$ computes the set $\tilde{\mathcal{B}}_{\mathcal{N}}$, then $\mathcal{L}_2$ cannot successfully test them as they become probabilistically equivalent. As shown in the left part of the plot, only for low values of the samples ratio $\mathcal{L}_2$ can distinguish the two hypotheses, as their respective probabilities have a ratio sufficiently different than 1.

B. Privatized Distributed Anomaly Detection

In this example, a 22 tanks system is considered (Fig. 9), decomposed into two subsystems. Its structural graph has been obtained by application of the Barabási-Albert model [44], and finally an edge between nodes 1 and 3 have been added in order to introduce an asymmetry.

The actual tank cross sections have been chosen equal to $1\,\mathrm{m}^2$, while pipe cross sections are 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 Hz, has been instead connected to tank number 1. All tank levels are assumed to be measured, with a Gaussian measurement uncertainty with zero mean and a standard deviation equal to 0.05 m. When building the LD estimators, a Gaussian parametric uncertainty is introduced, having zero mean and a variance equal to 5% and 7.5%, respectively, of the tanks and pipe cross sections. The privacy mechanism M has been generated using $\tilde{N}_s = \tilde{N}'_s = 16$. Finally, to implement the detection scheme C1 summarized in Section III-D, each LD will generate $N_s = 512$ samples for computing their threshold sets, using a fourth-order polynomial as an indicator function.

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 only interconnection variables, and as such it may be hidden, that is made undetectable, by the introduction of the privacy mechanism. The following figures will present the results obtained by simulating such fault occurring at time $T_f = 250$ s. In order to make it possible to represent graphically the 11-dimensional residuals and threshold sets for the two LDs, we have chosen to consider only their projection on the multitank components numbers 1, 3, and 5, respectively. As only the components numbers 1 and 3 will be affected by the fault, this is not going to hide any information, with component number 5 presented only for reference.

In Fig. 10, the residual and the threshold set of the first agent, projected on the component corresponding to tank number 1, are depicted. As in this case we are considering only one dimension, the residual can be plotted as a curve, and the threshold set projection, being a time-varying interval, can be represented by plotting two more curves corresponding to its bounds. Detection is successfully achieved shortly after the fault time.

In Fig. 11, we instead depict the residual and threshold set for the second agent. As in this case, we want to present their behaviour along the components corresponding to tanks numbers 3 and 5, a time-sequence of two-dimensional plots are given. Here we can notice that occasionally the residual value can fall outside the threshold set, e.g., in Fig. 11(c), as we may expect given the current probabilistic approach in designing the threshold set. After the fault time the residual is consistently outside the threshold set [see Fig. 11(d)], thus allowing for detection.

V. Concluding Remarks

In this article, we developed a probabilistic, set-based distributed anomaly detection framework for a large-scale uncertain nonlinear system. The designed threshold sets are guaranteed to be robust against uncertainties up to a user-desired probability. By using the Scenario Approach, such probability is met with a confidence level that depends on the number of uncertainty samples used during the design. On top of this framework, a novel privatized communication scheme has been proposed, that allows neighboring local detectors to exchange the data needed to compute the thresholds, while protecting the privacy of their subsystem local input. The proposed scheme is based on communicating sets that contain the results of queries based on adjacent inputs, with a given probability. As such, this scheme is not based on additive Laplacian noise nor does require the computation of query sensitivities. Theoretical results were provided to link the desired level of privacy to the loss of performance of the distributed detection scheme. As future work, we plan to address several research questions that were left unanswered. A first question regards the possibility to derive an analytical detectability theorem that characterizes the proposed detection scheme. Indeed, even if the cascaded optimization problem proposed to determine the threshold is meant to increase detectability, still what is the probability of detecting faults belonging to a given class is an open question. A second question would be directed at extending the distributed detection scheme in order to allow for anomaly isolation as well. Finally, an extended numerical study that compares the proposed approach, with and without privacy, to established deterministic and probabilistic detection schemes would allow to quantify the expected benefits of the approach.

Appendix

Proof of Theorem 1: Due to the nonconvexity introduced by the Chebyshev distance, we have to recast the second stage problem (10b) into ξ subprograms. By denoting with Ψ_j the feasible solution set of the jth subproblem it is clear that the optimizer of (10b) can be found in $\bigcup_{j=1}^{\xi} \Psi_j$ [15].

For clarity the proof will be broken down into three steps: a) application of the scenario approach of [40] to each individual subprogram; b) extension to the ξ subprograms; c) theoretical conditions for the optimizer $\upsilon^* := [\theta_b^*, \gamma^*]^\top$ to be a feasible solution of (9).

Let us now denote with $\mathcal{T}(\theta_b^*)$ the threshold set $\mathcal{T}_k$ obtained when $\mathbf{1}_{\mathcal{T}_k}$ is parameterized by a given θ_b^*, and recall that $\mathcal{V}(\mathcal{T}(\theta_b^*))$ is the violation probability as in Definition 6.

a) Applying the existing results in [40] to each subprogram, we have $\forall j \in \{1, \ldots, \xi\}$:

$$\mathbb{P}^{N_s}\left[\mathcal{V}(\mathcal{T}(\theta_{b_j}^*)) \leq 1-\alpha\right] \leq \sum_{i=0}^{\ell-1} \binom{N_s}{i} (1-\alpha)^i \alpha^{N_s-i}.$$

b) Considering that $\mathcal{V}(\mathcal{T}(\theta_b^*)) \subseteq \bigcup_{j=1}^{\xi} \mathcal{V}(\mathcal{T}(\theta_{b_j}^*))$, we can readily extend the aforesaid results to ξ subprograms as follows:

$$\begin{aligned}\mathbb{P}^N\left[\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\right] &\leq \mathbb{P}^N\left[\bigcup_{j=1}^{\xi} \mathcal{V}(\mathcal{T}(\theta_{b_j}^*)) \leq 1-\alpha\right] \\ &\leq \sum_{j=1}^{\xi} \mathbb{P}^N\left[\mathcal{V}(\mathcal{T}(\theta_{b_j}^*)) \leq 1-\alpha\right] \\ &< \xi \sum_{i=0}^{\ell-1} \binom{N}{i} (1-\alpha)^i \alpha^{N-i} \leq \beta.\end{aligned}$$

Note that the obtained bound is the desired assertion as it is stated in the theorem. However, the most important part of the proof is to extend this result to the cascade setup of the present optimization problem in (10).

c) In order to proceed let us first define another indicator function $\mathbf{1}_{\{\cdot\}} : [0,1] \mapsto \{0,1\}$ that indicates whether the inequality in its argument, which is a function of a random variable, holds or not. We now have to provide a new bound for the following N_s-fold product conditional probability $\mathbb{P}^{N_s}[\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha|\gamma^*]$ which is a random variable with respect to γ^* due to the fact that γ^* is an optimal solution of the first step optimization problem and it depends on specific random samples. To this end consider the following N_s-fold product conditional expectation problem:

$$\mathbb{E}^N\left[\mathbf{1}_{\{\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\}} \middle| \gamma^*\right] = \mathbb{P}^N\left[\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha \middle| \gamma^*\right]. \tag{25}$$

The best approximation of $\mathbb{P}^N[\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha|\gamma^*]$ is given by $\mathbb{E}^N[\mathbf{1}_{\{\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\}}|\gamma^*]$ which is a function of random variable γ^*. The best here means that one cannot do any better than this due to the fact that $\mathbb{P}^N[\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha|\gamma^*]$ is itself a function of random variable γ^*. Finally, we calculate the above quantity by the law of the unconscious [45], as follows:

$$\begin{aligned}&\mathbb{E}^N\left[\mathbb{E}^N\left[\mathbf{1}_{\{\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\}} \middle| \gamma^*\right]\right] \\ &= \sum_{\nu} \mathbb{E}^N\left[\mathbf{1}_{\{\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\}} \middle| \gamma^* = \nu\right] \mathbb{P}^N\left[\gamma^* = \nu\right] \\ &= \mathbb{E}^N\left[\mathbf{1}_{\{\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\}}\right] = \mathbb{P}^N\left[\mathcal{V}(\mathcal{T}(\theta_b^*)) \leq 1-\alpha\right]\end{aligned}$$

where the last equation is due to the partition theorem.

The proof is completed by noting that the final expression is already bounded in part (b) of the proof. □

Proof of Theorem 4: Following Definition 6, we have the following updated situation:

$$\alpha \leq \mathbb{P}\left[r_{k+1} \in \mathcal{T}_{k+1} \;\; \tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\right]$$

which is a joint probability of $r_{k+1} \in \mathcal{T}_{k+1}$ and $\tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}$. Such a joint probability can be equivalently written as a joint cumulative distribution function (CDF):

$$\begin{aligned}\alpha &\leq \mathbb{P}\left[r_{k+1} \in \mathcal{T}_{k+1} \;\; \tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}}\right] \\ &= \int_{\mathcal{T}_{k+1}} \int_{\tilde{\mathcal{B}}_{\mathcal{N}}} p(r_{k+1}, \tilde{d}) \, \mathrm{d}r_{k+1} \, \mathrm{d}\tilde{d} \\ &= F_{r_{k+1}, \tilde{d}}(\mathcal{T}_{k+1} \;\; \tilde{\mathcal{B}}_{\mathcal{N}})\end{aligned} \tag{26}$$

where $F_{r_{k+1}, \tilde{d}}(\mathcal{T}_{k+1} \;\; \tilde{\mathcal{B}}_{\mathcal{N}})$ and $p(r_{k+1}, \tilde{d})$ are a joint CDF and a joint probability density function (PDF) of r_{k+1} and $\tilde{d}$, respectively.

Our goal is to calculate:

$$\mathbb{P}\left[\, r_{k+1} \in \mathcal{T}_{k+1} \,\right] = \int_{\mathcal{T}_{k+1}} p(r_{k+1})\, \mathrm{d}r_{k+1} = F_{r_{k+1}}\left(\mathcal{T}_{k+1}\right)$$

where $p(r_{k+1})$ is the PDF of r_{k+1}. In order to transform the joint CDF into the marginal CDF of r_{k+1}, one can take the limit of the joint CDF as $\tilde{\mathcal{B}}_{\mathcal{N}}$ approaches $\mathbb{R}^{n_{\mathcal{N}}}$

$$\begin{aligned}
\mathbb{P}[\, r_{k+1} \in \mathcal{T}_{k+1} \,] &= F_{r_{k+1}}\left(\mathcal{T}_{k+1}\right) \\
&= \lim_{\tilde{\mathcal{B}}_{\mathcal{N}} \to \mathbb{R}^{n_{\mathcal{N}}}} F_{r_{k+1},\, \tilde{d}}\left(\mathcal{T}_{k+1}\ \tilde{\mathcal{B}}_{\mathcal{N}}\right) \\
&= \lim_{\tilde{\mathcal{B}}_{\mathcal{N}} \to \mathbb{R}^{n_{\mathcal{N}}}} F_{r_{k+1}\,|\,\tilde{d}}\left(\mathcal{T}_{k+1} \mid \tilde{\mathcal{B}}_{\mathcal{N}}\right) F_{\tilde{d}}\left(\tilde{\mathcal{B}}_{\mathcal{N}}\right) \\
&= F_{r_{k+1}}\left(\mathcal{T}_{k+1}\right) \lim_{\tilde{\mathcal{B}}_{\mathcal{N}} \to \mathbb{R}^{n_{\mathcal{N}}}} F_{\tilde{d}}\left(\tilde{\mathcal{B}}_{\mathcal{N}}\right) \qquad (27)
\end{aligned}$$

where the last equality is due to the independency of r_{k+1} and $\tilde{d}$. To determine $\lim_{\tilde{\mathcal{B}}_{\mathcal{N}} \to \mathbb{R}^{n_{\mathcal{N}}}} F_{\tilde{d}}\left(\tilde{\mathcal{B}}_{\mathcal{N}}\right)$, one can calculate

$$\begin{aligned}
\lim_{\tilde{\mathcal{B}}_{\mathcal{N}} \to \mathbb{R}^{n_{\mathcal{N}}}} F_{\tilde{d}}\left(\tilde{\mathcal{B}}_{\mathcal{N}}\right) &= \int_{\mathbb{R}^{n_{\mathcal{N}}}} p(\tilde{d})\mathrm{d}\tilde{d} \\
&= \int_{\mathbb{R}^{n_{\mathcal{N}}} \setminus \tilde{\mathcal{B}}_{\mathcal{N}}} p(\tilde{d})\mathrm{d}\tilde{d} + \int_{\tilde{\mathcal{B}}_{\mathcal{N}}} p(\tilde{d})\mathrm{d}\tilde{d} \\
&= \mathbb{P}\left[\, \tilde{d} \notin \tilde{\mathcal{B}}_{\mathcal{N}} \,\right] + \mathbb{P}\left[\, \tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}} \,\right] \\
&= (1 - \tilde{\alpha}_{\mathcal{N}}) + \tilde{\alpha}_{\mathcal{N}} = 1 \qquad (28)
\end{aligned}$$

where $p(\tilde{d})$ is the PDF of $\tilde{d}$, and the last equality is a direct result of Theorem 2. We now put all the steps together as follows:

$$\begin{aligned}
\alpha &\leq \mathbb{P}\left[\, r_{k+1} \in \mathcal{T}_{k+1}\ \ \tilde{d} \in \tilde{\mathcal{B}}_{\mathcal{N}} \,\right] = F_{r_{k+1},\, \tilde{d}}\left(\mathcal{T}_{k+1}\ \ \tilde{\mathcal{B}}_{\mathcal{N}}\right) \\
&\leq F_{r_{k+1}}\left(\mathcal{T}_{k+1}\right) \lim_{\tilde{\mathcal{B}}_{\mathcal{N}} \to \mathbb{R}^{n_{\mathcal{N}}}} F_{\tilde{d}}\left(\tilde{\mathcal{B}}_{\mathcal{N}}\right) \\
&= \mathbb{P}\left[\, r_{k+1} \in \mathcal{T}_{k+1} \,\right] \left(\int_{\mathbb{R}^{n_{\mathcal{N}}} \setminus \tilde{\mathcal{B}}_{\mathcal{N}}} p(\tilde{d})\mathrm{d}\tilde{d} + \int_{\tilde{\mathcal{B}}_{\mathcal{N}}} p(\tilde{d})\mathrm{d}\tilde{d} \right) \\
&\leq \int_{\mathbb{R}^{n_{\mathcal{N}}} \setminus \tilde{\mathcal{B}}_{\mathcal{N}}} p(\tilde{d})\mathrm{d}\tilde{d} + \mathbb{P}\left[\, r_{k+1} \in \mathcal{T}_{k+1} \,\right] \int_{\tilde{\mathcal{B}}_{\mathcal{N}}} p(\tilde{d})\mathrm{d}\tilde{d} \\
&= (1 - \tilde{\alpha}_{\mathcal{N}}) + \tilde{\alpha}_{\mathcal{N}}\, \mathbb{P}\left[\, r_{k+1} \in \mathcal{T}_{k+1} \,\right]
\end{aligned}$$

where the first inequality and equality is due to (26), the second inequality is due to (27), the second and last equality is due to (28), and the last inequality is due to the fact that $\mathbb{P}[\, r_{k+1} \in \mathcal{T}_{k+1} \,] \leq 1$. Rearranging the last equation results in

$$\frac{\alpha - (1 - \tilde{\alpha}_{\mathcal{N}})}{\tilde{\alpha}_{\mathcal{N}}} = 1 - \frac{1 - \alpha}{\tilde{\alpha}_{\mathcal{N}}} = \bar{\alpha} \leq \mathbb{P}\left[\, r_{k+1} \in \mathcal{T}_{k+1} \,\right].$$

The proof is completed by noting that the final equation is our desired assertion. □

References

[1] E. Kyriakides and M. Polycarpou, *Intelligent Monitoring, Control, and Security of Critical Infrastructure Systems*. Berlin, Germany: Springer, 2014, vol. 565.

[2] S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Berlin, Germany: Springer, 2008.

[3] J. Chen and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer, 2012, vol. 3.

[4] C. De Persis and A. Isidori, "A geometric approach to nonlinear fault detection and isolation," *IEEE Trans. Autom. Control*, vol. 46, no. 6, pp. 853–865, 2001.

[5] X. Zhang, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Trans. Autom. Control*, vol. 47, no. 4, pp. 576–593, 2002.

[6] R. M. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 275–290, Feb. 2012.

[7] Q. Zhang and X. Zhang, "Distributed sensor fault diagnosis in a class of interconnected nonlinear uncertain systems," *Annu. Rev. Control*, vol. 37, no. 1, pp. 170–179, 2013.

[8] D. Zhang, Q. Wang, L. Yu, and H. Song, "Fuzzy-model-based fault detection for a class of nonlinear systems with networked measurements," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 12, pp. 3148–3159, Dec. 2013.

[9] X. Ge and Q. Han, "Distributed fault detection over sensor networks with Markovian switching topologies," *Int. J. Gen. Syst.*, vol. 43, nos. 3/4, pp. 305–318, 2014.

[10] S. Riverso, F. Boem, G. Ferrari-Trecate, and T. Parisini, "Plug-and-play fault detection and control-reconfiguration for a class of nonlinear large-scale constrained systems," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3963–3978, Dec. 2016.

[11] F. Boem, R. M. Ferrari, C. Keliris, T. Parisini, and M. M. Polycarpou, "A distributed networked approach for fault detection of large-scale systems," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 18–33, Jan. 2017.

[12] E. Noursadeghi and I. Raptis, "Reduced-order distributed fault diagnosis for large-scale nonlinear stochastic systems," *J. Dyn. Syst. Meas. Control*, 2017.

[13] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed protocol for electric vehicle charging," in *Proc. Conf. Commun. Control Comput.*, IEEE, 2014, pp. 242–249.

[14] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Proc. IEEE Conf. Smart Grid Commun.*, 2011, pp. 220–225.

[15] P. Mohajerin Esfahani and J. Lygeros, "A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance," *IEEE Trans. Autom. Control*, vol. 61, no. 3, pp. 633–647, Mar. 2016.

[16] F. Boem, R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Optimal topology for distributed fault detection of large-scale systems," *Proc. IFAC Symp. Fault Detection Supervision Saf. Tech. Process.*, vol. 48, no. 21, pp. 60–65, 2015.

[17] M. Milanese, J. Norton, H. Piet-Lahanier, and É. Walter, *Bounding Approaches to System Identification*. Berlin, Germany: Springer, 2013.

[18] J. Blesa, V. Puig, and J. Saludes, "Robust fault detection using polytope-based set-membership consistency test," *IET Control Theory Appl.*, vol. 6, no. 12, pp. 1767–1777, 2012.

[19] A. Ingimundarson, J. M. Bravo, V. Puig, T. Alamo, and P. Guerra, "Robust fault detection using zonotope-based set-membership consistency test," *Int. J. Adaptive Control Signal Process.*, vol. 23, no. 4, pp. 311–330, 2009.

[20] I. Fagarasan, S. Ploix, and S. Gentil, "Causal fault detection and isolation based on a set-membership approach," *Automatica*, vol. 40, no. 12, pp. 2099–2110, 2004.

[21] G. R. Marseglia, J. Scott, L. Magni, R. D. Braatz, and D. M. Raimondo, "A hybrid stochastic-deterministic approach for active fault diagnosis using scenario optimization," *IFAC World Congr.*, vol. 47, no. 3, pp. 1102–1107, 2014.

[22] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptography Conf.*, Berlin, Germany, 2006, pp. 265–284.

[23] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3/4, 2014, pp. 211–407.

[24] S. Han, U. Topcu, and G. J. Pappas, "Differentially private convex optimization with piecewise affine objectives," in *Proc. IEEE Conf. Decis. Control*, 2014, pp. 2160–2166.

[25] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 50–64, Jan. 2017.

[26] E. Akyol, C. Langbort, and T. Başar, "Privacy constrained information processing," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, 2015, pp. 4511–4516.

[27] F. Farokhi and H. Sandberg, "Ensuring privacy with constrained additive noise by minimizing fisher information," *Automatica*, vol. 99, pp. 275–288, 2019.

[28] V. Rostampour, R. Ferrari, A. Teixeira, and T. Keviczky, "Differentially private distributed fault diagnosis for large-scale nonlinear uncertain systems," in *Proc. IFAC Conf. Fault Detection, Supervision Saf. (SAFEPROCESS)*, 2018.

[29] R. Anguluri, V. Katewa, and F. Pasqualetti, "On the role of information sharing in the security of interconnected systems," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, 2018, pp. 1168–1173.

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

[31] V. Rostampour, R. Ferrari, and T. Keviczky, "A set based probabilistic approach to threshold design for optimal fault detection," in *Proc. IEEE Amer. Control Conf. (ACC)*, 2017, pp. 5422–5429.

[32] K. J. Åström and B. Wittenmark, *Computer-Controlled Systems: Theorey and Design*, 3rd ed. Prentice-Hall, 1997.

[33] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault Tolerant Control*. Springer, 2003.

[34] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *J. Process Control*, vol. 7, no. 6, pp. 403–424, 1997.

[35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[36] L. J. Guibas, A. Nguyen, and L. Zhang, "Zonotopes as bounding volumes," in *Proc. ACM-SIAM Symp. Discrete Algorithms*, Soc. Ind. Appl. Math., 2003, pp. 803–812.

[37] F. Dabbene and D. Henrion, "Set approximation via minimum-volume polynomial sublevel sets," in *Proc. IEEE Eur. Control Conf.*, 2013, pp. 1114–1119.

[38] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscipl. Optim.*, vol. 26, no. 6, pp. 369–395, 2004.

[39] R. Zippel, *Effective Polynomial Computation*, vol. 241. Berlin, Germany: Springer, 2012.

[40] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.

[41] V. Rostampour and T. Keviczky, "Probabilistic energy management for building climate comfort in smart thermal grids with seasonal storage systems," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3687–3697, Jul. 2019.

[42] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, May 2006.

[43] R. M. Ferrari, T. Parisini, and M. M. Polycarpou, "A fault detection and isolation scheme for nonlinear uncertain discrete-time sytems," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2007, pp. 1009–1014.

[44] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Modern Phys.*, vol. 74, no. 1, pp. 47–97, 2002.

[45] O. Kallenberg, *Foundations of Modern Probability*. Springer, 2006.

Vahab Rostampour (Member, IEEE) received the M.Sc. degree in automation and control engineering from the University of Politecnico di Milano, Milan, Italy in 2013, and the Ph.D. degree in systems and control from the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands, in 2018.

He is currently a Postdoctoral Researcher at the Engineering and Technology Institute Groningen, Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands. His research interests include control synthesis, analysis, distributed learning and optimization of dynamical systems, and decision making in uncertain dynamical environments, with applications to large-scale interconnected systems such as financial and energy systems.

Dr. Rostampour was the finalist for the Paul M. Frank Award from IFAC SAFEPROCESS, Warsaw, Poland, in 2018.

Riccardo M.G. Ferrari (Member, IEEE) received the Laurea degree (*cum laude* and printing honours) in electronic engineering and the Ph.D. degree in information engineering both from the University of Trieste, Trieste, Italy, in 2004 and 2009, respectively.

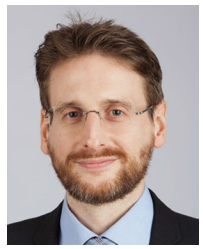
He is a Marie Curie Alumnus and is currently an Assistant Professor with the Delft Center for Systems and Control, Delft University of Technology, The Netherlands. His research interests include wind power fault tolerant control and fault diagnosis and attack detection in large-scale cyber-physical systems, with applications to electric vehicles, cooperative autonomous vehicles and industrial control systems.

Dr. Ferrari is the recipient of the 2005 Giacomini Award of the Italian Acoustic Society and he obtained the 2nd place in the Competition on Fault Detection and Fault Tolerant Control for Wind Turbines during IFAC 2011. Furthermore, he was awarded an Honorable Mention for the Paul M. Frank Award at the IFAC SAFEPROCESS in 2018 and won an Airbus Award at IFAC 2020 for the best contribution to the competition on Aerospace Industrial Fault Detection. He has held both academical and industrial R&D positions, in particular, as Researcher in the field of process instrumentation and control for the steel-making sector.

André M.H. Teixeira (Member, IEEE) received the M.Sc. degree in electrical and computer engineering from the Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, in 2009, and the Ph.D. degree in automatic control from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2014.

He is currently an Associate Senior Lecturer at the Division of Signals and Systems, Department of Electrical Engineering, Uppsala University, Sweden. From 2015 to 2017, he was an Assistant Professor at the Faculty of Technology, Policy and Management, Delft University of Technology. His current research interests include secure and resilient control systems, distributed fault detection and isolation, distributed optimization and power systems.

Dr. Teixeira was a recipient for the Best Student-Paper Award from the IEEE Multi-Conference on Systems and Control in 2014 and an Honorable Mention for the Paul M. Frank Award at the IFAC SAFEPROCESS in 2018. He was awarded a Starting Grant by the Swedish Research Council in 2019, and he is among the 20 young researchers in Sweden that received the Future Research Leaders 7 grant by the Swedish Foundation for Strategic Research in 2020.



Tamás Keviczky (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from the Budapest University of Technology and Economics, Budapest, Hungary, in 2001, and the Ph.D. degree in control science and dynamical systems from the Control Science and Dynamical Systems Center, University of Minnesota, Minneapolis, in 2005.

He is currently a Professor with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. He was a Postdoctoral Scholar in Control and Dynamical Systems, California Institute of Technology, Pasadena.

Prof. Keviczky served as Associate Editor of Automatica and he was a co-recipient of the AACC O. Hugo Schuck Best Paper Award for Practice in 2005. His research interests include distributed optimization and optimal control, model predictive control, embedded optimization-based control and estimation of large-scale systems with applications in aerospace, automotive and mobile robotics, industrial processes, and infrastructure systems such as water, heat, and electricity networks.