Distortion-Based Lightweight Security for Cyber-Physical Systems

Gaurav Kumar Agarwal, Mohammed Karmoose, Suhas Diggavi, *Fellow, IEEE*, Christina Fragouli, and Paulo Tabuada

***Abstract*—In cyber-physical systems (CPS), inference based on communicated data is of critical significance as it can be used to manipulate or damage the control operations by adversaries. This calls for efficient mechanisms for secure transmission of data since control systems are becoming increasingly distributed over larger geographical areas. Distortion-based security, recently proposed as one candidate for secure transmissions in CPS, is not only more appropriate for these applications but also quite frugal in terms of prior requirements on shared keys. In this article, we propose distortion-based metrics to protect CPS communication and show that it is possible to confuse adversaries with just a few bits of preshared keys. In particular, we will show that a linear dynamical system can communicate its state in a manner that prevents an eavesdropper from accurately learning the state.**

***Index Terms*—Cyber-physical systems, information security.**

I. Introduction

WIRELESS networked environments are a natural host for a number of cyber-physical control applications, ranging from autonomous cars and drones, to the Internet-of-Things (IoT), to immersive environments, such as augmented reality. It is well recognized that wireless networking is essential to realize the potential of new CPS applications, and is equally well recognized that private and secure exchange of information are necessary and not simply desirable conditions for the CPS ecosystem to thrive. For instance, personal health data in assisted environments, car positions and trajectories, and proprietary interests all need to be protected. This article introduces a new approach to secure communication in CPS, that aims to distort an adversary's view of a control system's states. In particular, we will show that a linear dynamical system can securely communicate its state to a trusted party in a manner that prevents a malicious adversary eavesdropping the communication from accurately learning the state.

Our starting observation is that information security measures (cryptographic and information theoretic secrecy) are not well matched to CPS applications as they impose unnecessary requirements, such as protecting all the raw data, and thus can cause high operational costs.[1] To illustrate this, we start by comparing existing techniques for ensuring CPS privacy, namely *cryptographic* and *information-theoretic* techniques. Cryptographic methods rely on computational complexity as a guarantee for the security of the underlying CPS, i.e., the system is secure against computationally limited adversaries. Cryptographic methods are universal and therefore are easy to integrate in any system under consideration. However, some of their shortcomings are as follows.

1) They do not provide guarantees against adversaries with unlimited computational power (e.g., quantum adversaries).
2) They utilize computationally heavy encryption/decryption algorithms.
3) They come at the cost of high overhead on short packet transmissions, therefore increasing delays [2]–[6].

In fact, those techniques have been previously studied in the context of secure CPSs. For example, homomorphic encryption [7]–[9] and public-encryption systems [10] have been used to provide security of networked control systems. Alternatively, information-theoretic methods rely on keys: they have low complexity and do not add packet overhead, but require the communicating nodes to share large keys—every communication link needs to use a shared secret key (for a one-time pad) of length equal to the entropy (effectively length) of the transmitted data [11], [12]. These costs accumulate rapidly given that large CPS applications can have dense communication patterns.

Instead, we propose a lightweight approach that provides security guarantees against computationally capable (i.e., with unlimited computational power) adversaries and uses small amounts of keys and low complexity operations. The main observation behind our approach is the following. Consider a general

Manuscript received July 22, 2019; revised February 15, 2020; accepted May 25, 2020. Date of publication July 3, 2020; date of current version March 29, 2021. This work was supported in part by NSF under Grant 1740047 and Grant 1705135, in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196, and in part by the UC-NL under Grant LFR-18-548554. This paper was presented in part at the 57th Conference on Decision and Control, Miami, FL, USA, December 2018. Recommended by Associate Editor W. X. Zheng. *(Gaurav Kumar Agarwal and Mohammed Karmoose are cofirst authors.) (Corresponding author: Gaurav Kumar Agarwal.)*

The authors are with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90024 USA (e-mail: gauravagarwal@ucla.edu; mkarmoose@ucla.edu; suhasdiggavi@ucla.edu; christina.fragouli@ucla.edu; tabuada@ucla.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2020.3006814

[1] Our article focuses on security against passive adversaries—alternatively referred to in the literature as *CPS Privacy*.

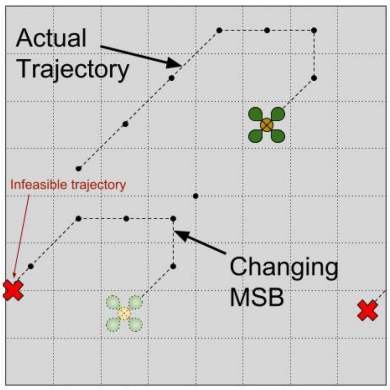

Fig. 1. Example of drone motion: protection of the most significant bit (MSB).

encryption scheme that uses a K-bit key to encrypt the states of a dynamical system. From an abstract point of view, such a scheme hides the true value of the state among a set of 2^K states; without knowing the value of the key, an outside observer of the encrypted state cannot resolve the ambiguity among these fake states—we refer to this set as the *ambiguity set*. General encryption (e.g., cryptographic or information-theoretic) schemes aim at increasing the size of the ambiguity set. Differently, in CPS applications, increasing the size of the ambiguity set may not be effective if all of these states are close to each other in a metric space. To make this idea concrete, assume that an adversary is trying to locate a drone in order to shoot it down with one missile in its possession. If a K-bit encryption scheme is used, the adversary would ideally have a set of 2^K possible locations for the drone—any set of 2^K possible locations are equivalent from an information-theoretic security point of view. However, if all locations are in close proximity, an adversary can possibly shoot the missile and hit the target regardless of the actual location. On the other hand, a different encryption scheme, which uses only 1 bit of information, but instead carefully chooses the two possible locations to be far apart, would be more secure against such an adversary. We therefore propose a distortion measure in order to capture this idea.

The following example illustrates the effect of maximizing distortion.[2] Consider the following simple example of a drone's flying motion, depicted in Fig. 1. The drone starts at any position, and moves between adjacent points within the grid. It regularly communicates its location to a legitimate receiver, Bob. A passive eavesdropper, Eve, wishes to infer the drone's locations, and can perfectly overhear all the transmissions the drone makes. We assume that the drone and Bob share just one bit of key, which is secret from Eve, and ask: what is the best use we can make of the key?

Using the 1-b of shared key to protect the MSB is not a good solution. The MSB can be protected by XORing a 1-b of shared key with the MSB. As shown in Fig. 1, the adversary can discover the fake trajectory after a few time steps since this scheme leads to trajectories that do not adhere to the dynamics or environment constraints. In particular, the fake trajectory abruptly moves from the left end of the grid to the right end. At this point, the adversary can learn the real trajectory by flipping back the MSB (we assume that the used scheme is known to everyone). Similar attacks can be made if we use a one-time pad [11] using the same keys over time: as time progresses, more fake trajectories can be discovered and discarded.

Conventional entropy measures also fail to provide insights on how to use the key. For instance, assume we label the 64 squares in Fig. 1 sequentially row per row, and consider two cases: in case I, Eve learns that the drone is in one of the neighboring squares $\{1, 2\}$, each with probability 1/2. For case II, Eve knows that the drone is in one of the squares $\{1, 64\}$, again each with probability 1/2. Both cases are equivalent from an information security perspective since in both cases Eve's uncertainty is a set of two equiprobable elements and hence its entropy is 1. However, the security risks in both situations are different. For example, if Eve aims to take a photo of the drone, in the first case, she knows where to turn her camera (squares 1 and 2 are close by), whereas in the second case, she does not (squares 1 and 64 are far apart).

Instead, we propose to use a Euclidean distance distortion measure: how far (in Euclidean space) is Eve's estimate from the actual location. We then propose encoding/decoding schemes that utilize the shared key to maximize this distance. We first consider an "average" distortion measure. Note that if Eve had not received any of the drone transmissions, then the best (adversarial) estimate of the drone's location at any given time is the center point of the confined region in Fig. 1. Therefore, a good encryption scheme would strive to maintain Eve's estimate to be as close to the center point as possible; we achieve the maximum possible distortion, if, after overhearing the drone's transmissions, Eve's best estimate still remains the center point.

The following scheme can achieve this maximum distortion by using exactly one bit of shared secret key. When encoding, the drone either sends its actual trajectory or its "mirrored" version, depending on the value of the secret key. The mirrored trajectory is obtained by reflecting the actual trajectory across a mirroring point in space; in this example, the mirroring point is the center point in Fig. 2. Since Eve does not know the value of the shared key, its best estimate of the drone's location—after receiving the drone's transmissions—would be the average location given the trajectory and its mirrored version, which is exactly the center point.

Our results in Section III extend this idea of mirroring to more general lightweight mappings for dynamical systems in higher dimensional spaces, and theoretically analyze the performance in terms of average distortion for a larger variety of distributions (with certain symmetry conditions). We also discuss a class of systems and controllers for which we can always achieve the perfect distortion with just 1-b of key.

Next, we consider a worst-case-sense distortion-based metric. In other words, our security metric is "in the worst case, how far is Eve's estimate from the actual location?" That is, the adversary's distortion may be different for different time instances and different instances of the actual trajectory, and we

[2] Although we illustrate our approach for a specific simple example, it extends to protecting general system states.

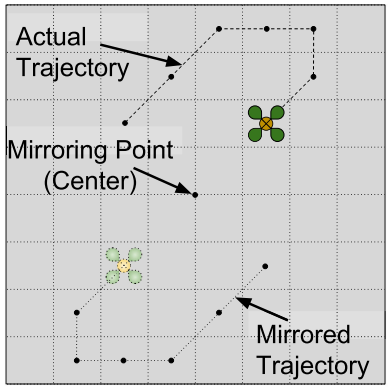

Fig. 2. Example of drone motion: mirroring-based scheme.

are interested in the minimum among these. For example when the drone is near the origin, both the actual and fake trajectories are close to origin and thus the adversary knows that the drone is near the origin. In this case, the overall (expected) distortion for adversary is still maximum, but at this particular instance of time, the adversary comes very close to the actual position. In Section V, we provide encryption schemes that are suitable for maximizing this distortion metric and show that with 3-b of shared key per dimension [i.e., nine for 3-D motion], our schemes achieve near-perfect worst case distortion. Our main contributions are as follows.

1) We define security measures that are based on assessing the distortion: in the average sense over time and over data, and in the minimum sense, providing worst case guarantees at any time and for any particular instances of data.
2) For the expected distortion, we develop a scheme that uses exactly 1-b of key and can provide maximum possible distortion (equivalent to Eve with no observations) in some cases. We also discuss the cases where it is not optimal and give an analytical characterization of the attained distortion. We then discuss a class of systems and controllers for which we can always guarantee the perfect distortion with just 1-b of shared key. Since for some applications, an ambiguity set of size two (corresponding to 1-b of key) may not be enough, we also derive an expression of attained distortion when we use larger keys.
3) For the worst case distortion, we design a scheme that uses 3-b of key per dimension and prove that it achieves the maximum possible distortion (equivalent to that of Eve with no observations) when the inputs to the systems are independent from the previous states.
4) For linear control systems, we provide a relation between the distortion in inputs to the distortion in states. This is particularly useful when inputs are easier to distort and analyze compared to the states.

A. Related Work

Secure data communication where the adversary has unlimited computational power is studied from the lens of information theory, most notably by Shannon [11] and Wyner [13]. The study of secure communication while using distortion as a measure of security is relatively new and is first studied by Yamamoto [14], where the goal is to maximize the distortion of an eavesdropper's estimate on a message, viewed from an asymptotic (in block length) information-theoretic approach. Schieler and Cuff [15] later showed that, in the limit of an infinite block length n code, only $\log(n)$ bits of secret keys are needed to achieve the maximum possible distortion. The idea of using finite block length (and even single-shot) distortion as a performance measure was initiated in [16], where schemes for single-shot communication were considered. It demonstrated the exponential benefits for each additional bit of shared key. The schemes examined were for single-shot sensor observations, and not for time-series data, which is the focus of this article.

Secure communication in control systems is studied in [17]–[21]. Securing the system state from an adversary was explored in [17] and [18], where an asymptotic steady-state analysis was explored. In contrast, this article also deals with transients and is not asymptotic. Information-theoretic security was explored in [19], where the mutual information was used as a privacy measure. Security of the terminal state is considered in [20] where an adversary makes partial noisy measurement of the state trajectory. Securing the states of an unstable system has been considered in [22] where the notion of secure capacity was used to characterize the level of secrecy against an adversary connected through a wiretap channel. Differential privacy for control systems was explored in [21], which uses standard statistical indistinguishability that is equally applicable to categorical (nonmetric space) data; in our article, we use the estimation error of the adversary in order to quantify privacy, utilizing the fact that CPS data lie in Euclidean space, as argued earlier.

B. Notation

For a matrix A, we denote by A' and A^H, the transpose and complex transpose of A, respectively; by A^r the rth power of A; X and X_a denote column vectors, and $X_a^b = [X_a' \; X_{a+1}' \; \cdots \; X_b']'$ for $b \geq a$ and $a, b \in \mathbb{Z}$; $f_X(x)$ denotes the probability density function of a random vector X; for any random vector Y, we denote the mean and covariance matrices of Y by μ_Y and R_Y respectively, thus for example, the mean and the covariance matrix of X_a^b will be denoted by $\mu_{X_a^b}$ and $R_{X_a^b}$ respectively; by $[m]$, we denote $\{1, 2, \ldots, m\}$ where $m \in \mathbb{Z}^+$; and by $[m_1 : m_2]$, we denote $\{m_1, m_1 + 1, \ldots, m_2\}$ where $m_1, m_2 \in \mathbb{Z}^+$ and $m_2 > m_1$; and a negative sign $(-)$ in the superscript of a function indicates the inverse of the function, i.e., the inverse of the functions $\alpha(x)$ and $\alpha^{(K)}(x)$ are $\alpha^{-}(x)$ and $\alpha^{-(K)}(x)$, respectively.

II. System Model

A. System Dynamics

We consider the linear dynamical system as

$$\widetilde{X}_{t+1} = A\widetilde{X}_t + BU_t + w_t, \qquad Y_t = C\widetilde{X}_t + v_t \quad (1)$$

where $\widetilde{X}_t \in \mathbb{R}^n$ is the state of the system at time $t \in \mathbb{N}$, $U_t \in \mathbb{R}^m$ is the input to the system at time t, $w_t \in \mathbb{R}^n$ is the process noise, Y_t are the system observations, and $v_t \in \mathbb{R}^p$ is the observation noise. We denote $\widetilde{X}_1^T$ by $\widetilde{X}$, U_1^{T-1} by U, and w_1^{T-1} by w. Based on the initial state $\widetilde{X}_1$ and target state $\widetilde{X}_T$, the controller computes a sequence of inputs that moves the state from the initial state $\widetilde{X}_1$ to the target state $\widetilde{X}_T$ in T time instances. We assume that the system uses the observations Y_1^T to optimally estimate the states $\widetilde{X}$. The optimal estimates of $\widetilde{X}$ made by the system are denoted by X—in the case of *perfect observation*, i.e., noiseless and observable systems, then $X = \widetilde{X}$.

B. Communication and Attacker/Defender Models

At each time instance, the system (Alice) transmits information about its state estimate to a legitimate receiver, which is referred to as Bob, via a noiseless link. This situation occurs, for example, when Bob is remotely monitoring the execution of the system as in Supervisory Control and Data Acquisition systems or in the remote operation of drones.

Attacker Model. A malicious receiver, referred to as Eve, is assumed to eavesdrop on the communication between the system and Bob and is able to receive all transmitted signals. The goal of Eve is to make an estimate that is as close to X as possible: since Bob receives X and makes control decisions with this information, Eve is interested in X. We assume that Eve knows the following:

1) the encoding/decoding functions used by Alice and Bob;
2) the dynamical system;
3) the controller design.

This information automatically implies knowing the prior probability distributions on the input and state vectors. With this information set, we assume that Eve uses the most adversarial eavesdropping strategy: one which minimizes our performance metric (see Section II-E). Eve is assumed to be passive: she does not actively communicate, but is interested in learning the system's states from $t = 1$ to T.

Defender Model. We assume that Alice and Bob have a shared k-bit key K that they use to encode/decode the transmitted messages. For a given encoding/decoding function, the assumed Eve adopts the most adversarial eavesdropping strategy (from the perspective of our chosen performance metric). Therefore, we assume that Alice/Bob attempts to design their encoding/decoding functions, which optimize this worst case performance. We elaborate more on that in Section II-E.

C. Inputs and States Random Process Model

We assume that both receivers are only aware of the system model, the matrices A, B, C, and the statistics of noises. Therefore, from the perspective of the receivers, the input and output sequences have random distributions which depend on A, B, C and the statistics of the noise. In addition to the process noise w, the joint distribution $f(X, U, w)$ depends on the following:

1) the initial and target states;
2) the control law of the system;
3) the state estimation process.

So, even in noiseless systems, X and U possess inherent randomness from a receiver's perspective due to its lack of knowledge about the initial and target states.

D. Encoding Model

The system encodes and transmits packets Z_1^T to ensure that Bob is able to accurately receive X_1^T, the optimal estimates of the system. To do so, the system transmits a packet Z_t at each time step t. In this article, we use lightweight memoryless encryption schemes. The tth transmitted packet is a function of only the current state estimate and the shared keys, thus, $Z_t := \mathcal{E}_t(X_t, K)$, where $\mathcal{E}_t$ is the encoding function used at time t. We will denote Z_1^T by Z.

E. Bob/Eve Models of Decoding

Bob noiselessly receives the transmitted packets from the system, and decodes them using the shared key. Then, using the decoded information, it generates an estimate of the state of the system at times $t \in [T]$. We require that Bob's estimate is as accurate as Alice's. If we assume that, at time $t \in [T]$, Bob's decoding function is $\Gamma_t(Z_1^t, K)$, then the previous condition is satisfied by ensuring that $\Gamma_t(Z_1^t, K) = X_t$ for all $t \in [T]$.

Similarly, Eve also receives all transmissions from the system. However, unlike Bob, she does not have the key K. Therefore, Eve's estimate of X_t is $\hat{X}_t := \phi_t(Z_1^T), t \in [T]$, where ϕ_t is the decoding function used by Eve at time t.

F. Distortion Metrics

We consider a distortion-based security metric that captures how far an estimate is from the actual value. In particular, our analysis is based on the Euclidean distance as our distance metric. However, our analysis can be extended to any p-norm, since other norms are just a constant factor away, i.e., $\|X\|_p \leq n^{\frac{1}{p}-\frac{1}{q}}\|X\|_q$. We assess the performance of Eve as how far its estimate $\hat{X}$ is from Alice's estimate X. Formally, for a given time instance t and a transmitted codeword Z_1^T, we define the following quantity:

$$D(t, Z_1^T) := \mathbb{E}_{X_t|Z_1^T}\left\|X_t - \hat{X}_t\right\|^2 \overset{(a)}{=} \operatorname{tr}\left(R_{X_t|Z_1^T}\right) \tag{2}$$

where (2) captures the distortion incurred by Eve while estimating X_t for transmitted symbols Z_1^T. Equality in (a) follows because the best (minimizing) estimates of Eve at time t are $\hat{X}_t = \phi_t(Z_1^T) = \mathbb{E}[X_t|Z_1^T]$.

Note that Bob is required to successfully estimate X_t knowing Z_1^t and the key. Therefore, for a given realization of the key, the encoding function can only map one X_t and that key realization to each value of Z_1^T. Therefore, Eve realizes that only trajectories from a particular subset can be the true trajectory for a given Z_1^T: those are the ones which correspond to each key realization. Therefore, the expectation in (2) is in fact taken over the randomness in the key taking into account posterior probabilities given Z_1^T. If Eve does not have observations, the expectation is taken over X_t with prior distribution and we get $D(t, Z_1^T) = \operatorname{tr}(R_{X_t})$.

As $D(t, Z_1^T)$ is a function of time t and the transmitted sequence Z_1^T, we consider two overall distortion metrics: the "average case" distortion (denoted by D_E) where we take expectation over all possible Z_1^T and average out over time; and the "worst case" distortion (denoted by D_W) where we take minimum over all possible Z_1^T and time instances

$$\text{Average Distortion} \quad - \quad D_E := \mathbb{E}_{Z_1^T}\left[\frac{1}{T}\sum_{t=1}^{T} D(t, Z_1^T)\right] \tag{3}$$

$$\text{Worst Case Distortion} \quad - \quad D_W := \min_{Z_1^T}\left[\min_{t\in[T]} D(t, Z_1^T)\right]. \tag{4}$$

It is worth to note that the definitions of D_E and D_W in (3) and (4) imply that Eve's state estimation must be associated to a time instance. In other words, making a random/constant estimate of the state hoping that it matches the actual state at some time will lead to high distortion values. Furthermore, D_W can be defined even when there is no prior distribution on X_1^T. However, to provide a baseline comparison with the case when the adversary has no observations, we assume that X_1^T always have a known prior distribution.

G. Design Goals

Our goal is to choose the encoding and decoding functions, $\mathcal{E}_t$ and γ_t, so that Bob can decode loselessly while the distortion is maximized for Eve's estimate. In addition, we seek to achieve this with the minimum amount of shared keys K. In the absence of any observation by Eve, these distortions will be

$$D_E^{\max} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{tr}(R_{X_t}), \qquad D_W^{\max} = \min_{t\in[T]} \mathrm{tr}(R_{X_t}).$$

These will serve as upper bounds as

$$D_E = \frac{1}{T}\mathbb{E}_{Z_1^T}\sum_{t=1}^{T}\mathrm{tr}(R_{X_t|Z_1^T}) \overset{(a)}{\leq} \frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}(R_{X_t}) = D_E^{\max} \tag{5}$$

$$D_W = \min_{Z_1^T}\min_{t\in[T]}\mathrm{tr}(R_{X_t|Z_1^T}) \leq \min_{t\in[T]}\mathbb{E}_{Z_1^T}\left[\mathrm{tr}(R_{X_t|Z_1^T})\right] \overset{(b)}{\leq} \min_{t\in[T]}\mathrm{tr}(R_{X_t}) = D_W^{\max} \tag{6}$$

where (a) and (b) follows from $\mathbb{E}_{Z_1^T}[\mathrm{tr}(R_{X_t}|Z_1^T)] \leq \mathrm{tr}(R_{X_t})$, which follows from the law of total variance.

III. Optimizing Average Distortion D_E

In this section, we will first discuss schemes to optimize the average distortion (D_E). We will initially analyze encoding schemes that use 1 b of secret key, and characterize their attained level of distortion. We then show that such schemes attain the maximum level of distortion for a family of distributions on X, which exhibit a certain class of symmetry. Later we describe how this analysis extends to the use of multiple keys, as for some application having an ambiguity set of size two might not be enough.

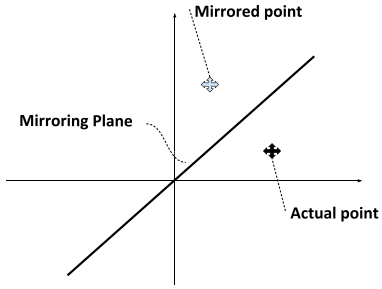

Fig. 3. Mirroring across the line passing through the origin and having a 45° angle with the X-axis.

A. Encoding Schemes With 1-b Shared Secret Key

We now discuss encoding schemes that use 1 b of shared key and show how the achieved distortion compares to the upper bound in (5). These encoding schemes work as follows:

$$Z_t = \begin{cases} X_t & \text{if } K = 0 \\ \alpha_t(X_t) & \text{if } K = 1 \end{cases} \quad \forall t \in [T] \tag{7}$$

where $K \in \{0, 1\}$ is the shared bit and $\alpha_t(X_t)$ is a transformation of the state vector X_t. We will next show the attained distortion of such schemes.

Theorem III.1 (Proof in Appendix A): The average distortion (D_E) attained by using the scheme in (7) is

$$\frac{1}{2T}\sum_{t=1}^{T}\mathbb{E}_X\left\{\frac{f_X(\alpha^-(X))}{f_X(X)+f_X(\alpha^-(X))}\left\|X_t - \alpha_t^-(X_t)\right\|^2\right\} \tag{8}$$

where $f\alpha^-(X) := [\alpha_1^-(X_1)'\ \alpha_2^-(X_2)'\ \cdots\ \alpha_T^-(X_T)']'$. Moreover, if the following condition holds:

$$f_X(x) = f_X(\alpha^-(x)), \qquad \text{for all } x \in \mathcal{X} \tag{9}$$

then the expression simplifies to

$$D_E = \frac{1}{4T}\sum_{t=1}^{T}\mathbb{E}_X\left\|X_t - \alpha_t(X_t)\right\|^2. \tag{10}$$

Condition (9) implies a general notion of symmetry in the distribution of $f_X(x)$. In the following, we focus on a particular notion of distribution symmetry, for which we show the corresponding choice of $\alpha_t(X_t)$ and how it can achieve high levels of distortion. Consider a transformation function $\alpha_t(x)$ that reflects a point x across an affine subspace of dimension d, defined by the equations $S_t x = b_t$, where $S_t \in \mathbb{R}^{d\times n}$ consists of $d \leq n$ orthonormal rows and $b_t \in \mathbb{R}^d$; the transformation is $\alpha_t(x) = (I - 2S_t'S_t)x + 2S_t'b_t$. The choice of the dimension d and the subspace (S_t, b_t) depends on the properties we would like the encoded trajectories to have. We refer to encoding schemes that are based on this transformation as *mirroring schemes*. For example, consider $X_t \in \mathbb{R}^2$ where $S_t = \frac{1}{\sqrt{2}}[-1\ \ 1]$ and $b_t = 0$. Then, $\alpha_t(X_t)$ corresponds to mirroring across a line that passes through the origin with a 45° angle. This is shown in Fig. 3. We are interested in *mirroring schemes* as they are lightweight and can be implemented on low-complexity IoT devices. Moreover,

such schemes can provide the maximum distortion level for a class of distributions with what we refer to as *point symmetry*.

Definition III.1 (Point Symmetry): A random vector X is said to have point symmetry if there exists a point v for which $f_X(x) = f_X(2v - x),\ \forall x \in \mathcal{X}$.

Lemma III.2: If X has point symmetry across v, then $v = \mu_X$.

Proof: Since X has point symmetry, then

$$f_X(x) = f_X(2v - x) \quad \Rightarrow \quad f_X(x) = f_{2v-X}(x)$$
$$\Rightarrow \quad \mu_X = 2v - \mu_X \quad \Rightarrow \quad \mu_X = v.$$

■

The following result characterizes the performance of the mirroring scheme, and shows that it achieves the maximum distortion for distributions with point symmetry.

Corollary III.3: If $\alpha_t(X_t)$ is based on a mirroring scheme along the planes given by $S_t x = b_t,\ t \in [T]$, and the condition (9) holds, then (10) becomes

$$D_E = \frac{1}{T}\sum_{i=1}^{T} \mathrm{tr}\left(S_t R_{X_t} S_t' + (b_t - S_t\mu_{X_t})(b_t - S_t\mu_{X_t})'\right). \tag{11}$$

Moreover, if X has point symmetry, then $D_E = \frac{1}{T}\sum_{t=1}^{T} tr(R_{X_t})$, the maximum possible distortion.

Proof: If condition (9) holds, then by simply plugging the expression of $\alpha_t(X_t)$ for the mirroring scheme along $S_t x = b_t$ that is $\alpha_t(X_t) = (I - 2S_t'S_t)X_t + 2S_t'b_t$ in (10), we get (11) (formal proof in Appendix A). Choosing $S_t = I$ and $b_t = \mu_{x_t}$ makes $\alpha^-(X_1^T) = 2\mu_{X_1^T} - X_1^T$, which by point symmetry satisfies (9). Therefore, we get $D_E = \frac{1}{T}\sum_{t=1}^{T} \mathrm{tr}(R_{X_t})$. Note that the optimal distortion, denoted as $D_E^\star$ and obtained by optimizing over S_t and b_t, satisfies $D_E^\star \geq \frac{1}{T}\sum_{t=1}^{T} \mathrm{tr}(R_{X_t})$. However, from (5), we have $D_E^\star \leq \frac{1}{T}\sum_{t=1}^{T} \mathrm{tr}(R_{X_t})$. Therefore, the selected S_t and b_t and the corresponding distortion values are optimal. ■

Now, we show the implications of our results for mirroring-based schemes in the context of a few examples.

Example 1: Consider an example where U is distributed as Gaussian with mean μ_U and covariance matrix R_U. Then, for a noiseless system with perfect observation and a zero initial state, X_2^T is also Gaussian distributed with mean $\mu_{X_2^T} = Q\mu_U$ and variance $R_{X_2^T} = QR_UQ^T$, where Q relates U to X_2^T after unfolding the time-dependent state equations into the form $X_2^T = QU$. A Gaussian random vector has point symmetry and therefore, according to Corollary III.3, we can get maximum distortion by setting $b_t = \mu_{X_t}$ and $S_t = I$.

The next example is based on a Markov-based model for the dynamical system. For this example, the following lemma is useful.

Lemma III.4: Consider the random vector X_1^T where the following conditions hold: $f_{X_1}(x_1)$ has point symmetry; and $f_{X_t|X_1^{t-1}}(x_t|x_1^{t-1})$ has point symmetry, then so does $f_X(X)$, where $X = X_1^T$ and $\mu = [\mu_{x_1}'\ \mu_{x_2}'\ \cdots\ \mu_{x_T}']'$. Therefore, by virtue of Corollary III.3, mirroring schemes can achieve the maximum distortion.

Lemma III.4 allows us to characterize the performance of the following example.

Example 2: Consider the following random walk mobility model. Let $a \in \mathbb{N}^+$, and X_t be its location at time t, then

$$X_1 \sim \mathrm{Uni}([-a : a])$$
$$X_t|X_{t-1} \sim \mathrm{Uni}([-a : a] \cap \{X_{t-1} - 1, X_{t-1}, X_{t-1} + 1\}).$$

This example follows the system model in (1) by assuming a noiseless system with U_t to be independent across t, and to be uniformly distributed among $\{-1, 0, 1\}$ when $X_t \in [-a + 1 : a - 1]$, U_t uniformly distributed in $\{0, 1\}$ when $X_t = -a$, and U_t uniformly distributed in $\{-1, 0\}$ when $X_t = a$. One can see that these distributions satisfy the conditions in Lemma III.4. Therefore, one can set $b_t = \mu_t = 0$ and $S_t = 1$, which will achieve maximum distortion of D_E.

Example 3: Here, we provide a numerical example that shows how our mirroring scheme performs for situations where we compute the state distributions using numerical simulations. In Section IV, we will also show that the controller used in this example falls under the class of controller where we do not need to compute the distribution on states and can directly apply our scheme to achieve the perfect distortion. We consider the quadrotor dynamical system provided in [23, eq. (4)]. The quadrotor moves in a 3-D cubed space with a width, length, and height of 2 m, where the origin is the center point of the space. The quadrotor starts its trajectory from an initial point $(-1, y_1, z_1)$ and finishes its trajectory at a target point $(1, y_T, z_T)$ after T time steps, where the points y_1, z_1, y_T, and z_T are picked uniformly at random in $[-1, 1]^4$. We assume that $T = 10$ time steps, and that the continuous model in [23, (4)] is discretized with a sample time of $T_s = 0.5$ s. We assume that the quadrotor encodes and transmits only the states that contain the location information (first three elements of the state vector X_t). The quadrotor is equipped with an LQR controller that designs the input sequence U_1^{T-1} as the solution of the following problem:

$$\begin{aligned} \text{minimize} \quad & \|U\|^2 + 10\left\|X_2^{T-1}\right\|^2 \\ \text{subject to} \quad & X_{t+1} = A^{\text{quad}}X_t + B^{\text{quad}}U_t \quad \forall t \in [T-1] \\ & X_1 = \begin{bmatrix}-1\ y_1\ z_1\ 0 \cdots 0\end{bmatrix}' \\ & X_T = \begin{bmatrix}1\ y_T\ z_T\ 0 \cdots 0\end{bmatrix}' \end{aligned} \tag{12}$$

where A^{quad} and B^{quad} define the quadrotor's discrete-time model. The remaining states of X_1 and X_T are set to zero to allow the drone to hover at the respective locations. We perform numerical simulation of the aforementioned setup: we run 2 millions iterations, where in each iteration, a new initial and target points are picked, and the resultant trajectory is recorded. Based on the recorded data, we consider different mirroring schemes and numerically evaluate the attained distortion. To facilitate numerical evaluations, the simulation space is gridded into bins with 0.2 m of separation, and the location of the drone at each trajectory is approximated to the nearest space bin.

Fig. 4 shows some of the drone trajectories obtained from our numerical simulation. It is clear that not all trajectories are equiprobable, and therefore the distribution of X_t is not uniform across all bins in space. Since the motion of the drone is mainly progressive in the positive x-axis direction, reflection across a

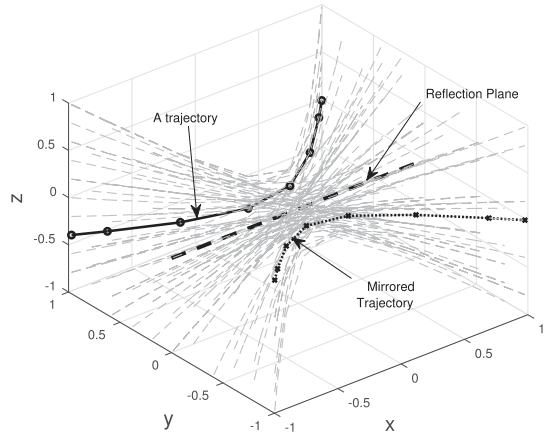

Fig. 4. Illustration of some trajectories. The reflection plane is shown as a dashed black line. One trajectory (solid black) is shown along with its mirrored image (dotted black).

fixed point results in mirrored trajectories that are progressing in the opposite direction, and therefore are identified to be fake automatically. Therefore, mirroring across a point here is useless: the numerically computed distortion for this scheme is equal to zero.

Next, we consider mirroring across the reflection plane shown in Fig. 4, where $b_t = 0$ and $S_t = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. As can be seen from the figure, the reflection plane is indeed an axis of symmetry for the distribution of the drones trajectories, and therefore is expected to provide high distortion values. We numerically evaluate the attained distortion using the scheme by using (8), which evaluates to $D_E = 0.3971$. This is slightly less than $D_E^{\max} = 0.3979$.

B. Encoding Schemes With k-bits Shared Secret Key

The scheme in (7) assumes the use of 1 b for encryption. However, it is straightforward to extend the scheme when we require a larger ambiguity set. For k bits, we denote the possible values of the shared key as $K \in [0 : 2^k - 1]$. Therefore, the scheme works as follows:

$$Z_t(K) = \alpha_t^{(K)}(X_t) \; \forall t \in [T] \tag{13}$$

where $\alpha_t^{(K)}$ is an invertible transformation function used at time t when the value of the key is K, and $\alpha_t^{(0)}(x) = \alpha_t^{-(0)}(x) = x$. The following theorem shows the achieved value of the distortion in this case, which is a direct extension of Theorem III.1.

Theorem III.5 (Proof in Appendix B): The average distortion D_E attained by using the scheme in (13) is

$$\frac{1}{2^k T} \sum_{t=1}^{T} \mathbb{E}_X \left\{ \frac{\sum_{K=0}^{2^k-1} f_X(\alpha^{-(K)}(X)) \left\| R_t^{(K)} \right\|^2}{\left[\sum_{K=0}^{2^k-1} f_X(\alpha^{-(K)}(X)) \right]^2 f_X(X)} \right\} \tag{14}$$

where $R_t^{(K)} = \sum_{\ell=0}^{2^k-1} f_X(\alpha^{-(\ell)}(X))(\alpha_t^{-(\ell)}(X_t) - \alpha_t^{-(K)}(X_t))$ and $\alpha^{-(K)}(X) := [\alpha_1^{-(K)}(X_1)' \; \alpha_2^{-(K)}(X_2)' \; \cdots \; \alpha_T^{-(K)}(X_T)']'$. Moreover, if the following condition holds:

$$f_X(\alpha^{-(K)}(x)) = C(x) \qquad \forall x \in \mathcal{X} \quad \forall K \in [0 : 2^k - 1] \tag{15}$$

where $C(x)$ is a constant, then D_E simplifies to

$$\frac{1}{2^{3k} T} \sum_{t=1}^{T} \mathbb{E}_X \left[\sum_{K=0}^{2^k-1} \left\| \sum_{\ell=0}^{2^k-1} \left(\alpha_t^{-(K)}(X_t) - \alpha_t^{-(\ell)}(X_t) \right) \right\|^2 \right]. \tag{16}$$

Theorem III.5 shows the average distortion attained for general schemes that use k bits. In addition, we can generalize the mirroring scheme in Section III to utilize k bits as follows. Given a k-bit key, then we select a set of k hyperplanes (i.e., a set of k parameters, $S_t^{(K)} \in \mathbb{R}^{d \times n}$, and $b_t^{(K)} \in \mathbb{R}^d$) for each time step t. Then, let $\mathcal{K}$ be a set of binary values corresponding to the binary representation of the k-bit key. The mirroring scheme would transform the point x to $\prod_{K \in \mathcal{K}} (I - 2S_t^{(K)'} S_t^{(K)}) x + 2S_t^{(K)'} b_t$, i.e., x is mirrored across the hyperplanes corresponding to the 1-valued bits in the binary representation of the shared key. It is not difficult to see that this scheme can achieve the maximum distortion when X_t is Gaussian distributed with zero mean and covariance matrix R_t and independent across t: for this case, $S_t^{(K)}$, $K \in \mathcal{K}$, are chosen as the eigenvectors of R_t and $b_t^{(K)} = 0$.

Using multiple bits of shared keys can provide benefits beyond having a larger ambiguity set. In fact, while we show the optimality of 1-b mirroring schemes for distributions with point symmetries, using multiple bits of shared key can provide a better distortion for general distributions. For example, it was shown that, for a general finite alphabet: 1-b schemes are not sufficient to achieve the maximum distortion; and with just 5 b of shared keys, a scheme achieves more than 97% of the maximum possible distortion [16].

IV. Transformations Maintaining Point Symmetry

Encoding and decoding schemes, such as the ones mentioned in Section III, can be generally used for any dynamical system with arbitrary distributions on the inputs U, the state vectors $\widetilde{X}_t$, and the state estimates X_t. However, characterizing the attained level of average distortion [using expressions (8) and (14)] requires the knowledge of the distribution of the state estimate. While a distribution can be obtained for the initial and target state vectors, it may be difficult to incorporate the system dynamics, the estimation method as well as the controller into the process of finding a distribution of the inputs, states, and states estimate. In such cases, numerical evaluations can aid into finding the needed distribution, as was shown in Example 3 in Section III. Although it is necessary to find the state distribution in order to characterize the distortion, the knowledge of existing symmetries in the distribution can directly give possible choices for the transformation function $\alpha_t(\cdot)$, which may attain high levels of distortion; for example, if there is a point symmetry

in the distribution, mirroring across the symmetry point attains the maximum possible distortion. In this section, we ask the following question: "*under which conditions on the dynamical system, does point symmetry in the initial and target states results into point symmetry on the states estimate*?"

A general answer to the aforementioned question appears to be difficult. Therefore, we limit our answer in this article to the scope of linear controllers. For a given initial and target states, let $X^{(\text{ref})}$ be the reference trajectory that the control system ideally wishes to follow. We assume that the system controller selects an input vector that is a linear function of $X^{(\text{ref})}$. In many cases, $X^{(\text{ref})}$ is also a linear function of the initial and target states (e.g., when the reference trajectory is the solution of an LQR problem for the noiseless version of the system). Then, we can write $U_t = K_t(X_t - X_t^{(\text{ref})})$. Moreover, we assume that the optimal estimation function that the system uses a linear one in the observations, i.e., we assume that X_t is a linear function of Y_1^t, X_{init}, and X_{target}. By incorporating the controller and estimation equations into the system dynamics, one can arrive at the relation $X = MQ$, where $Q = [X'_{\text{init}} \;\; X'_{\text{target}} \;\; w_1^{T'} \;\; v_1^{T'}]$, and the matrix M is a function of the matrices A, B, C, and K_t and the linear function used in the estimation of state X_t from the observations. We assume that w_1^T and v_1^T are uncorrelated Gaussian random vectors. We first prove the following lemma.

Lemma IV.1: If a random vector $V_1 \in \mathbb{R}^n$ has point symmetry across μ_{V_1}, and g is an affine function, then the random vector $V_2 = g(V_1)$ has point symmetry across $g(\mu_{V_1})$.

Proof: If V_1 has point symmetry, then the following conditions are equivalent:

$$f_{V_1}(v_1) = f_{V_1}(2\mu_{V_1} - v_1) \qquad \forall v_1 \in V_1$$
$$f_{V_1}(v_1) = f_{2\mu_{V_1} - V_1}(v_1) \qquad \forall v_1 \in V_1.$$

Thus, to prove that V_2 also has point symmetry, it suffices to prove that the densities of V_2 and $2\mu_{V_2} - V_2$ are the same. Consider the two random vectors W_1 and W_2. If they have the same support and the same density function, then $g(W_1)$ and $g(W_2)$ will also have the same density for any function g; we denote this by writing $W_1 \sim W_2$. Thus

$$\begin{aligned} V_1 &\sim 2\mu_{V_1} - V_1 \\ g(V_1) &\sim g(2\mu_{V_1} - V_1) \\ M_1V_1 + M_2 &\sim 2M_1\mu_{V_1} - M_1V_1 + M_2 \\ V_2 &\sim 2(M_1\mu_{V_1} + M_2) - (M_1V_1 + M_2) \\ V_2 &\sim 2(\mu_{V_2}) - V_2. \end{aligned}$$

Thus, V_2 has a point of symmetry. ■

Theorem IV.2: If X_{init} and X_{target} are independent of w_1^T and v_1^T, and both have point symmetries, then the vectors X_t as well as X will all have point symmetries for any matrix M.

Proof: First, note that w_1^T and v_1^T are Gaussian random vectors, and, therefore, have point symmetries across their mean points. Since X_{init} and X_{target} are independent of w_1^T and v_1^T, then the vector Q also has a point symmetry across the mean point (which is the concatenation of the mean points of the respective components of Q); we denote this point by μ_Q. Then, by virtue of Lemma IV.1, X (respectively X_t) will also have point symmetry across the point $M\mu_Q$ (respectively across the point μ_Q left multiplied by the corresponding section of the matrix M). ■

Revisiting Example-3 of Section III. Example 3 in Section III shows an example where the initial and target points exhibit point symmetry. In such an example, the LQR controller is a linear function of the previous states (one can find such a controller by applying the KKT conditions). Since the system is noiseless, then the estimated states are equal to the observations. Therefore, the conditions for Theorem IV.2 are met, and point symmetry is preserved for X_t and the whole trajectory X. Note, however, that the point of symmetry for X_t changes with t, i.e., it progresses along the x-axis as shown in Fig. 4.

V. Optimizing the Worst Case Distortion D_W

The expected distortion metric might not be well suited for some applications (for example, if an adversary wants to shoot a drone). In this case, the adversary's estimate needs to be far from the actual state *at all* time instances. Therefore, a more appropriate metric would be to consider the worst case distortion for the adversary. Consider, for example, the scheme in Fig. 2. Here, the adversary's estimate is always the center point and therefore the maximum expected distortion is achieved. However, when the drone is close to the center, its mirror image will also be close to the center. At this particular time instance, the adversary's distortion will be very small and thus the adversary will essentially know the position.

In this section, we present an encryption scheme that attempts to maximize the worst case distortion for Eve. The main idea is to obfuscate the initial state in such a way that Eve, even if she optimally, uses her knowledge about the dynamics and her observations, her best estimate is close to the maximal distortion. We start by studying the problem of distorting the transmission of a single random variable in Theorems V.2 and V.3. These results then form the basis for maximizing the worst case distortion of a trajectory, as described in Theorem V.4.

A. Building Step: Scalar Case

Consider the case where the system wants to communicate a single scalar random variable X to Bob by transmitting Z. The worst case distortion D_W for Eve will be $D_W = \min_Z \text{Var}(X|Z)$. Note that if Eve does not overhear Z, Eve uses the minimum mean square error estimate (i.e., the mean value) as her estimate, and thus experience a worst case distortion equal to the variance of X.

We first assume that $X \sim \mathcal{N}(0, 1)$, and, thus the worst case distortion cannot be larger than 1 by (6). We next develop our scheme progressively, from simple to more sophisticated steps. We will also use the following lemma.

Lemma V.1: The variance of a Bernoulli random variable taking values a and b with probabilities p_a and p_b, respectively, is given by $p_a p_b (a-b)^2$.

Mirroring. Reflecting around the origin (as we did for optimizing the average case distortion in Section III) does not work well when X takes small values: indeed $\text{Var}(X|Z)$ is $\Pr(X = Z|Z)(\Pr(X = -Z|Z))(Z - (-Z))^2$ using Lemma V.1

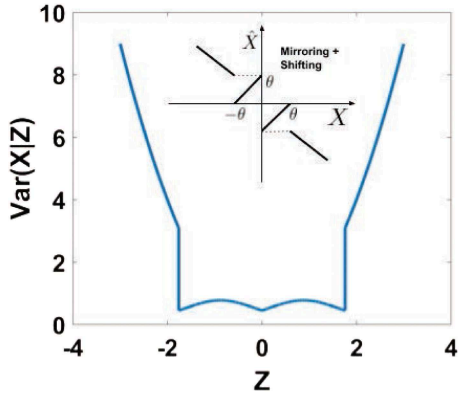

Fig. 5. Var$(X|Z)$ versus Z for the shifting+mirroring scheme with $\theta_1 = 1.76$; $D_W = 0.4477$. Var$(X|Z)$ is Z^2 for $|Z| > \theta_1$ and is $pZ^2 + (1-p)\tilde{Z}^2 - (pZ + (1-p)\tilde{Z})^2$ for $|Z| \leq \theta_1$, where $p = f(Z)/(f(Z) + f(\tilde{Z}))$, with $f(Z) \sim \mathcal{N}(0,1)$, and $\tilde{Z} = Z + \theta_1 \text{mod}\,[-\theta_1, \theta_1)$.

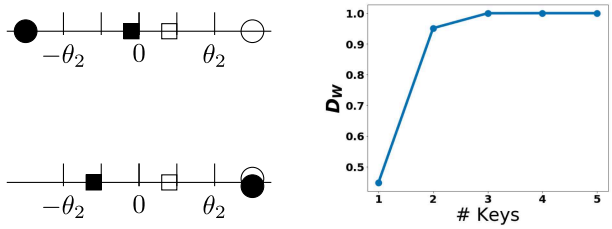

Fig. 6. (a) Transparent shapes represent true values and solid shapes represent their respective mapping when 2-b key is 11 and 10, respectively. (b) D_W as a function of number of keys for optimal choice of θ_k.

and has a worst case value that goes to zero as Z approaches zero.

Shifting. To avoid this, we could try to use a "shifting" scheme where we add a constant θ to X whenever the shared key bit is one; but now this scheme does not perform well for large values of Z: as Z increases Var$(X|Z)$ goes to zero. This is because using Lemma V.1

$$\begin{aligned}\text{Var}(X|Z) &= \Pr(X = Z|Z)(\Pr(X = Z-\theta|Z))(Z-(Z-\theta))^2 \\ &= \Pr(X = Z|Z)(\Pr(X = Z-\theta|Z))(\theta)^2\end{aligned}$$

and $\Pr(X = Z|Z)(Pr(X = Z - \theta|Z))$ goes to zero for large value of Z.

Shifting+Mirroring. We here combine shifting and mirroring, in order to achieve a good performance for both small and large values of X. We start from the case where we have $k = 1$ b of key and then go to the case $k \geq 1$.

1) $k = 1$. We select $\theta_1 \in \mathbb{R}$ that determines a window size (θ_1 is public and known by Eve). The encoding function is

$$Z = \mathcal{E}(X, K) = \begin{cases} X & \text{if } K = 0 \\ -X & \text{if } K = 1,\ |X| > \theta_1 \\ X + \theta_1 & \text{if } K = 1,\ -\theta_1 \leq X < 0 \\ X - \theta_1 & \text{if } K = 1,\ 0 \leq X < \theta_1. \end{cases}$$

We note that there is one particular value of X, $X = \theta_1$, which we do not transmit. Since this is of zero probability measure, it can be safely ignored. Given Z, there are two possibilities for X

$$X \in \begin{cases} \{Z, -Z\} & \text{if } |Z| > \theta_1 \\ \{Z, Z + \theta_1\} & \text{if } -\theta_1 \leq Z < 0 \\ \{Z, Z - \theta_1\} & \text{if } 0 \leq Z < \theta_1. \end{cases}$$

Using the fact that $X \sim \mathcal{N}(0,1)$, we can calculate the posterior probabilities $\Pr(X|Z)$ and use Lemma V.1 to compute Var$(X|Z)$. Fig. 5 plots Var$(X|Z)$ for $\theta = 1.76$. The worst case distortion in this case becomes 0.4477, which is the best we can hope for if we have only 1 b of shared key. This follows because for any mapping from X to Z, a transmitted symbol Z can have at most two preimages (as Bob needs to reliably decode with 1 b of key), and if one of these is $X = 0$, then no matter what the second one is, the distortion corresponding to Z will be at most 0.4477. Equality occurs when the second preimage of Z is ± 1.76. Note that our scheme also maps 0 to -1.76 (for $\theta_k = 1.76$).

2) $k \geq 1$. For $K \in \{0,1\}^k$, we use the following encoding:

$$\begin{aligned} Z = \mathcal{E}(X, K) \qquad\qquad (17) \\ = \begin{cases} \begin{cases} X & \text{if } K < 2^{k-1} \\ -X & \text{if } K \geq 2^{k-1} \end{cases} & |X| > \theta_k \\ X + K\frac{2\theta_k}{2^k} \text{mod}\,[-\theta_k, \theta_k) & X \in [-\theta_k, \theta_k) \end{cases}\end{aligned}$$

where the optimal value of the constant θ_k depends on the number k of keys we have, K is the decimal equivalent of a binary string of length k, and $r \text{mod}\,[a,b) = r - i(b-a)$ is such that i is an integer and $r - i(b-a) \in [a,b)$ for $r, a, b \in \mathbb{R}$. Intuitively, if $|X| > \theta_k$, then for half of the keys, we reflect across origin, and for the other half, we do nothing; if $|X| < \theta_k$, we divide this window of size $2\theta_k$ into 2^k equal size windows and shift a point from one window to another by jumping K (in decimal) windows. An example for $k = 2$ is shown in Fig. 6(a) for the key values $K = 11$ and $K = 10$. Fig. 6(b) plots D_W as a function of the number of keys k. Using $k = 3$ and $\theta_3 = 4.84$, we achieve $D_W = 0.9998$, which is very close to 1, the best we can hope for.

Remark: We optimize the parameter θ_k of our scheme assuming Gaussian distribution. In particular, $\theta_k = \arg\max_{\theta_k}(\min_Z D_W(Z))$ and $D_W(Z)$ is Z^2 if $|Z| > \theta_k$, and is $\sum_{K\in\{0,1\}^k}(p_K(Z_K)^2) - (\sum_{K\in\{0,1\}^k} p_K Z_K)^2$, if $|Z| \leq \theta_k$. Here, $p_K = f(Z_K)/\sum_{L\in\{0,1\}^k} f(Z_L)$ with $f(Z_K) \sim \mathcal{N}(0,1)$, and Z_K for $K \in \{0,1\}^k$ is defined as $Z_K = Z + K_d\frac{2\theta_k}{2^k}$ mod $[-\theta_k, \theta_k]$, with K_d being the decimal equivalent of K.

We pick a choice of θ_k by computationally iterating over the values of θ_k to find one that maximizes D_W, i.e., $\theta_k = \arg\max_{\theta_k} \min_Z(D_W(Z))$.

For other distributions, the optimal choice of θ_k and the corresponding worst case distortion would be different.

Theorem V.2: A Gaussian random variable with mean μ and variance σ^2 can be near perfectly (~ 0.9998 times the perfect

distortion) distorted in worst case settings by just using three bits of shared keys.

Proof: Generate the random variable $V \sim \mathcal{N}(0,1)$ as $V = (X-\mu)/\sigma$ and encrypt it using $k=3$ key bits and the previously described scheme. We transmit the mean μ and the variance σ^2 uncoded, and show that near perfectly distorting the standard Gaussian V results in near perfect distortion of X for Eve. For $c = 0.9998$, we have

$$D_W = \min_Z \mathrm{Var}(X|Z) = \min_Z \mathrm{Var}(\sigma V + \mu | Z)$$
$$= \sigma^2 \min_Z \mathrm{Var}(V|Z) = c\sigma^2.$$

■

B. Vector Case and Time Series

Theorem V.3 (Proof in Appendix C): For a Gaussian random vector $X \in \mathbb{R}^n$ with mean μ and a diagonal covariance matrix Σ, we can achieve D_W within 0.9998 of the optimal by using $3n$ bits of shared keys.

This theorem uses our 3-b encryption for each element in the vector. Assume now that this vector captures the probability distribution of the initial state of dynamical system; by encrypting this state, we can guarantee the following.

Theorem V.4 (Complete Proof in Appendix D): Using $3n$ bits of shared keys, the shifting+mirroring scheme achieves $D_W \geq c \cdot \mathrm{tr}(|\Lambda|^{2t}\Sigma)$ with $c = 0.9998$ for the dynamical systems (1) with $C = I$, $v_t = 0$, the singular value decomposition of A is $A = \Phi \Lambda V^H$, and initial state $X_1 \sim \mathcal{N}(\mu, \Sigma)$, where Σ is diagonal covariance matrix, and U_t and w_t are independent of X_t. Moreover, if $|\lambda_i| \geq 1, \forall i$, where λ_i is the ith singular value of A, then D_W is within 0.9998 of the maximum distortion.

Remark: Although the independence assumption on the inputs is rather restrictive, the result serves as a stepping stone toward understanding general cases.

Proof: The system transmits $Z_1 = f(Y_1, K) = f(X_1, K)$ where f is the encoding in Theorem V.3, and for $t \in [T-1]$

$$Z_{t+1} = AZ_t + (Y_{t+1} - AY_t) = AZ_t + BU_t + w_t.$$

Bob can decode X_1 using Z_1 and K. Then

$$\hat{X}_{t+1} = Z_{t+1} - AZ_t + A\hat{X}_t$$
$$= (AZ_t + BU_t + w_t) - AZ_t + A\hat{X}_t$$
$$= AX_t + BU_t + w_t = X_{t+1} \qquad \forall t \in [T-1].$$

Eve's distortion is calculated in Appendix D. ■

Complexity. $\mathcal{O}(n^2)$ per time instance for both encoding and decoding.

Case study. We take three choices of A of sizes 2×2, first having all singular values no smaller than one, in particular $[1.01, 1]$, second having singular values $[1.5, 0.5]$, and third having singular values of $[0.8, 0.9]$. For a given co-variance matrix $\Sigma = [2, 0; 0, 3]$ for the initial state, we plot the evolution of distortion at the adversary's end corresponding to our encryption scheme and compare with $c\mathrm{tr}(\Sigma)$. This evolution is shown in Fig. 7. As we can observe, when A has all the singular values more than one, the distortion at adversary's end is always at least $c\mathrm{tr}(\Sigma)$, whereas for other cases in eventually goes to zero.

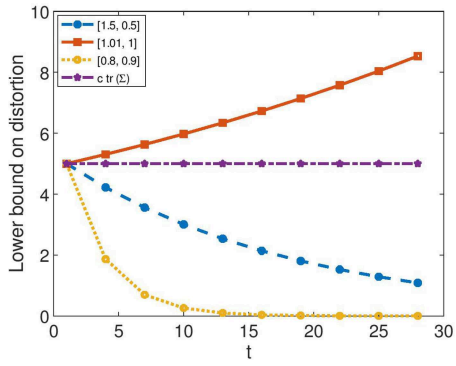

Fig. 7. Evolution of the distortion for Eve for shifting+mirroring-based scheme.

VI. Distorting the Inputs and the Implication on States

In many situations, it is easier to obtain a handle on the distribution of the input sequence than on the distribution of the state transition sequence. Moreover, in some situations, a simple transformation of the original trajectory would lead to a fake trajectory that does not obey the system dynamics and thus can be detected by the adversary—alternatively, the state trajectory distribution does not have useful symmetry properties. Motivated by these, here we consider a different setup where Alice encodes and transmits the input sequence to Bob instead of the state transition sequence, i.e., $Z_t := \mathcal{E}_t(U_t, K)$. Under this setup, using the mirroring-based scheme on inputs, one can provide guarantees on the level of average/worst case distortion for Eve's estimate of the inputs. We then ask the following question: if Alice encodes and transmits the input vectors, how does the guarantees on average and worst case distortions on the *inputs* translate to the guarantees on average and worst case distortions on the *states*?

Formally, we consider the system model in (1) with zero noise, i.e., $w_t = 0$. Following the definition of average and worst case distortions in (3) and (4), respectively, the distortions on inputs and states vectors are as follows:

Expected distortions:

$$D_E^{(X)} = \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} \mathrm{tr}\left(R_{X_t|Z}\right), \quad D_E^{(U)} = \frac{1}{T}\mathbb{E}_Z \sum_{t=0}^{T-1} \mathrm{tr}\left(R_{U_t|Z}\right)$$

Worst case distortions:

$$D_W^{(X)} = \min_Z \min_t \mathrm{tr}\left(R_{X_t|Z}\right), \quad D_W^{(U)} = \min_Z \min_t \mathrm{tr}\left(R_{U_t|Z}\right).$$

The following theorem provides a relation between the distortion on the input vector and the distortion on the state vector for both expected and worst case scenarios. In particular, it provides relations between D_E^X, D_E^U and D_W^X, D_W^U.

Theorem VI.1: If the input vectors U_t, $t \in [T-1] \cup \{0\}$, satisfy the following condition:

$$\mathbb{E}_Z \left[\operatorname{tr} \left(\underbrace{\sum_{i=1}^{t} \sum_{j=i+1}^{t} B'(A^{t-j})' A^{t-i} B R_{U_{i-1}U_{j-1}|Z}}_{\Phi_t} \right) \right] \geq 0 \tag{18}$$

then, for given distortions $D_E^{(U)}$ and $D_W^{(U)}$, the following bounds holds:

$$D_E^{(X)} \geq \lambda_{\min}(B'B)\, D_E^{(U)}, \quad D_W^{(X)} \geq \lambda_{\min}(B'B)\, D_W^{(U)}$$

where $\lambda_{\min}(B'B)$ is the minimum eigenvalue of $B'B$.

Theorem VI.1 gives a lower bound on the distortion level of the state vectors when distorting the inputs. The bound holds when the condition (18) holds. Examples where condition (18) holds are open-loop control systems where the distribution on the inputs has a point of symmetry. We expand more on condition (18) after the proof of Theorem VI.1.

Proof: We start by introducing the following notation: $Z = Z_0^{T-1}$, $U = U_0^{T-1}$, and $X = X_1^T$. Moreover, without loss of generality, we assume that $X_0 = 0$. The states of the noiseless dynamical system can be expressed as $X_t = \sum_{i=1}^{t} A^{t-i} B U_{i-1}$. Then, we can write

$$R_{X_t|Z} = \sum_{i=1}^{t} A^{t-i} B R_{U_{i-1}|Z} B'(A^{t-i})' + 2\Phi_t. \tag{19}$$

Therefore, combining (18) and (19), we express $D_E^{(X)}$ as

$$\begin{aligned} D_E^{(X)} &\geq \frac{1}{T} \mathbb{E}_Z \sum_{t=1}^{T} \sum_{i=1}^{t} \operatorname{tr}\left(A^{t-i} B R_{U_{i-1}|Z} B'(A^{t-i})'\right) \\ &\overset{(a)}{\geq} \frac{1}{T} \mathbb{E}_Z \sum_{t=1}^{T} \operatorname{tr}\left(B R_{U_{t-1}|Z} B'\right) = \frac{1}{T} \mathbb{E}_Z \sum_{t=1}^{T} \operatorname{tr}\left(B' B R_{U_{t-1}|Z}\right) \\ &\geq \lambda_{\min}(B'B) \frac{1}{T} \mathbb{E}_Z \sum_{t=1}^{T} \operatorname{tr}\left(R_{U_{t-1}|Z}\right) = \lambda_{\min}(B'B) D_E^{(U)} \end{aligned}$$

where (a) follows by noting that the matrices $B'(A^{t-i})' R_{U_{i-1}|Z} A^{t-i} B$ are positive semidefinite, and therefore their trace is greater than or equal to zero. To see that they are indeed positive semidefinite, note that $R_{U_{i-1}|Z}$ is positive semidefinite, therefore it has the eigendecomposition $R_{U_{i-1}|Z} = \Sigma \Lambda \Sigma'$. Therefore, the claim follows by noting that, for any vector z, we have

$$\begin{aligned} &z' B' \left(A^{t-i}\right)' R_{U_{i-1}|Z} A^{t-i} B z \\ &\quad = z' B' \left(A^{t-i}\right)' \Sigma \Lambda^{1/2} \Lambda^{1/2} \Sigma' A^{t-i} B z \\ &\quad = \|\Lambda^{1/2} \Sigma' A^{t-i} B z\|^2 \geq 0. \end{aligned}$$

Identical arguments can be made to show the bound on $D_W^{(X)}$. ∎

Next, we show some sufficient conditions which ensure that condition (18) holds.

1) U_i and U_j are uncorrelated for $i \neq j$ and $U := U_0^{T-1}$ has point symmetry: In this case, the optimal mirroring scheme is to mirror the point U across the point of symmetry. Therefore, given Z, U_i takes two values: Z_i with probability p_Z and $\tilde{Z}_i$ with probability $1 - p_Z$, where p_Z is equal to $p_Z = f_U(Z)/(f_U(Z) + f_U(\tilde{Z})) = 0.5$, which follows from the point symmetry assumption on U. So we have $\mathbb{E}[U_i|Z] = \frac{Z_i + \tilde{Z}_i}{2}$. Therefore, $R_{U_iU_j|Z}$ can be computed as follows:

$$\begin{aligned} R_{U_iU_j|Z} &= \mathbb{E}_{U|Z}\left[(U_i - \mathbb{E}[U_i|Z])\,(U_j - \mathbb{E}[U_j|Z])'\right] \\ &= \frac{1}{4}\left(Z_i - \tilde{Z}_i\right)\left(Z_j - \tilde{Z}_j\right)' = (Z_i - \mu_{Z_i})\,(Z_j - \mu_{Z_j})'. \end{aligned}$$

Then, we have $E_Z R_{U_iU_j|Z} = R_{U_iU_j} = 0$ by noting that Z_i and Z_j have the same distribution as U_i and U_j

$$\begin{aligned} f_{Z_i,Z_j}(z_i, z_j) &= \frac{1}{2}\left(f_{U_i,U_j}(z_i, z_j) + f_{U_i,U_j}(\tilde{z}_i, \tilde{z}_j)\right) \\ &= \frac{1}{2}\left(f_{U_i}(z_i) f_{U_j}(z_j) + f_{U_i}(\tilde{z}_i) f_{U_j}(\tilde{z}_j)\right) = f_{U_i,U_j}(z_i, z_j). \end{aligned}$$

2) A and $R_{U_i,U_j|Z}$ are positive semidefinite matrices for all i, j, and Z: This follows because if A is positive semidefinite, then so is A^i for any value of i. Therefore, we can write

$$\begin{aligned} &\operatorname{tr}\left(B'(A^{t-j})' A^{t-i} B R_{U_{i-1}U_{j-1}|Z}\right) \\ &= \operatorname{tr}\left(B' A^{t-j} A^{t-i} B R_{U_{i-1}U_{j-1}|Z}\right) \\ &= \operatorname{tr}\left(B' A^{2t-i-j} B R_{U_{i-1}U_{j-1}|Z}\right) \\ &= \operatorname{tr}\left(B' \Sigma \Lambda^{1/2} \Lambda^{1/2} \Sigma' B R_{U_{i-1}U_{j-1}}\right) \\ &= \operatorname{tr}\left(\Lambda^{1/2} \Sigma' B R_{U_{i-1}U_{j-1}|Z} B' \Sigma \Lambda^{1/2}\right) \\ &= \operatorname{tr}\left(\Lambda^{1/2} \Sigma' B \Phi \Gamma^{1/2} \underbrace{\Gamma^{1/2} \Phi' B' \Sigma \Lambda^{1/2}}_{D}\right) = \operatorname{tr}(D'D) \geq 0. \end{aligned}$$

VII. Conclusion

In this article, we considered distortion-based security for CPSs as a complementary security approach that optimizes an alternative security goal. This approach for security is suitable for CPS applications where the estimation of the adversary about the states is required to be "far" from the actual state value. We provided security schemes that aim to optimize for both the average and worst case distortions. For the average distortion, we showed the surprising result that 1-b schemes are optimal for certain distributions. We then provided the expression for the attained level of distortion for a general security scheme. For worst case distortion, we considered an initial situation where we proposed an encryption scheme that achieves near optimal distortion.

Appendix

A. Proof of Theorem III.1 and Corollary III.3

We start by computing $R_{X_t|Z_1^T}$. Note that given a sequence of transmitted symbol Z_1^T, there are two possible values of

sequence of message symbols X_1^T that are $X_1^T = Z_1^T$ and $X_1^T = \tilde{Z}_1^T$, where $\tilde{Z}_t$ is $\alpha_t^-(Z_t)$ and $\tilde{X}_t$ is $\alpha_t^-(X_t)$.

The posterior probability of $X_t = Z_t$ given Z_1^T, i.e., $\Pr(X_t = Z_t|Z_1^T)$, will be equal to $\Pr(X_1^T = Z_1^T|Z_1^T) := p_Z$. We note that $p_Z = \frac{f(Z)}{f(Z)+f(\tilde{Z})}$, where $\tilde{Z} := [\tilde{Z}_1' \ \tilde{Z}_2' \ \cdots \ \tilde{Z}_T']'$. Then, $\mathbb{E}(X_t|Z_1^T) = p_Z Z_t + (1-p_Z)(\tilde{Z}_t)$. With this

$$\begin{aligned}
R_{X_t|Z_1^T} &= \mathbb{E}_{X_t|Z_1^T}\left[\left(X_t - \mathbb{E}(X_t|Z_1^T)\right)\left(X_t - \mathbb{E}(X_t|Z_1^T)\right)'\right] \\
&= p_Z(1-p_Z)^2(Z_t-\tilde{Z}_t)(Z_t-\tilde{Z}_t)' \\
&\quad + (1-p_Z)p_Z^2(Z_t-\tilde{Z}_t)(Z_t-\tilde{Z}_t)' \\
&= p_Z(1-p_Z)(Z_t-\tilde{Z}_t)(Z_t-\tilde{Z}_t)' \\
D_E &= \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} \text{tr}\left(R_{X_t|Z_1^T}\right) \\
&= \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} \text{tr}\left(p_Z(1-p_Z)(Z_t-\tilde{Z}_t)(Z_t-\tilde{Z}_t)'\right) \\
&= \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} p_Z(1-p_Z)\text{tr}\left((Z_t-\tilde{Z}_t)(Z_t-\tilde{Z}_t)'\right) \\
&= \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} p_Z(1-p_Z)\|Z_t-\tilde{Z}_t\|^2 \\
&= \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} \frac{f_X(Z)f_X(\tilde{Z})}{(f_X(Z)+f_X(\tilde{Z}))^2}\|Z_t-\tilde{Z}_t\|^2.
\end{aligned}$$

Now, Z_1^T is the transmitted symbols if $X_1^T = Z_1^T$ and key was zero or if $\{X_t = \tilde{Z}_t, \ \forall t \in [T]\}$ and key was one. So, $f_Z(Z) = \frac{f_X(Z)+f_X(\tilde{Z})}{2}$. Thus, D_E

$$\begin{aligned}
&= \frac{1}{T}\mathbb{E}_Z \sum_{t=1}^{T} \frac{f_X(Z)f_X(\tilde{Z})}{(f_X(Z)+f_X(\tilde{Z}))^2}\|Z_t-\tilde{Z}_t\|^2 \\
&= \frac{1}{T}\int f_Z(Z) \sum_{t=1}^{T} \frac{f_X(Z)f_X(\tilde{Z})}{(f_X(Z)+f_X(\tilde{Z}))^2}\|Z_t-\tilde{Z}_t\|^2 dZ \\
&= \frac{1}{2T}\int \sum_{t=1}^{T} \frac{f_X(Z)f_X(\tilde{Z})}{f_X(Z)+f_X(\tilde{Z})}\|Z_t-\tilde{Z}_t\|^2 \ dZ \\
&= \frac{1}{2T}\mathbb{E}_X \sum_{t=1}^{T} \frac{f_X(\tilde{X})}{f_X(X)+f_X(\tilde{X})}\|Z_t-\tilde{Z}_t\|^2 \\
&= \frac{1}{2T}\mathbb{E}_X \sum_{t=1}^{T} \frac{f_X(\alpha^-(X))}{f_X(X)+f_X(\alpha^-(X))}\|X_t-\alpha_t^-(X_t)\|^2
\end{aligned}$$

which proves (8). Again, if we can choose S_ts, b_ts where $\alpha_t()$ is mirroring across planes given by $S_t x = b_t$ such that

$$f_X(X) = f_X(\alpha^{-1}(X)) \quad \forall X \in \mathbb{R}^{nT}$$

the distortion D_E becomes

$$\begin{aligned}
D_E &= \frac{1}{4T}\mathbb{E}_X \sum_{t=1}^{T} \|X_t - \alpha_t^-(X_t)\|^2 \overset{(a)}{=} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{X_t}\|S_t X_t - b_t\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \text{tr}\left(S_t R_{X_t} S_t' + (b_t - S_t\mu_{X_t})(b_t - S_t\mu_{X_t})'\right)
\end{aligned}$$

where (a) follows as $\alpha_t(.)$ is mirroring across plane given by $S_t x = b_t$, and thus $\alpha_t(x) = \alpha_t^{-1}(x) = (I - 2S_t'S_t)X_t + 2S_t'b_t$. This proves (11).

B. Proof of Theorem III.5

Since given Z, there are 2^k possibilities of X_1^T; $X_1^T = \alpha^{-1(K)}(Z), K \in [0 : 2^k - 1]$, we start by computing

$$\begin{aligned}
p_Z^{(K)} &:= \Pr(X_t = \alpha_t^{-(K)}(Z_t)|Z) = \Pr(X = \alpha^{-(K)}(Z)|Z) \\
&= \frac{1}{f_Z(Z)}\Pr(Z|X = \alpha^{-(K)}(Z))f_X(\alpha^{-(K)}(Z)) \\
&\overset{(a)}{=} \frac{f_X(\alpha^{-(K)}(Z))}{\sum_{j=0}^{2^k-1} f_X(\alpha^{-(j)}(Z))}, \ K \in [0 : 2^k - 1]
\end{aligned}$$

where (a) follows by noting that $\Pr(Z|X = \alpha^{-(K)}(Z)|Z)$ is equal to the probability of the key being equal to K, which is $1/2^k$. Let $S = \sum_{j=0}^{2^k-1} f_X(\alpha^{-(j)}(Z))$. Then, $\mathbb{E}(X_t|Z)$ equals

$$\sum_{K=0}^{2^k-1} \alpha^{-(K)}(Z_t)p_Z^{(K)} = \frac{1}{S}\sum_{K=0}^{2^k-1} \alpha^{-(K)}(Z_t)f_X(\alpha^{-(K)}(Z)).$$

We can then compute $\text{tr}(R_{X_t|Z})$ as

$$\mathbb{E}_{X_t|Z}\|X_t - \mathbb{E}(X_t|Z)\|^2 = \frac{1}{S^3}\sum_{K=0}^{2^k-1} f_X\left(\alpha^{-(K)}(Z)\right)\left\|R_t^{(K)}\right\|^2$$

where $R_t^{(K)} = \sum_{\ell=0}^{2^k-1} f_X(\alpha^{-(\ell)}(X))(\alpha_t^{-(\ell)}(X_t) - \alpha_t^{-(K)}(X_t))$. Plugging $\text{tr}(R_{X_t|Z})$ in the expression of D_E gives (14). Moreover, if condition (15) is met, (14) simplifies to (16).

C. Proof for Theorem V.3

Let the shared key K is $(K_1, K_2, \ldots, K_n)$ where all K_is are independent identically distributed and uniformly distributed in $\{0,1\}^3$. Let us also assume that $X = (X^{(1)}, X^{(2)}, \ldots, X^{(n)})$, where each $X^{(i)} \in \mathbb{R}$. Similar to the scheme for scalar case, we create a random vector $V = (V^{(1)}, \ldots, V^{(n)})$ where $V^{(i)} = (X^{(i)} - \mu^{(i)})/\sqrt{\Sigma_{ii}}$, and encode $V^{(i)}$ using key K_i as in the case of a scalar for all $i \in [n]$. Thus, the distortion D_W will be

$$\begin{aligned}
D_W &= \min_Z \text{tr}(R_{X|Z}) = \min_Z \sum_{i=1}^{n} \text{Var}(X^{(i)}|Z) \\
&= \min_Z \sum_{i=1}^{n} (\Sigma_{ii})\text{Var}(V^{(i)}|Z) = \sum_{i=1}^{n} (\Sigma_{ii}) \min_Z \text{Var}(V^{(i)}|Z) \\
&= \sum_{i=1}^{n} (\Sigma_{ii}) \min_{Z^{(i)}} \text{Var}(V^{(i)}|Z^{(i)}) = c\sum_{i=1}^{n} (\Sigma_{ii}) = c\,\text{tr}(\Sigma)
\end{aligned}$$

where $c = 0.9998$. Since $\text{tr}(\Sigma)$ is the expected distortion even when the adversary has no observations, and as we can not beat this by (6), this is optimal.

D. Proof for Theorem V.4

Distortion at the adversary's end. Based on the coding scheme, we can see that the adversary get $BU_t + w_t$ by just subtracting AZ_t from Z_{t+1} for $t \in [1 : T-1]$. So, the adversary's information is given by the following set:

$$\begin{aligned} E_{\text{info}} &= \{Z_1, BU_t + w_t,\ t \in [1 : T-1]\} \\ &= \{f(X_1, K), BU_t + w_t,\ t \in [1 : T-1]\}. \end{aligned}$$

Thus, $D(t, Z_1^T) = D(t, E_{\text{info}}) = \text{tr}(R_{X_t|E_{\text{info}}})$. Next, we can write

$$\begin{aligned} D(t+1, Z_1^T) &= \text{tr}(R_{X_{t+1}|E_{\text{info}}}) = \text{tr}(R_{(AX_t+BU_t+w_t)|E_{\text{info}}}) \\ &\overset{(a)}{=} \text{tr}(R_{A^tX_1|E_{\text{info}}}) \overset{(b)}{=} \text{tr}(R_{A^tX_1|f(X_1,K)}) \\ &= \text{tr}(A^t R_{X_1|E_{\text{info}}}(A^t)') = \text{tr}((A^t)'A^t R_{X_1|E_{\text{info}}}) \\ &\overset{(c)}{\geq} c \cdot \text{tr}\left((A^t)'A^t\Sigma\right) = c \cdot \text{tr}\left((A^t)^H A^t\Sigma\right) \\ &\overset{(d)}{=} c \cdot \text{tr}\left(V(\Lambda^H)^t\Lambda^t V^H\Sigma\right) = c \cdot \text{tr}\left(V|\Lambda|^{2t}V^H\Sigma\right) \\ &= c \cdot \text{tr}\left(|\Lambda|^{2t}V^H\Sigma V\right) \overset{(e)}{=} c \cdot \text{tr}\left(|\Lambda|^{2t}\Sigma\right) \overset{(f)}{\geq} = c \cdot \text{tr}(\Sigma) \end{aligned}$$

where (a) follows by noting that E_{info} contains $BU_t + w_t, \forall t \in [1 : T-1]$; (b) follows because U_t and w_t are independent on X_t; (c) follows because $(A^t)'A^t$ is a positive semidefinite matrix and $R_{X_1|E_{\text{info}}}$ being a diagonal matrix with the diagonal entries being elementwise greater than $c\Sigma$; (d) follows by writing the singular value decomposition of A as $A = \Phi\Lambda V^H$; (e) follows by noting that λ_i is the ith singular value of A, and V is a unitary matrix; and (f) follows for dynamic systems where $|\lambda_i| \geq 1$.

ACKNOWLEDGMENT

This article contains new results not present in the CDC paper (Sections III-B, IV, and VI), extended proofs, and additional examples.

REFERENCES

[1] G. K. Agarwal, M. Karmoose, S. Diggavi, C. Fragouli, and P. Tabuada, "Distorting an adversary's view in cyber-physical systems," in *Proc. IEEE Conf. Decis. Control*, Dec. 2018, pp. 1476–1481.

[2] J. Wan, A. B. Lopez, and M. A. A. Faruque, "Exploiting wireless channel randomness to generate keys for automotive cyber-physical system security," in *Proc. ACM/IEEE 7th Int. Conf. Cyber-Phys. Syst.*, Apr. 2016, pp. 1–10.

[3] B. Zan, M. Gruteser, and F. Hu, "Key agreement algorithms for vehicular communication networks based on reciprocity and diversity theorems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4020–4027, Oct. 2013.

[4] W. Trappe, R. Howard, and R. S. Moore, "Low-energy security: Limits and opportunities in the internet of things," *IEEE Secur. Privacy*, vol. 13, no. 1, pp. 14–21, Jan. 2015.

[5] P. Koopman and C. Szilagyi, "Integrity in embedded control networks," *IEEE Secur. Privacy*, vol. 11, no. 3, pp. 61–63, May 2013.

[6] A. B. Alexandru and G. J. Pappas, *Secure Multi-party Computation for Cloud-Based Control*. Singapore: Springer, 2020, pp. 179–207. [Online]. Available: https://doi.org/10.1007/978-981-15-0493-8_9

[7] K. Kogiso and T. Fujita, "Cyber-security enhancement of networked control systems using homomorphic encryption," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 6836–6843.

[8] F. Farokhi, I. Shames, and N. Batterham, "Secure and private cloud-based control using semi-homomorphic encryption," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 163–168, 2016.

[9] Y. Shoukry *et al.*, "Privacy-aware quadratic optimization using partially homomorphic encryption," in *Proc. IEEE 55th Conf. Decis. Control*, Dec. 2016, pp. 5053–5058.

[10] T. Fujita, K. Kogiso, K. Sawada, and S. Shin, "Security enhancements of networked control systems using RSA public-key cryptosystem," in *Proc. 10th Asian Control Conf.*, May 2015, pp. 1–6.

[11] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Techn. J.*, vol. 28, no. 4, pp. 656–715, 1949.

[12] A. Tsiamis, K. Gatsis, and G. J. Pappas, "An information matrix approach for state secrecy," in *Proc. IEEE Conf. Decis. Control*, Dec. 2018, pp. 2062–2067.

[13] A. D. Wyner, "The wire-tap channel," *Bell Syst. Techn. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[14] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 835–842, Jul. 1988.

[15] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Trans. Info. Theory*, vol. 60, no. 12, pp. 7584–7605, Dec. 2014.

[16] C. Tsai, G. K. Agarwal, C. Fragouli, and S. Diggavi, "A distortion based approach for protecting inferences," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1913–1917.

[17] A. Tsiamis, K. Gatsis, and G. J. Pappas, "State-secrecy codes for networked linear systems," *IEEE Trans. Auto. Cont.*, vol. 65, no. 5, pp. 2001–2015, 2020.

[18] A. Tsiamis, K. Gatsis, and G. J. Pappas, "State estimation with secrecy against eavesdroppers," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8385–8392, 2017.

[19] T. Tanaka, M. Skoglund, H. Sandberg, and K. H. Johansson, "Directed information and privacy loss in cloud-based control," in *Proc. Amer. Control Conf.*, May 2017, pp. 1666–1672.

[20] W. A. Malik, N. C. Martins, and A. Swami, LQ Control Under Security Constraints. *Berlin*, Germany: Springer, 2013, pp. 101–120.

[21] J. Corts, G. E. Dullerud, S. Han, J. L. Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *Proc. IEEE 55th Conf. Decis. Control*, Dec. 2016, pp. 4252–4272.

[22] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Secure estimation and zero-error secrecy capacity," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1047–1062, Mar. 2019.

[23] V. Kumar and N. Michael, "Opportunities and challenges with autonomous micro aerial vehicles," *Int. J. Robot. Res.*, vol. 31, no. 11, pp. 1279–1291, 2012.

Gaurav Kumar Agarwal received the B.Tech. degree in electronics and communication engineering from the Indian Institute of Technology Roorkee, Roorkee, India, and the M.Eng. degree in telecommunication engineering from the Indian Institute of Science, Bangalore, Bengaluru, India. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA, USA.

He has been intern with Cranfield University, Shrivenham, U.K., and Technicolor Research, Los Altos, CA, USA, in 2011 and 2016, respectively.

Mohammed Karmoose received the B.S. and M.S. degrees in electrical and electronics engineering from Alexandria University, Alexandria, Egypt, in 2009 and 2013, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA, USA.

He is currently a Senior Engineer with Samsung Semiconductor, Inc., San Diego, CA, USA. He was an intern with the Security Research Group, Intel Labs in 2017. His research interests include information theory, security and privacy in communication and cyber-physical systems, and distributed detection.

Dr. Karmoose was the recipient of the Annual Distinguished Student Award from Alexandria University for the years 2005–2009.

Suhas N. Diggavi (Fellow, IEEE) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Delhi, New Delhi, India, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1998.

After completing his Ph.D. degree, he was a Principal Member Technical Staff in the Information Sciences Center, AT&T Shannon Laboratories, Florham Park, NJ, USA. After that, he was on the faculty of the School of Computer and Communication Sciences, EPFL, where he directed the Laboratory for Information and Communication Systems. He is currently a Professor with the Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA, USA, where he directs the Information Theory and Systems laboratory. He has eight issued patents. His research interests include information theory and its applications to several areas including wireless networks, cyber-physical systems, distributed computation and learning, security and privacy, genomics, data compression.

Dr. Diggavi was an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, ACM/IEEE TRANSACTIONS ON NETWORKING, IEEE COMMUNICATION LETTERS, a Guest Editor for the IEEE SELECTED TOPICS IN SIGNAL PROCESSING, and in the program committees of several IEEE conferences. He was a Distinguished Lecturer and also currently on board of governors for the IEEE Information Theory Society. He also helped organize IEEE and ACM conferences including serving as the Technical Program Co-Chair for 2012 IEEE Information Theory Workshop, the Technical Program Co-Chair for the 2015 IEEE International Symposium on Information Theory, and the General Co-Chair for Mobihoc 2018. He has received several recognitions for his research including the 2013 IEEE Information Theory Society & Communications Society Joint Paper Award, the 2013 ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) Best Paper Award, the 2006 IEEE Donald Fink Prize Paper Award.

Christina Fragouli received the B.S. degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, Los Angeles, CA, USA.

She is currently a Professor with the Department of Electrical and Computer Engineering, University of California, Los Angeles. Her current research interests include network security and privacy, wireless networks, and machine learning under communication constraints.

Dr. Fragouli was an Information Theory Society Distinguished Lecturer and an Associate Editor for the IEEE COMMUNICATIONS LETTERS, Elsevier's *Journal on Computer Communication*, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON INFORMATION THEORY, and IEEE TRANSACTIONS ON MOBILE COMMUNICATIONS.

Paulo Tabuada was born in Lisbon, Portugal, one year after the Carnation Revolution. He received the "Licenciatura" degree in aerospace engineering from the Instituto Superior Tecnico, Lisbon, Portugal, in 1998, and the Ph.D. degree in electrical and computer engineering in 2002 from the Institute for Systems and Robotics, a private research institute associated with Instituto Superior Tecnico.

Between January 2002 and July 2003, he was a Postdoctoral Researcher with the University of Pennsylvania. After spending three years at the University of Notre Dame, as an Assistant Professor, he joined the Electrical and Computer Engineering Department, University of California, Los Angeles, Los Angeles, CA, USA, where he is currently Vijay K. Dhir Professor of engineering.

Dr. Tabuada was the Program Chair and General Chair for several conferences in the areas of control and of cyber-physical systems, such as NecSys, HSCC, and ICCPS. He is currently the Chair for HSCCs steering committee and was on the Editorial Board of the IEEE EMBEDDED SYSTEMS LETTERS and the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. His contributions to cyber-physical systems have been recognized by multiple awards including the NSF CAREER Award in 2005, the Donald P. Eckman Award in 2009, the George S. Axelby Award in 2011, the Antonio Ruberti Prize in 2015.