Plausible Deniability As a Notion of Privacy

Monshizadeh, Nima; Tabuada, Paulo

Published in:
Proceedings of the Conference on Decision and Control 2019

DOI:
10.1109/CDC40024.2019.9030201

2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Plausible deniability as a notion of privacy

Nima Monshizadeh and Paulo Tabuada

Abstract— **This work is motivated by privacy concerns as a result of the growing rate of information exchange among components of complex cyber-physical systems, agents in a network, or actuators/sensors of a process. We propose a deterministic notion of privacy for a dynamical system, and completely characterize it for linear time-invariant dynamics. The proposed notion relies on a "plausible deniability" principle, which implies that a curious party will always be in doubt about the actual value of private variables of the system. In case privacy is guaranteed, we propose analytical metrics to assess the degree of privacy or privacy margin of the system. The size of the latter depends on the amount and structure of the information on the system which can be accessed by a curious party. We study the proposed notions and metrics for a class of distributed averaging algorithms.**

I. Introduction

Recent advances in technology, internet of things, and big data, have led to increasing concerns about privacy. Modern complex systems include several components interacting with each other to achieve a certain goal. This is often accomplished by exchanging data through communication networks in order to perform certain tasks. A case in point are networked control systems where sensors, controller, and actuators are not co-located and need to exchange messages to close the loop. Our digital society is tantamount to extensive data transfer among users. The cornerstone of all these cases is exchange of information, which in turn raises issues on the privacy of the individuals or participating agents.

One approach to privacy relies on *data encryption* techniques [1], including homomorphic encryption [2], [3], data obfuscation [4], and multi-party computation schemes [5]. Typical challenges in those schemes are the large computational overhead and vulnerability during the public key distribution. Another related approach is based on algebraic transformations, where the original problem is mapped into an equivalent problem in which the private data is hidden, see [6], [7], in the context of linear programming, and [8] in dynamical systems.

A different class of methods to enhance privacy stems from *data perturbations*. The most popular tool in this category is differential privacy [9], [10], which serves as the basis of many of the results reported in the literature, see e.g. [11]–[16]. In most cases, this amounts to adding noise with appropriate statistical properties to the process under investigation. Consequently, the ability of a curious party to estimate the private variables will be limited to a predetermined precision, leading to a compromise between privacy and performance.

Inserting stochastic time-varying parameters in the control system has been pursued in [17]. Among other relevant methods, mostly tailored for averaging algorithms, are [18], [19], where the communication weights are appropriately synthesized to ensure privacy, and [20] where time-varying output masks are carefully devised to protect the initial opinion of the agents in a consensus protocol.

While the notion of differential privacy serves as a foundation for statistical analysis and design of data perturbation based techniques, the works developed in a deterministic setting are mostly diverse, and are based on suitable adaptations or ad-hoc ideas adjusted for the application of interest. Motivated by this, we consider a general dynamical system, and put forwards a deterministic notion of privacy, which relies on a plausible deniability principle. This means that a curious adversary cannot distinguish the actual private variables of the system from their "replicas" due to insufficient accessible knowledge. To ensure privacy, there should exists copies of private variables that are different from the original ones, yet exhibit the same external behavior. A conceptually similar notion has been studied under the name of *opacity* for discrete event systems modeled by automata, see e.g. [21] for a recent overview on the topic. From the technical perspective, the proposed notion relates to the indistinguishability property and the unobservable subspace of switched systems, see e.g. [22]–[24]. The proposed definition of privacy is then characterized for linear time-invariant systems. In case privacy is guaranteed, we investigate the degree of privacy or privacy margin of the system by leveraging suitably constructed metrics. As will be observed, the privacy margin is directly related to the amount and the structure of the information that can be assessed by a curious adversary. We illustrate the results on a class of distributed averaging algorithm. The focus of the current paper will be primarily on analysis, and we postpone a systematic design of privacy-preserving controllers to a future work.

The structure of the paper is as follows. In Section II, we propose and characterize a notion of privacy for dynamical systems. Privacy margin of the system is examined in Section III. Privacy aspects of a class of averaging algorithms are studied in Section IV, and concluding remarks are provided in Section V.

II. Keeping private variables private

We consider a linear time-invariant system given by

$$\dot{x}(t) = Ax(t), \tag{1}$$

N. Monshizadeh is with the Engineering and Technology Institute, University of Groningen, The Netherlands. Email: n.monshizadeh@rug.nl. Paulo Tabuada is with Department of Electrical and Computer Engineering, University of California, Los Angeles, USA. Email: tabuada@ucla.edu

978-1-7281-1398-2/19/$31.00 ©2019 IEEE

where $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. While we consider continuous-time systems here, we note that most of the subsequent developments apply to both continuous and discrete time systems. The variables that can be measured are collected in a vector $y \in \mathbb{R}^\ell$ and are given by

$$y(t) = Cx(t). \tag{2}$$

We denote the variables whose values should be kept *private* by

$$p(t) = Px(t), \tag{3}$$

where $P \in \mathbb{R}^{r\times n}$, $r \leq n$. The value of p should be protected for almost all time $t \geq t_0$.[1] We use the term *public* to refer to knowledge, variables, or quantities that are not considered private.

If the pair (C, A) is observable and all the matrices involved are known, then the vector p can be reconstructed by exploiting the measurements y and devising a suitable observer. Therefore, we allow the system to have a degree of *secrecy* in its dynamics. In particular, we assume that while the matrices C and P are known (public), the system dynamics is only partially known. In particular, we consider that the dynamics belong to a system class, i.e.,

$$A \in \mathcal{A}. \tag{4}$$

We treat $\mathcal{A}$ as public knowledge, while it may also reflect any (side) knowledge of a curious adversary on the system. The class of matrices $\mathcal{A}$ can be finite or infinite. To rule out the trivial (uninteresting) cases, we assume that (C, A) is observable.

Next, we define a deterministic notion of privacy which is based on *plausible deniability*, i.e, the ability to cast enough doubt over your culpability that it cannot be ascertained. This principle is also the basis for the notion of opacity in discrete time event systems modeled by finite automata, see e.g. [21] and the references therein. We use the notation $x_X(t, x_0)$ to denote the state response and $y_X(t, x_0)$ to denote the output response resulting from the initial condition x_0 and the state matrix $X \in \mathcal{A}$.

Definition 1: We say that privacy is preserved for the system (1)–(4) if for any $x_0 \in \mathbb{R}^n$ there exists $\hat{x}_0 \in \mathbb{R}^n$ and $\hat{A} \in \mathcal{A}$, with $\hat{A} \neq A$, such that

$$y_A(t, x_0) = y_{\hat{A}}(t, \hat{x}_0) \tag{5}$$

for all time $t \geq t_0$, and

$$Px_A(t, x_0) \neq Px_{\hat{A}}(t, \hat{x}_0). \tag{6}$$

for $t = t_0$ and almost all time $t > t_0$.

Remark 1: We remark that Definition 1 does not rely on linearity, and can be analogously stated for nonlinear systems. However, linearity is assumed for the subsequent characterisation and development of the results.

Remark 2: The conditions in Definition 1 can be equivalently stated for a nonempty finite interval of time since p and y are both real analytic functions of time (see e.g. [25]). Due to the same reason, the condition (6) needs to be checked only at time $t = t_0$ (see Proposition 1), however we keep the definition as it is for the sake of generality and to highlight the intuition behind it.

The rationale behind Definition 1 is that, from the measurements y and class $\mathcal{A}$, one cannot distinguish a system with the state matrix A and private variables p from the one with the state matrix $\hat{A}$ and the private variables $\hat{p}$, with $\hat{p}$ denoting the right hand side of (6). The condition (5) is closely related to the notion of observability of switched systems (see e.g. [22]), whereas (6) is additionally needed to ensure privacy. The latter is due to the fact that $x_0 \neq \hat{x}_0$ does not in general imply $Px_0 \neq P\hat{x}_0$.

The following characterization follows from the the notion of privacy posed in Definition 1:

Proposition 1: Privacy is preserved for the system (1)–(4) if and only if for any $x_0 \in \mathbb{R}^n$ there exists $\hat{x}_0 \in \mathbb{R}^n$ and $\hat{A} \in \mathcal{A}$, with $\hat{A} \neq A$ such that

$$CA^{k-1}x_0 = C\hat{A}^{k-1}\hat{x}_0, \qquad \forall k \geq 1, \tag{7}$$

and

$$Px_0 \neq P\hat{x}_0. \tag{8}$$

Proof: The equivalence of (5) and (7) is well-known. Namely, computing (high order) time derivatives of (5) and taking the limit as $t \to t_0$ yields (7), and the converse result follows by using the Taylor series of the state-transition matrices corresponding to A and $\hat{A}$. Moreover, the condition in (6) imposes $Px_0 \neq P\hat{x}_0$. To complete the proof, it suffices to show that if (6) does not hold, then $Px_0 = P\hat{x}_0$. Suppose that (6) does not hold, then either $Px_0 = P\hat{x}_0$ or there exists a nonzero interval of time for which $Px_A(t, x_0) = Px_{\hat{A}}(t, \hat{x}_0)$. As solutions of the system can be expressed as a real analytic function of time, the latter equality must hold for all time $t \in (t_0, +\infty)$ (see e.g. [25]), which results in $Px_0 = P\hat{x}_0$. ■

As is well-known, one needs to check the condition in (7) only in a finite set $k \in \{1, 2, \ldots, K\}$. In this case, we have $K \leq 2n$ as (7) can be seen as the (un)observability property of the pair

$$(\begin{bmatrix} C & -C \end{bmatrix}, \begin{bmatrix} A & 0 \\ 0 & \hat{A} \end{bmatrix}).$$

In fact, K can be taken as small as the "joint-observability index" of the pairs (C, A) and $(C, \hat{A})$ (see e.g. [22]).

III. Privacy Margin

In the previous section, we have provided a notion of privacy together with its characterization. Assuming that privacy is preserved, the next important question is to quantify the *degree of privacy* or *privacy margin*. Here, we assume that the matrix A is Hurwitz, and this is regarded as public knowledge. Consequently, all matrices in $\mathcal{A}$ are Hurwitz.

[1] By "almost all time", we mean with the exception of a set of measure zero.

Motivated by Definition 1, given $x_0 \in \mathbb{R}^n$, we quantify the gap between p and $\hat{p}$, with $\hat{p}$ denoting the right hand side of (6). To this end, let

$$\xi(t) := p(t) - \hat{p}(t) = Px_A(t, x_0) - P\hat{x}_{\hat{A}}(t, \hat{x}_0). \quad (9)$$

Intuitively, the energy of the signal $\xi(t)$ is a measure of privacy, and reflects the mismatch between the private variable p and its replica $\hat{p}$. For simplicity, we set $t_0 = 0$. We denote the squared L_2-norm of ξ by Ξ, namely

$$\Xi(p, \hat{p}) := \int_0^{+\infty} \xi(t)^\top \xi(t) dt. \quad (10)$$

Note that while Definition 1 allows for existence of multiple (and possibly infinite) pairs of $\hat{x}_0$ and $\hat{A}$, we focus first on one such pair. We get back to this point towards the end of this section.

We have now the following proposition:

Proposition 2: Assume that the matrices A and $\hat{A}$ are both Hurwitz, and the pairs (C, A) and $(C, \hat{A})$ are observable, with their observability matrices denoted by $\mathcal{O}(C, A)$ and $\mathcal{O}(C, \hat{A})$, respectively. Let

$$\Pi := \mathcal{O}(C, A)^+ \mathcal{O}(C, \hat{A}), \quad (11)$$

with $\mathcal{O}(\cdot)^+$ denoting the left-inverse of $\mathcal{O}(\cdot)$. Suppose that privacy is preserved in the sense of Definition 1. Then, we have

$$\Xi(p, \hat{p}) = x_0^\top Q x_0 \quad (12)$$

where Ξ is given by (10), and $Q = Q^\top$ is the unique solution to the Lyapunov equation

$$A^\top Q + QA + (I - \Pi)P^\top P(I - \Pi) = 0. \quad (13)$$

Proof: By (5), and its corresponding high-order time derivatives, we have that

$$\mathcal{O}(C, A)x_A(t, x_0) = \mathcal{O}(C, \hat{A})x_{\hat{A}}(t, \hat{x}_0).$$

Since $(C, \hat{A})$ is observable, we find that

$$x_{\hat{A}}(t, \hat{x}_0) = \mathcal{O}(C, \hat{A})^+ \mathcal{O}(C, A)x_A(t, x_0) = \Pi x_A(t, x_0).$$

Therefore, we find that

$$\xi(t) = P(I - \Pi)x_A(t, x_0).$$

Noting that $x_A(t, x_0) = e^{At}x_0$ and A is Hurwitz, it is well-known that the squared L_2-norm can be written as in (12), which completes the proof. ■

Remark 3: The assumption of A being Hurwitz is made in order to write $\Xi(p, \hat{p})$ in terms of the unique solution of the Lyapunov equation in (13). Note that this is sufficient, but not necessary, for the integral in (10) to be well-defined. More generally, $\Xi(p, \hat{p})$ can be defined in a finite interval as

$$\Xi(p, \hat{p}) := \int_0^{t_f} \xi(t)^\top \xi(t) dt, \quad (14)$$

for some $t_f > 0$. In this case, $\Xi(p, \hat{p})$ still admits the quadratic from in (12), but with Q given by the finite gramian

$$\int_0^{t_f} e^{A^\top t}(I - \Pi)P^\top P(I - \Pi)e^{At} dt. \quad (15)$$

The projection matrix $I - \Pi$ can be seen as a measure of similarity/dissimilarity between the observability matrices of $\mathcal{O}(C, A)$ and $\mathcal{O}(C, \hat{A})$. Then, as can be seen from the result of Proposition 2, the value of Ξ becomes larger when the observability matrices become more dissimilar.

Note that having more choices of $\hat{x}$ and $\hat{A}$ satisfying the conditions of Definition 1 enhances the privacy, and in fact adds to the *confusion* of a curious party. To take this into account, given $x_0 \in \mathbb{R}^n$, we define

$$\Xi^*(x_0) := \sup\nolimits_{\hat{p} \in \mathcal{P}} \Xi(p, \hat{p}), \quad (16)$$

where $\mathcal{P}$ is the collection of all vectors $Px_{\hat{A}}(t, \hat{x}_0)$ for which $\hat{A}$ and $\hat{x}_0$ satisfy the conditions of Definition 1.

While the values of Ξ and Ξ^* quantify a degree of privacy, there is still one major drawback to adopt these measures, and that is their dependency on the vector x_0. To overcome this issue, one can appeal to *average* or *worst-case* measures. In particular, in view of (12), $\operatorname{trace}(Q)$, $\det(Q)$, or the minimum eigenvalue of Q are prime candidates for quantifying the privacy margin available in the system. Motivated by this, we define the *privacy margin* of the system as

$$\Xi_\mu := \sup_{Q \in \mathcal{Q}} \mu(Q) \quad (17)$$

where the metric $\mu : \mathbb{R}^{n \times n} \to \mathbb{R}_{\geq 0}$ can be taken as $\operatorname{trace}(\cdot)$, $\det(\cdot)$, or the smallest eigenvalue. The class $\mathcal{Q}$ is the collection of all solutions Q of the Lyapunov equation (13) with $\hat{A} \in \mathcal{A}$ being any matrix satisfying the conditions in Definition 1.[2] Finally, we note that one can bypass the Lyapunov equation and work directly with the integral in (15). The latter is particularly useful in case the state matrix is not Hurwitz.

IV. Case study: distributed averaging

Here, we study the proposed notions of privacy and privacy margins on a disturbed averaging algorithm. The idea of averaging algorithms is to compute the average or a weighted average of initial states (opinion) of agents via exchanging information in a distributed fashion. Here, we consider

$$T\dot{x}(t) = -Lx(t), \quad (18)$$

where $L \in \mathbb{R}^{N \times N}$ is the Laplacian matrix of a connected graph, and $T > 0$ is diagonal. The dynamics in (18) can represent mass-damper systems as well. It is well-known that solutions of (18) converge to a weighted average of initial state $x(0)$, namely to $\mathbb{1}x^*$, with

$$x^* := \frac{1}{\operatorname{trace}(T)} \sum_{i=1}^{N} T_i x_i(0), \quad (19)$$

[2] Note that the dependency of Q on $\hat{A}$ stems from the matrix Π in (11).

where $\mathbb{1}$ denotes a vector of all ones with an appropriate dimension. We treat the opinion/value of the agents as the variables which we would like to keep private, namely let $p(t) = x(t)$. However, exchanging the state variables of the agents can immediately reveal such private information. To avoid this, one could introduce additional state variables, namely ζ, run the consensus algorithm on ζ, and make ζ track x asymptotically. Such augmented dynamics can be written as

$$T\dot{x}(t) = -L\zeta(t) \tag{20a}$$
$$\dot{\zeta}(t) = K_x x(t) - K_\zeta \zeta(t) \tag{20b}$$
$$y(t) = \zeta(t) \tag{20c}$$
$$p(t) = x(t). \tag{20d}$$

where K_x and K_y are diagonal and positive definite. Note that the values of the communicated variables are assumed to be known (public), and thus are collected in y. For illustration purposes, we first examine the privacy aspects of the dynamics above, and postpone the discussion on stability and convergence to a later moment. The time constant matrix T is considered to be unknown for a curious party, and, to avoid too much secrecy in the dynamics, we assume that one of the matrices K_x and K_ξ is set to the identity matrix. This gives rise to the following two distinct cases:

(i) $K_x = I$,

(ii) $K_\zeta = I$.

Note that the matrices K_ζ and K_x account for the *secrecy* in the dynamics in case i) and case ii), respectively. The public information is the measurement y and the system dynamics modulo the value of T and K_ζ in case (i), and T and K_x in case (ii).

First, we consider case (i). It is public that the state matrix belongs to a set $\mathcal{A}$ whose elements are parametrized by

$$\hat{A} = \begin{bmatrix} 0 & -\hat{T}^{-1}L \\ I & -\hat{K}_\zeta \end{bmatrix},$$

for some diagonal matrices $\hat{K}_\zeta, \hat{T} > 0$. Now, given (x_0, ζ_0), verifying the condition (7) with $k = 1$ yields $\zeta_0 = \hat{\zeta}_0$. For $k = 2$, we obtain that

$$x_0 - \hat{x}_0 = (K_\zeta - \hat{K}_\zeta)\zeta_0,$$

where we used the fact that $\zeta_0 = \hat{\zeta}_0$. If $\zeta_0 = 0$, then $x_0 = \hat{x}_0$, which violates the condition (8), and thus privacy is lost, noting that Definition 1 is stated for an arbitrary initial condition of the original system.

Next, we consider case (ii), namely $K_\zeta = I$. The public information in this case is that the state matrix belongs to a set $\mathcal{A}$ whose elements are parametrized by

$$\hat{A} = \begin{bmatrix} 0 & -\hat{T}^{-1}L \\ \hat{K}_x & -I \end{bmatrix}, \ \hat{K}_x > 0. \tag{21}$$

Given (x_0, ζ_0), verifying the condition (7) with $k = 1$ results in

$$\zeta_0 = \hat{\zeta}_0. \tag{22}$$

For $k = 2$, we find that

$$K_x x_0 = \hat{K}_x \hat{x}_0. \tag{23}$$

For $k = 3$, we have

$$K_x T^{-1} = \hat{K}_x \hat{T}^{-1}. \tag{24}$$

By computation, one can verify that (23) and (24) are sufficient to satisfy the conditions of Proposition 1. Moreover, these equalities highlight the fact that keeping both K_x and T "secret" is necessary. In fact, if either $K_x = \hat{K}_x$ or $T = \hat{T}$, then $x_0 = \hat{x}_0$ which implies the lack of privacy by (8).

Next, we discuss convergence properties of the dynamics in (20) for case (ii).

Proposition 3: The algorithm (20) with $K_x > 0$ and $K_\zeta = I$, preserves privacy and the vector x globally converges to $K_x^{-1}\mathbb{1}\overline{x}$ with

$$\overline{x} := \frac{1}{\text{trace}(TK_x^{-1})} \sum_{i=1}^{N} T_i x_i(0). \tag{25}$$

Proof: The proof for privacy was established preceding the proposition. To prove the convergence result let L be decomposed as $B\Gamma B^\top$ where B is the incidence matrix of the graph and Γ is a diagonal matrix with positive diagonal elements indicating the weights of the coupling. Let $z := Tx$ and $\eta := B^\top \zeta$. Then, we have

$$\begin{bmatrix} \dot{z} \\ \dot{\eta} \end{bmatrix} = \begin{bmatrix} 0 & -B \\ B^\top & -\Gamma^{-1} \end{bmatrix} \begin{bmatrix} K_x T^{-1} z \\ \Gamma \eta \end{bmatrix}$$

which is a port-Hamiltonian system with the quadratic Hamiltonian [26]

$$H(x, \eta) := \frac{1}{2} x^\top K_x T^{-1} x + \frac{1}{2} \eta^\top \Gamma \eta.$$

Taking the Hamiltonian as the Lyapunov function, we have that

$$\dot{H} = -\eta^\top \Gamma \eta \le 0,$$

and thus the solutions are bounded. Then, by invoking LaSalle's invariance principle, we conclude that $\eta = 0$ and thus $B^\top K_x T^{-1} z = 0$ on the invariant set. Therefore, on this set $z = \alpha T K_x^{-1} \mathbb{1}$, for some $\alpha \in \mathbb{R}$. Noting that $\mathbb{1}^\top z$ is a conserved quantity of the system, we conclude that z converges to $\alpha T K_x^{-1} \mathbb{1}$ with

$$\alpha = \frac{1}{\text{trace}(TK_x^{-1})} \sum_{i=1}^{N} z_i(0).$$

Consequently, x converges to $K_x^{-1}\mathbb{1}\overline{x}$ with $\overline{x}$ given by (25) as claimed. Finally note that the convergence of z, and thus x is global since the Hamiltonian is radially unbounded. This completes the proof. ■

While (20) preserves privacy and its set of equilibria is attractive, the vector x converges to the same weighted average value in (19) if and only if K_x be chosen as a multiple of the identity matrix. While this is admissible, a smart curious party might do the same reasoning and figure out that K_x is a multiple of the identity matrix. Then, by

(23), this means that the value of the private vector x will be revealed up to a scaling factor.

To avoid this, an alternative idea is to abandon the requirement that x in (20) must converge to the exact same vector as the one in (18). Along this line, next we quantify the privacy margin of (20).

Recall that the state matrix in this case is given by

$$A = \begin{bmatrix} 0 & -T^{-1}L \\ K_x & -I \end{bmatrix},$$

and the matrix $\hat{A}$ is given by (21), where $\hat{K}_x$ and $\hat{T}$ satisfy (23) and (24). Let

$$\Delta := \hat{K}_x^{-1} K_x = \hat{T}^{-1} T. \tag{26}$$

Then, it is easy to verify that

$$\hat{A} = \begin{bmatrix} \Delta & 0 \\ 0 & I \end{bmatrix} \underbrace{\begin{bmatrix} 0 & -T^{-1}L \\ K_x & -I \end{bmatrix}}_{A} \begin{bmatrix} \Delta & 0 \\ 0 & I \end{bmatrix}^{-1}. \tag{27}$$

Noting (9), we have

$$\xi(t) = x(t) - \hat{x}(t) = \begin{bmatrix} I & 0 \end{bmatrix} e^{At} \begin{bmatrix} x_0 \\ \zeta_0 \end{bmatrix} - \begin{bmatrix} I & 0 \end{bmatrix} e^{\hat{A}t} \begin{bmatrix} \hat{x}_0 \\ \hat{\zeta}_0 \end{bmatrix}$$

By (27), we find that

$$\xi(t) = \begin{bmatrix} I & 0 \end{bmatrix} e^{At} \begin{bmatrix} x_0 \\ \zeta_0 \end{bmatrix} - \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & I \end{bmatrix} e^{At} \begin{bmatrix} \Delta & 0 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \hat{x}_0 \\ \hat{\zeta}_0 \end{bmatrix},$$

which, noting (22), (23), and (26), simplifies to

$$\begin{aligned} \xi(t) &= \begin{bmatrix} I & 0 \end{bmatrix} e^{At} \begin{bmatrix} x_0 \\ \zeta_0 \end{bmatrix} - \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} \Delta & 0 \\ 0 & I \end{bmatrix} e^{At} \begin{bmatrix} x_0 \\ \zeta_0 \end{bmatrix} \\ &= \begin{bmatrix} I - \Delta & 0 \end{bmatrix} e^{At} \begin{bmatrix} x_0 \\ \zeta_0 \end{bmatrix}. \end{aligned}$$

Noting that the state matrix is not Hurwitz, we follow the definition in (14), which yields

$$\Xi(x, \hat{x}) = \begin{bmatrix} x_0 \\ \xi_0 \end{bmatrix}^\top \left(\int_0^{t_f} e^{A^\top t} \begin{bmatrix} (I-\Delta)^2 & 0 \\ 0 & 0 \end{bmatrix} e^{At} dt \right) \begin{bmatrix} x_0 \\ \xi_0 \end{bmatrix}. \tag{28}$$

The privacy margin Ξ_μ can be obtained in view of (17) by looking at the trace, determinant, or the minimum eigenvalue of the integral in equality (28). Clearly, noting (26) and (28), to enhance privacy the matrices K_x and T must be selected from a larger pool to cause more confusion for a curious party. On the other hand, a larger set of admissible matrices, particularly for K_x, results in convergence to a point which might be far from the desired one, suggesting a compromise between privacy margin and accuracy. For illustration purposes, we have computed the trace of the matrix

$$\int_0^{t_f} e^{A^\top t} \begin{bmatrix} (I-\Delta)^2 & 0 \\ 0 & 0 \end{bmatrix} e^{At} dt, \tag{29}$$

with $t_f = 10$, for different values of Δ in a network of four nodes, with a line graph topology. The result of this computation is shown in Figure 1. As expected, the privacy margin improves as the amount of secrecy in the dynamics, which in this case has been quantified in terms of $\|I - \Delta\|$, increases.

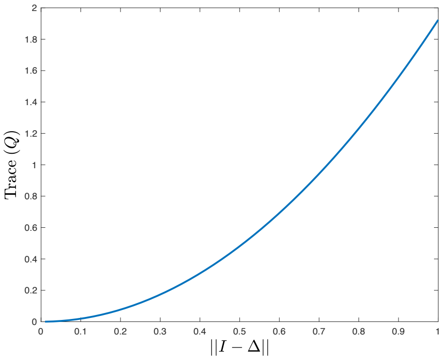

Fig. 1. The privacy margin computed from (29).

V. Conclusions

Relying on a a "plausible deniability" principle and inspired by observability of switched systems as well as opacity in discrete event systems, a deterministic notion of privacy has been provided for dynamical systems. Privacy is preserved if there exists at least one permissible choice of private variables (different to the actual one) that exactly mimics the behavior of the actual private variables. A complete algebraic characterization of such property has been provided. In case privacy is guaranteed for the system, we have proposed quantifiable metrics that assess the degree of privacy or privacy margin. In principle, the privacy margin of the system reveals how close a curious adversary can get to the true value of the private variables. As observed, such margin depends on the amount and structure of the information on the system which is accessible to an adversary. The proposed notions and metrics have been studied for a class of distributed averaging algorithm. We have shown how variations of the averaging algorithms can lead to different results in view of privacy and the proposed metrics. While the focus of the current work has been primarily on analysis, developing designing methodologies to guarantee privacy is part of the ongoing research. Extending the results to nonlinear systems is another challenging task for future research.

References

[1] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[2] K. Kogiso and T. Fujita, "Cyber-security enhancement of networked control systems using homomorphic encryption," in *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 6836–6843.

[3] Y. Shoukry, K. Gatsis, A. Alanwar, G. J. Pappas, S. A. Seshia, M. Srivastava, and P. Tabuada, "Privacy-aware quadratic optimization using partially homomorphic encryption," in *55th IEEE Conference on Decision and Control (CDC)*, 2016, pp. 5053–5058.

[4] C. Wang, K. Ren, and J. Wang, "Secure and practical outsourcing of linear programming in cloud computing," in *Proceedings IEEE INFOCOM*, 2011, pp. 820–828.

[5] W. Du and M. J. Atallah, "Secure multi-party computation problems and their applications: a review and open problems," in *Proceedings of the 2001 workshop on New security paradigms*. ACM, 2001, pp. 13–22.

[6] J. Dreier and F. Kerschbaum, "Practical privacy-preserving multiparty linear programming based on problem transformation," in *IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*, 2011, pp. 916–924.

[7] P. C. Weeraddana, G. Athanasiou, C. Fischione, and J. S. Baras, "Per-se privacy preserving solution methods based on optimization," in *52nd IEEE Conference on Decision and Control (CDC)*, 2013, pp. 206–211.

[8] A. Sultangazin and P. Tabuada, "Towards the use of Symmetries to Ensure Privacy in Control Over the Cloud," in *57th IEEE Conference on Decision and Control (CDC)*, 2018, pp. 5008–5013.

[9] C. Dwork, "Differential privacy," *Encyclopedia of Cryptography and Security*, pp. 338–340, 2011.

[10] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[11] J. Le Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2014.

[12] J. Cortes, G. E. Dullerud, S. Han, J. L. Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *55th IEEE Conference on Decision and Control (CDC)*, 2016, pp. 4252–4272.

[13] H. Sandberg, G. Dán, and R. Thobaben, "Differentially private state estimation in distribution networks with smart meters," in *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 4492–4498.

[14] F. Koufogiannis and G. J. Pappas, "Differential privacy for dynamical sensitive data," in *56th IEEE Conference on Decision and Control (CDC)*, 2017, pp. 1118–1125.

[15] E. Nozari, P. Tallapragada, and J. Cortes, "Differentially Private Distributed Convex Optimization via Functional Perturbation," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 395–408, mar 2018.

[16] Y. Kawano and M. Cao, "Design of differentially private dynamic controllers," *arXiv preprint arXiv:1901.07384*, 2019.

[17] P. Griffioen, S. Weerakkody, and B. Sinopoli, "A moving target defense for securing cyber-physical systems," *arXiv preprint arXiv:1902.01423*, 2019.

[18] A. Alaeddini, K. Morgansen, and M. Mesbahi, "Adaptive communication networks with privacy guarantees," in *American Control Conference (ACC)*, 2017, pp. 4460–4465.

[19] S. Pequito, S. Kar, S. Sundaram, and A. P. Aguiar, "Design of communication networks for distributed computation with privacy guarantees," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 1370–1376.

[20] C. Altafini, "A dynamical approach to privacy preserving average consensus," *arXiv preprint arXiv:1808.08085*, 2018.

[21] S. Lafortune, F. Lin, and C. N. Hadjicostis, "On the history of diagnosability and opacity in discrete event systems," *Annual Reviews in Control*, vol. 45, pp. 257–266, 2018.

[22] R. Vidal, A. Chiuso, S. Soatto, and S. Sastry, "Observability of linear hybrid systems," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2003, pp. 526–539.

[23] M. Babaali and M. Egerstedt, "Observability of switched linear systems," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2004, pp. 48–63.

[24] M. Baglietto, G. Battistelli, and P. Tesi, "Mode-observability degree in discrete-time switching linear systems," *Systems & Control Letters*, vol. 70, pp. 69–76, 2014.

[25] S. G. Krantz and H. R. Parks, *A primer of real analytic functions*. Springer Science & Business Media, 2002.

[26] A. van der Schaft and D. Jeltsema, *Port-Hamiltonian Systems Theory: An Introductory Overview*. Now Publishers Incorporated, 2014.