A Key Agreement Scheme for Cyber-Physical Systems

Walter Lucia, *Member, IEEE,* Amr Youssef, *Senior Member, IEEE*

***Abstract*—Traditional cryptographic approaches may not always be suitable for cyber-physical systems (CPSs). In this correspondence, we present a control theoretic approach allowing a networked controller and a smart actuator of a stochastic CPS to agree on a common secret key without resorting to classical cryptographic approaches. More precisely, by utilizing the asymmetry in the system model knowledge available to the control system defender and to the eavesdropper, we propose a key agreement scheme utilizing a simultaneous input and state estimation algorithm. The validity of the proposed solution is shown through a numerical example.**

***Index Terms*—Cyber-physical systems security, key agreement protocols.**

I. Introduction

In recent years, cybersecurity of cyber-physical systems (CPSs) has attracted significant attention from both the academia and industry [1]–[4]. Different detection schemes have been proposed to detect networked False Data Injection (FDI) attacks [5]. Of particular interest are the class of attacks, such as covert attacks, and zero-dynamics attacks, which are capable of remaining stealthy if the detection strategy is only implemented in the control center, see [6]. To detect such attacks, some of the existing approaches (e.g. moving target, sensor coding) require the exchange of a secret message between the plant and the controller [7]. Moreover, in any security mechanism that aims to establish any of the Confidentially, Integrity, and Availability (CIA) triad components between the plant and the networked controller, a common secret key is needed.

The establishment of a common secret key between two entities is traditionally achieved through the use of symmetric key or public key cryptographic protocols [8]. However, such cryptographic approaches might not always be suitable for CPSs, e.g., because of the lack of a practical mechanism to securely distribute/store long term keys required for symmetric key-based solutions or because of the large computational requirements associated with public key-based solutions. Furthermore, the support of a key revocation mechanism in public key infrastructure (PKI) might be hard to deploy in some CPSs applications [9]. For example, the size of the certificate revocation list (CRL) might be too large to fit in the memory of the underlying resource-constrained CPS devices and the frequency of distribution of CRLs may lead to a non-acceptable tradeoff between performance overhead (if this frequency is too high) and longer vulnerability period for less frequent updates. Outsourcing the task of checking the revoked certificates to online servers requires always-available communication with this server, which can be a bottleneck for real time operations and the security also depends on the freshness of the revocation information stored at the server. Finally, these schemes may require the operation technology network to be connected to the servers that manage the certificates, which introduces additional vulnerabilities.

On the other hand, in the classic model of cryptosystems introduced by Shannon [10], the adversary is assumed to have access to precisely the same information as the legitimate receiver. As argued in [11], in some situations, this assumption is rather restrictive. For example, in the case of CPSs, one can utilize the asymmetry in the system model knowledge available to control system designer/operator (the defender) and to the eavesdropper. This can be justified by noting that the defender has a larger degree of freedom in designing/executing a system identification procedure. On the other hand, the adversary cannot design the input signals and can perform system identification only using the data observed during the online closed-loop operations. We show that this asymmetry can be exploited to play a role similar to the one performed by the noise in Wyner's wiretap channel [12] and present a control theoretic approach that allows the networked controller and the smart actuator to agree on a common secret key.

Walter Lucia and Amr Youssef are with the Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, H3G-1M8, Canada. (e-mail: walter.lucia@concordia.ca; youssef@ciise.concordia.ca).

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

A. Background and related works.

Wyner [12] introduced the wiretap channel model, which provides an approach to secure communications without relying on classical cryptographic approaches. Wyner showed that if the channel between the sender and intended legitimate receiver is statistically better, i.e., less noisy, than the channel from the sender to the adversary, then perfect secrecy is possible. Maurer [11], and Ahlswede and Csiszár [13] considered the problem of secret-key generation using public discussion. In this setting, the two communicating parties observe correlated versions of a common random source. Based on these observations both parties want to agree on the same secret key.

The key agreement problem in CPSs has traditionally been considered to be an issue to be addressed separately from the control-theoretic/physical aspects of the underlying control systems. For example, Lara-Nino et al. [14] presented three cryptographic key-establishment protocols for resource constrained CPSs, including one constructed using isogeny-based cryptography, which makes it well-suited for the post-quantum computing scenario. Sutrala et al. [15] proposed a three-factor user authenticated key agreement protocol which achieves user anonymity and untraceability for 5G-enabled softwarized industrial CPSs. Zhang et al. [16] proposed a cross-layer key establishment model for wireless devices in CPSs, where wireless devices extract master keys at the physical layer using ambient wireless signals. Following an information-theoretic approach, Rawat et al. [17] studied the outage probability for secrecy rate in multiple-input multiple-output (MIMO) systems for CPSs.

Although the above reviewed key-agreement solutions target CPSs' applications, such approaches do not explicitly take any advantage of the peculiar underlying physical dynamics of these systems. In recent years, there has been an increased interest towards the use of the physical properties of the underling control systems to provide security. Similar to the role of information theory in physical layer security [18], control theory provides a framework for the study of this issue in CPSs. Li et al. [19] presented a key establishment scheme for networked control systems, where they exploited the common information of the physical system state for the key establishment between the sensor and the controller. Since the sensors can observe changes in the system state which are caused by the controller actions, hence, the controller and the sensor can utilize this implicit channel to exchange messages to find the common random bits in the predicted and observed system states, respectively. Finally, the secret key is generated from the common bits. Through out the analysis, the authors assume that the eavesdropper is able to observe the system state $x(t)$ via $z(t) = Ex(t) + n(t)$, where E is the observation matrix for the eavesdropper and $n(t)$ denotes the noise. However, they focused on the special case in which $E = 0$, i.e., the eavesdropper cannot observe the system state. In [20], the authors considered a CPS where the networked controller wants to send a secrete message to the plant without resorting to cryptographic solutions. The presented robust control-theoretic approach exploits a

Wyner wiretap-like encoding scheme taking advantage of the closed-loop operations typical of feedback control systems, specifically, by resorting to the control concept of one-step reachable sets, which is applicable only to systems with bounded noise (also see [21], [22]).

B. Contribution

Our contributions can be summarized as follows: (i) we show that a key-agreement scheme in CPSs can be obtained by leveraging the asymmetry in the plant model knowledge available to the defender (controller) and the attacker (eavesdropper), (ii) we leverage an Unknown Input Observer (UIO) [23], [24], as the main decoding mechanism, and (iii) we utilize Error Correcting Code (ECC) to improve its robustness against the process and measurement noise realizations. Differently from [19], which focuses on the case where the attacker does not have access to a state measurement signal, our proposed key-agreement scheme is effective in the presence of attackers with perfect access to the state measurement signal. Furthermore, the solution in [20] proposes a key-agreement scheme for systems subject to bounded disturbances. In contrast, our proposed approach is capable of dealing with a more general class of stochastic CPSs. The robust decoding mechanism in [20], which exploits one-step robust reachability arguments, is not suitable for stochastic systems. Indeed, while in [20] the decoding mechanism has the property of ensuring, by design, the absence of decoding errors, such a property does not hold in stochastic systems where the presence of, possibly unbounded, disturbances/noise can lead to unpredictable decoding errors. Therefore, we address such a problem with an entirely different approach that leverages an unknown input observer, an error-correcting code and a key-distillation procedure.

C. Notation and Paper's Organization

In the sequel, the sets of real numbers and real-valued column vectors of dimension $n_v > 0$ are denoted with $\mathbb{R}$ and $\mathbb{R}^{n_v}$, respectively. The sets of two dimensions real-valued matrices M of size $n_1 \times n_2 > 0$, $n_1, n_2 > 0$, and the non-negative integer numbers are denoted with $\mathbb{R}^{n_1 \times n_2}$ and $\mathbb{Z}_+ := \{0, 1, ...\}$, respectively. Given a vector $v \in \mathbb{R}^{n_v}$, v_k denotes the values of v at the discrete sampling time instant $k \in \mathbb{Z}_+$, while $v[k]$ selects the $k-th$ entry of v. The transpose of a matrix M and its inverse (for a squared matrix) are denoted with M^T, and M^{-1}, respectively. Given a random variable $v \in \mathbb{R}^{n_v}$, $v \sim \mathcal{N}(\mu_v, \Sigma_v)$ indicates that v is a random variable normally distributed with mean $\mu_v \in \mathbb{R}$ and covariance matrix $\Sigma_v > 0 \in \mathbb{R}^{n_v \times n_v}$. Given an event E, the probability of realization of such an event is denoted as $P(E)$. Table I shows a summary of the symbols used throughout the paper. The rest of the paper is organized as follows. The system setup and problem formulation are presented in section II. Then the description of our proposed protocol is provided in section III. A proof of concept implementation and simulation results are presented in section IV. Finally, the paper is concluded in section V.

TABLE I: List of symbols.

Symbol	Description
$\mathcal{N}(\mu_v, \Sigma_v)$	Normal distribution with mean μ_v and covariance Σ_v
A, B, C	System matrices (state-space representation)
A_e, B_e, C_e	System matrices estimated by the attacker
$x_k, \hat{x}_k, \hat{x}^e_k$	State vector (x_k), estimated state ($\hat{x}_k$), attacker's estimated state ($\hat{x}^e_k$)
$u_k, \hat{u}_k, \hat{u}^e_k$	Control input vector (u_k), estimated control input (u_k), attacker's estimated input (u^e_k)
y_k	Sensor measurement vector
w_k, v_k	Process (w_k) and measurement (v_k) noises
w^e_k, v^e_k	Process (w^e_k) and measurement (v^e_k) noises estimated by the attacker
$\mathcal{W}, \mathcal{V}$	Process and measurement noise covariance matrices
$\mathcal{W}_e, \mathcal{V}_e$	Process and measurement noise covariance matrices estimated by the attacker
$\mathcal{M}$, $\mathcal{M}_e$	Model knowledge of the defender and attacker
UIO	Unknown Input Observer
ECC	Error Correcting Code
d_H	Hamming distance
n_c, k_c, d_c	ECC parameters: length of input message (k_c), length of codeword (n_c), minimum Hamming distance (d_c) between two codewords.
$\mathcal{K}_c, \mathcal{K}_{sa}, \mathcal{K}_e$	Binary keys identified by the controller ($\mathcal{K}_c$), smart actuator ($\mathcal{K}_{sa}$), and eavesdropper ($\mathcal{K}_e$)
δ	Parameter used by the controller to impose a state shift in the computation of the admissible control inputs
η	Parameter used by the attacker to increase the distance between the two admissible control inputs
α	Max percentage difference between the non-zero entries of A, B, C and A_e, B_e, C_e.

II. System Setup and Problem Formulation

A. Networked control system setup

Consider the following Linear Time Invariant stochastic system

$$x_{k+1} = Ax_k + Bu_k + w_k, \;\; y_k = Cx_k + v_k \tag{1}$$

where $x_k \in \mathbb{R}^{n_x}$, $u_k \in \mathbb{R}^{n_u}$, and $y_k \in \mathbb{R}^{n_y}$, are the plant's state, control input and measurement vectors, respectively, and A, B, C denote the system matrices of appropriate dimensions. Furthermore, the variables $w_k \sim \mathcal{N}(0, \mathcal{W})$ and $v_k \sim \mathcal{N}(0, \mathcal{V})$ are independent and identically distributed (i.i.d), uncorrelated with each other and with the plant's initial state. By assuming a standard control setup, the networked controller is a dynamic compensator consisting of a state estimator (e.g,. Kalman filter) providing an unbiased estimation of x_k, namely $\hat{x}_k$, and a stabilizing state-feedback controller, whose actions are generically described as

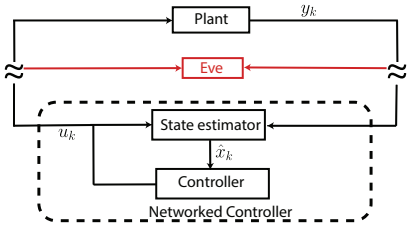

Fig. 1: Networked control system setup.

$$u_k = f(\hat{x}_k), \quad f(\cdot) : \mathbb{R}^{n_x} \to \mathbb{R}^{n_m} \tag{2}$$

B. Adversary model

We assume that an attacker (Eve) can eavesdrop the signals transmitted in the communication channels between the plant and the controller (see Fig. 1). Moreover, Eve has the following knowledge of the system dynamical model

$$x_{k+1} = A_e x_k + B_e u_k + w^e_k, \;\; y_k = C_e x_k + v^e_k \tag{3}$$

where A_e, B_e, C_e are matrices of appropriate dimensions and $w^e_k \sim \mathcal{N}(0, \mathcal{W}_e)$ and $v^e_k \sim \mathcal{N}(0, \mathcal{V}_e)$ are i.i.d, uncorrelated with each other and with the plant's initial state.

Assumption 1: The defender possesses the system model (1), while Eve has a non perfect approximation of (1). In particular, by defining the sets $\mathcal{M} = \{A, B, C, \mathcal{W}, \mathcal{V}\}$, $\mathcal{M}_e = \{A_e, B_e, C_e, \mathcal{W}_e, \mathcal{V}_e\}$ collecting the plant's model knowledge for the defender and attacker, we assume that

$$\mathcal{M} \neq \mathcal{M}_e. \tag{4}$$

The asymmetry in the system model (4) can be justified by noting that the defender can offline design a proper system identification phase by adequately selecting sufficiently exciting input signals [25]. However, the adversary can only identify the system dynamics by using the data accessible by observing the online closed-loop operations.

C. Problem formulation

Given the system model (1)-(2) and attacker model (3)-(4), design a key-agreement scheme between a plant's actuator (hereafter referred to as the smart actuator or just actuator) and the controller such that:

- The key-agreement is achieved by leveraging the asymmetry (4) in the system model knowledge (i.e., no cryptography schemes are explicitly used);

- Let $\mathcal{K}_c \in \{0,1\}^n$, $\mathcal{K}_{sa} \in \{0,1\}^n$ and $\mathcal{K}_e \in \{0,1\}^n$ be the binary keys of length $n > 0$ identified by the controller, smart actuator and Eve, respectively. Then, $P(\mathcal{K}_c = \mathcal{K}_{sa}) \approx 1$ and $P(\mathcal{K}_c \neq \mathcal{K}_e) \approx 1$.

In this correspondence, a solution to the above problem is given under the assumption that an unknown input observer for (1) can be defined to simultaneously estimate the state x_k (namely $\hat{x_k}$) and the input signal u_k (namely $\hat{u}_k$) from the sensor measurement y_k. In what follows, the UIO algorithm is abstractly described by means of the following recursive UIO function:

$$[\hat{u}_{k-1}, \hat{x}_k] = UIO(\hat{u}_{k-2}, \hat{x}_{k-1}, y_k, \mathcal{M}) \tag{5}$$

where the pairs $(\hat{u}_{k-1}, \hat{x}_k)$ and $(\hat{u}_{k-2}, \hat{x}_{k-1})$ define the available estimation at the time k and $k-1$, respectively.

III. Protocol Design

To motivate for our key agreement protocol, we start by assuming that the system's model (1) is shared between the actuator and the controller. In this case, a key agreement protocol can be obtained as follows:

- In the networked controller, two different stabilizing control laws $u_k^0 = f_0(\hat{x}_k)$ and $u_k^1 = f_1(\hat{x}_k)$ are used to compute two admissible control inputs. The controller sends (u_k^0, u_k^1) to the smart actuator, which will be in charge of deciding the control action u_k.

- Given (u_k^0, u_k^1), the smart actuator generates a random bit $b_k \in \{0,1\}$ and uses it to pick the control action u_k to apply to the plant, i.e.

$$u_k = \begin{cases} u_k^0 & \text{if } b_k = 0 \\ u_k^1 & \text{otherwise} \end{cases} \tag{6}$$

- Since the controller is not aware of u_k, an UIO is used to obtain the pair $(\hat{x}_k, \hat{u}_{k-1})$ and estimate the bit b_{k-1} encoded into u_{k-1} as:

$$\hat{b}_{k-1} = \begin{cases} 0 & \text{if } d_0 < d_1 \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

where $d_0 = \|\hat{u}_{k-1} - u_{k-1}^0\|_2$, $d_1 = \|\hat{u}_k - u_{k-1}^1\|_2$. At each iteration, the controller appends $\hat{b}_{k-1}$ into the local key $\mathcal{K}_c$.

- Since the UIO-based decoding scheme (7) is not robust against process and measurement noise realizations, it is possible that $\hat{b}_{k-1} \neq b_{k-1}$. Such a drawback can be solved by implementing a copy of the UIO estimation (5) local to the smart actuator. By doing so, the smart actuator can determine the bit $\hat{b}_{k-1}$ estimated by the controller and hence appends $\hat{b}_{k-1}$ (instead of b_{k-1}) in its local key $\mathcal{K}_{sa}$.

Eve can eavesdrop $(u_{k-1}^0, u_{k-1}^1, y_k)$ and run an UIO algorithm to decode the transmitted key. However, given the asymmetry in the system model between the attacker and the defender (Assumption 1), Eve cannot implement the exact UIO used by the defender. In particular, it can use the same UIO algorithm but with a different system model $\mathcal{M}_e \neq \mathcal{M}$, i.e.

$$[\hat{u}_k^e, \hat{x}_{k+1}^e] = UIO(\hat{x}_k, u_k, y_{k+1}, \mathcal{M}_e) \tag{8}$$

Therefore, the adversary reconstructs the unknown input u_k with a covariance of the error, namely $cov(u_k - \hat{u}_{k-1}^e)$, larger than the one of the defender (controller and smart actuator). As a consequence, for a proper choice of the switching control law $f_0(\cdot)$ and $f_1(\cdot)$, and for a sufficiently long key $\mathcal{K}_c$, the probability $P(\mathcal{K}_c \neq \mathcal{K}_e) \approx 1$.

The protocol described above guarantees, by constructions, that the key decoded by the controller is without bit errors, i.e. $P(\mathcal{K}_c = \mathcal{K}_{sa}) = 1$. However, one can argue that the information inherently represented by the model shared between the actuator and controller can be looked at as an initial secret key and hence this protocol is better described as a key stretching scheme, rather than a key establishment scheme. In what follows, we show that such pre-shared model assumption is not required and that a key agreement protocol can be built by exploiting only the fact that the attacker's equivocation (uncertainty) about the key bit encoded in the control signal (6) is greater of the one available to the defender, i.e. the implicit communication channel between the plant and the controller enjoys the properties of a Wyner's channel [12].

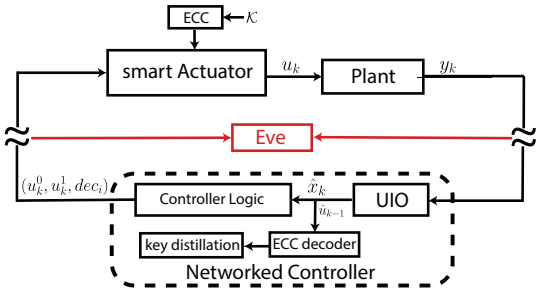

Fig. 2: System setup for the proposed key agreement protocol.

A. Key-agreement protocol without a shared model

Similar to the above protocol, at each time step k, we assume that the controller sends two possible control actions (u_k^0 and u_k^1) and that the decision of which input apply is left to the smart actuator (see Fig. 2). As a consequence, the controller still needs to implement the UIO (5) to be able to estimate the applied input u_k. However, differently from the previous protocol, now the smart actuator cannot implement a copy of (5) to correct possible decoding errors. To overcome such a drawback and improve the protocol robustness, we assume that the controller and actuator publicly agree (i.e., known to Eve) on the use of a specific error correcting scheme (ECC). For the purpose of this work, we assume the use of a linear ECC. A linear ECC with parameters $[n_c, k_c, d_c]$ defines a linear transformation of a binary string $s \in \{0,1\}^{k_c}$ into a subspace $\mathcal{C} \in \{0,1\}^{n_c}$ of cardinality 2^{k_c} such that for any pair of words $c_1, c_2 \in \mathcal{C}$, $c_1 \neq c_2$, the Hamming distance $d_H(c_1, c_2) < d_c$ and the maximum number of correctable error is $\frac{d_c - 1}{2}$. The proposed key-agreement protocol can be summarized as shown in Algorithm 1. The complexity of the steps run at the smart actuator is dominated by the ECC encoding process, which, when using redundancy code, can be performed in a negligible time (independent of the underlying system state space complexity). On the other hand, the complexity of the steps run by the actuator is dominated by the UIO. Given the dimensions of the matrices used in the UIO algorithm, and by noting that typically $n_x \geq n_y \geq n_u$, the time complexity of the UIO (Algorithm 1) is given by $O(n_x^3)$.

B. Key distillation

Although the adversary also receives the codewords encoded by the actuator (by running the UIO algorithm with the model available to it), the adversary conceptual channel is nevertheless going to be

Algorithm 1 Key agreement protocol: Smart Actuator and Controller

—Smart Actuator—

1: A random binary string $\mathcal{K}$ is generated at $k = 0$;
2: The string $\mathcal{K}$ is divided into substrings $\{s_i\}, s_i \in \{0,1\}^{k_c}$;
3: According to the utilized ECC scheme, each s_i is sequentially encoded into a codeword $\{c_i\}$, $c_i \in \{0,1\}^{n_c}$, $c_i \in \mathcal{C}$;
4: Each bit $c_i[j]$, $j = 1, \ldots, n_c$, is sequentially used to define u_k,

$$u_k = \begin{cases} u_k^0 & \text{if } c_i[j] = 0 \\ u_k^1 & \text{otherwise} \end{cases} \tag{9}$$

5: Every time each substring s_i is sent, the actuator expects to publicly receive, through the actuation channel, an extra bit, namely $dec_i \in \{0,1\}$, announcing if the decoding operations on the controller's side were successful (see the below Controller's lines 5-7);
if $dec_i == 1$ **then** s_i is appended in the key $\mathcal{K}_{sa}$; **else** s_i is discarded.

—Controller—

1: At each time instant, the pair $(\hat{u}_{k-1}, \hat{x}_k)$ is estimated: $[\hat{u}_{k-1}, \hat{x}_k] = UIO(\hat{u}_{k-2}, \hat{x}_{k-1}, y_k, \mathcal{M})$;
2: The distances d_0 and d_1 are computed: $d_0 = \|\hat{u}_{k-1} - u_{k-1}^0\|_2$, $d_1 = \|\hat{u}_{k-1} - u_{k-1}^1\|_2$;
3: The UIO-based decoding rule (7) is used to obtain $\hat{b}_k$;
4: $\hat{b}_k$ is appended in the estimated codeword $\hat{c}_i$;
5: Every time a string $\hat{c}_i$ of size n_c is received, the Hamming distance is evaluated

$$d_{\hat{c}_i} = \min_{c \in \mathcal{C}} d_H(\hat{c}_i, c) \tag{10}$$

if $d_{\hat{c}_i} \ll \frac{d_c - 1}{2}$, **then**
- the codeword $\hat{c}_i$ is accepted (i.e., a very reliable decision about the codeword chosen by the actuator can be made); $\hat{s}_i$ decoded and appended in the key $\mathcal{K}_c$; dec_i is set to 1;

else the codeword $\hat{c}_i$ is discarded; dec_i is set to 0;
6: Compute the two admissible control inputs: $u_k^0 = f_0(\hat{x}_k)$, $u_k^1 = f_1(\hat{x}_k)$;
7: Publicly send (u_k^0, u_k^1, dec_i) over the actuation channel.

worse for appropriate choices of a code C and for an appropriate reliability decision than the controller's channel, if one averages only over those instances accepted by the controller. Also, the switching control law (6) and the ECC are two design parameters that can be tuned to obtain a very small bit error rate in order to achieve $P(\mathcal{K}_c = \mathcal{K}_{sa}) \approx 1$. Further assurance that $\mathcal{K}_c = \mathcal{K}_{sa}$ can be achieved by exchanging the hash value of these two variables over the public actuation/measurement channels, and verifying that they match. Assuming that the length of the utilized hash function, H, is l bits, then, for $\mathcal{K}_c \neq \mathcal{K}_{sa}$, we have $P(H(\mathcal{K}_c) = H(\mathcal{K}_{sa})) \approx \frac{1}{2^l}$, which is negligible for practical hash functions with $l \geq 128$ bits [8].

To eliminate the knowledge that Eve has gained, the controller and actuator proceeds to the privacy amplification stage [26]. To exploit what Eve does not know about the key, the controller and plant can calculate a function of their key elements so as to spread Eve's partial ignorance over the entire result. More precisely, privacy amplification is a process which allows two parties to distill a secret key from a common random string, which an eavesdropper has partial information about it. In this work, we utilize universal hashing [27], which is a simple technique for achieving privacy amplification [26]. Let n denote the length of $\mathcal{K}_{sa}$, and $\mathcal{K}_c$, and let a be an element of $GF(2^n)$, which denotes Galois Field of 2^n elements [28]. Interpret z as an element of $GF(2^n)$. Consider the function $\{0,l\}^n \to \{0,1\}^r$ which assigns to an argument z the first r bits of the element az of $GF(2^n)$. The class of all such functions for $a \in GF(2^n)$ is a universal class of functions for $1 \leq r \leq n$ [29]. Thus, for our protocol, both the controller and the actuator can agree on the final key, $\mathcal{K}_s$ of length r by applying this universal hash function with input $z = \mathcal{K}_c$, and $z = \mathcal{K}_{sa}$, respectively.

IV. Proof-of-Concept and Simulation Results

A. Proposed implementation

Switching control logic: we assume that the plant (1) is regulated by the linear controller

$$u_k = -K\hat{x}_k \tag{11}$$

with $K \in \mathbb{R}^{n_m \times n_x}$ the controller gain. Therefore, we proposing obtaining (u_k^0, u_k^1) by introducing a small state shift $\delta > 0$ into (11), i.e.,

$$u_k^0 = -K(\hat{x}_k - \delta \mathbf{1}_{n_x}),\ u_k^1 = -K(\hat{x}_k + \delta \mathbf{1}_{n_x}) \tag{12}$$

with $\mathbf{1}_{n_x}$ denoting the all-ones column vector of size n_x.
UIO algorithm: Without loss of generality and under standard UIO conditions (see Assumption2), we proposed using *Algorithm* 2 (adapted from [23]) to obtain the unbiased estimations $\hat{u}_k$ and $\hat{x}_k$.

Assumption 2: The system matrices (A, B, C) satisfy the following conditions [23]:

- $rank(C) = n_y$, $rank(B) = n_u$, $n_u \leq n_y$, $rank(CB) = n_u$,
- $C(zI - A)^{-1}B$ is left invertible and strictly minimum phase, i.e.

$$rank\left(\begin{bmatrix} zI - A & -zB \\ C & 0 \end{bmatrix}\right) = n_x + m_u, \forall z \in \mathbb{C}, |z| \geq 1 \tag{13}$$

Algorithm 2 UIO: Unknown Input and State Estimation [23]

Input: The measurement vector y_{k+1}.
Output: The estimated state and input vectors $\hat{x}_{k+1}$ and $\hat{u}_k$.

- (State and input covariance matrices estimation) -

1: $\bar{P}_{k/k} = AP_{k/k}^x A^T + \mathcal{W}$
2: $P_{k/k+1}^u = \left(B^T C^T \left(\mathcal{V} + C\bar{P}_{k/k}C^T\right)^{-1} CB\right)^{-1}$
3: $P_{k+1/k+1}^{xu} = P_{k/k}^x \bar{P}_{k/k}^{-1} B \left(B^T \bar{P}_{k/k}^{-1} B\right)^{-1}$
4: $P_{k+1/k+1}^{ux} = P_{k/k+1}^u B^T \bar{P}_{k/k}^{-1} \left(\bar{P}_{k/k}^{-1} + C^T \mathcal{V}^{-1} C\right)^{-1}$
5: $\mathcal{P}_{k+1/k+1}^x = \left(\bar{P}_{k/k}^{-1} + C^T \mathcal{V}^{-1} C\right)^{-1} + P_{k+1/k+1}^{xu} (P_{k/k+1}^u)^{-1} P_{k+1/k+1}^{ux}$

- (State and input correction gains computation) -

6: $L_{k+1}^x = \left(\bar{P}_{k/k}^{-1} + C^T \mathcal{V} C\right)^{-1} C^T \mathcal{V}^{-1}$
7: $L_{k+1}^u = P_{k+1/k+1}^{ux} C^T \mathcal{V}^{-1}$

- (Unknown input estimation) -

8: $\hat{u}_k = L_{k+1}^u (y_{k+1} - CA\hat{x}_k)$

- (State estimation) -

9: $\hat{x}_{k+1} = A\hat{x}_k + B\hat{u}_k + L_{k+1}^x (y_{k+1} - C(A\hat{x}_k + B\hat{u}_k))$

ECC scheme: For simplicity we use the simplest form of ECC known as repetition code (see [30] for more efficient ECC schemes and [31] for the use of neural network in the process of error reconciliation). We assume that the string $\mathcal{K}$ is divided into strings s_i containing a single bit, i.e. $k_c = 1$. Moreover, s_i is encoded into a codeword $c_i = [s_i, \ldots, s_i]$ containing $n_c > 1$ repetitions of s_i. Furthermore, we assume that the received codeword $\hat{c}_i$ is accepted by the controller if and only if $d_{\hat{c}_i} = 0$, i.e., all the bits are either 0 or 1.

B. Numerical results for a quadruple-tank water system

The proposed key agreement scheme is based on a control-theoretical approach. Hence, it can be used as long as the mathematical model of the CPS can be abstracted as a stochastic linear time-invariant system. For our smulation example, we consider the quadruple-tank water benchmark system described in [32] is used as a testbed. The system consists of four tanks where the water levels h_i, $i = 1, \dots, 4$ are the state components of the systems and the two pumps' valves v_1 and v_2 are the control inputs. Two sensors are available to measure the water's levels in the first two tanks. The nonlinear system dynamics have been discretized using a sampling time $T_s = 0.1\,\text{sec}$ and linearized around the operating equilibrium point. The resulting system matrices are

$$A = \begin{bmatrix} 0.9984 & 0 & 0.0042 & 0 \\ 0 & 0.9989 & 0 & 0.0033 \\ 0 & 0 & 0.9958 & 0 \\ 0 & 0 & 0 & 0.9967 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.0083 & 9.9969 \times 10^{-6} \\ 5.1966 \times 10^{-6} & 0.0063 \\ 0 & 0.0048 \\ 0.0031 & 0 \end{bmatrix}, C = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{bmatrix}.$$

Moreover, the covariance matrices for the process and measurement noises are $\mathcal{W} = 10^{-6} I_4$ and $\mathcal{V} = 10^{-4} I_2$, respectively. The system is regulated by a state-feedback Linear Quadratic (LQ) controller $u_k = -K\hat{x}_k$, with

$$K = \begin{bmatrix} 0.0124 & 0.0046 & 0.0087 & 0.0068 \\ 0.0048 & 0.0133 & 0.0075 & 0.0096 \end{bmatrix} \tag{14}$$

The asymmetry in the knowledge of the adversary ($\mathcal{M} \neq \mathcal{M}_e$) is modeled by perturbing by a percentage $\pm\alpha$ the non-zero elements of the system matrices A, B, C. The zero elements are not perturbed since they present structural properties of the water tank system, which are assumed to be known to the attacker. We utilize a repetition code with $n_c = 3$. Moreover, the adversary's optimal decoding strategy consists in intercepting the controller decoding announcements dec_i and consider only the codewords accepted by the defender and uses a majority decoding rule.

The performance of the proposed key-agreement protocol (see Section III-A) have been investigated for ten equally spaced values of $\delta \in [0.003, 0.09]$ in (12) and six equally spaced values of $\alpha \in \pm[0, 4]\%$. The results are obtained by averaging over $50,000$ randomly generated bits for each choice of α and δ.

Figure 3 shows the percentage of bits accepted by the controller and the corresponding percentage of bit correctly decoded for $\delta \in [0.003, 0.09]$. From the obtained numerical results, for $\delta > 0.0127$, the percentage of bit correctly decoded bits is equal to 100%. Also, the number of accepted bits shows a monotonically increasing behaviour, implying that the capacity of the key exchange protocol improves with the magnitude of the state shift δ. On the other hand, it should be noted, as depicted in the figure, that increasing δ leads to an increase in the control cost $J_x = \frac{1}{N_s}\sum_{k=1}^{N_s} \|x(k)\|_2^2$, where N_s is the number of steps for which the index is evaluated.

Figure 4 depicts how the percentage of bit disagreement between the key decoded by the controller and Eve varies with α and δ. It is clear that the adversary conceptual channel becomes worse as the model uncertainty, i.e. α, increases. Moreover, with a relatively small model discrepancy equal to $\pm 4\%$, roughly 30% of the adversary key will be wrong. In this case, running a privacy amplification procedure using the universal hash function described above, with $r < 0.30n$, eliminates this partial knowledge of Eve about the final agreed upon key.

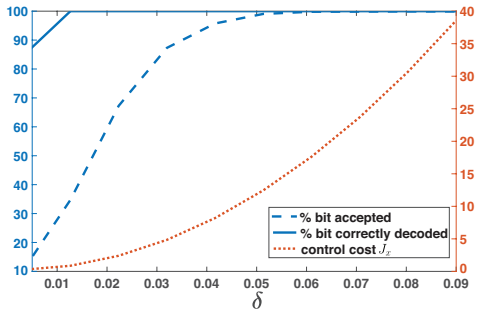

Fig. 3: Percentage of bits accepted by the controller (blue dashed line), % of bits correctly decoded by the controller (blue solid line), and control cost J_x (dotted red line), for $\delta \in [0.003, 0.09]$.

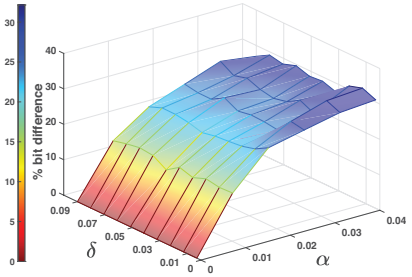

Fig. 4: Bits difference (% disagreement) between the key decoded by the controller and Eve for $\alpha \in \pm[0, 4]\%$ and $\delta \in [0.003, 0.09]$.

C. Resistance to active attackers

This section shows how our proposed key-agreement protocol can be enhanced to provide resistance against active man-in-the-middle attackers. We do not consider Denial of Service (DoS) attacks that disrupt the key agreement process, simply because all key agreement protocols are trivially susceptible to this sort of attacks (e.g., by an attacker who drops the communication packets). Instead, we consider a stealthy active attacker who tries to modify the data transmitted between the controller and the smart actuator to deceive them into agreeing on a key that the adversary can successfully decode. In particular, the following attack strategy is investigated: if the attacker has access to the signals u_k^0 and u_k^1 transmitted by the controller, then the attacker can maximize the separation of these two signals to improve his/her chance of correctly decoding the signal applied by the smart actuator (i.e., to compensate for the not accurate model knowledge exploited by the attacker's UIO). More precisely, according to the switching control logic (12), the norm-2 distance between u_k^0 and u_k^1 is equal, by design, to $\Delta = \|2K\delta \mathbf{1}_{n_x}\|_2$. Then, the attacker can increase the distance by a quantity $\|2K\eta \mathbf{1}_{n_x}\|_2$, $\eta > 0$, by replacing the pair (u_k^0, u_k^1), with $(u_k^{0'}, u_k^{1'})$, where

$$u_k^{0'} = u_k^{0'} + K\eta \mathbf{1}_{n_x}, \quad u_k^{0'} = u_k^{1'} - K\eta \mathbf{1}_{n_x}, \quad \eta > 0 \tag{15}$$

Increasing η has two consequences. On one hand, the attacker increases its capability of correctly decoding each bit. On the other hand, the number of bits correctly decoded and accepted by the defender also increases (see Fig. 3). In principle, for a sufficiently large η, the attacker could reduce the bit difference shown in Fig. 4. Nevertheless, such an attack scenario can be easily detected by the defender. In particular, given $\hat{u}_k$ (estimated by the UIO), the defender can compute $d_0^k = \|\hat{u}_{k-1} - u_{k-1}^0\|_2$, $d_1^k = \|\hat{u}_k - u_{k-1}^1\|_2$. In the absence of the attack, $\hat{u}_{k-1}$, estimated by the defender's UIO, has an expected value that is centered around u_{k-1}^0 or u_{k-1}^1. Therefore, the defender can check at each time instant k if the following anomaly

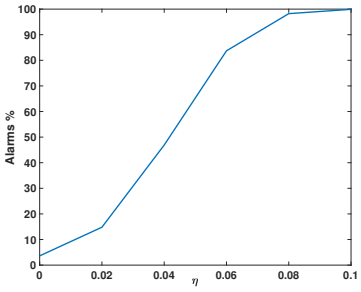

Fig. 5: Percentage of alarms raised for $\eta \in [0,\ 0.1]$ during the key bit transmission time interval.

condition (which simply verifies with a given tolerance taking into account the presence of noise, if $\hat{u}_{k-1}$, is within its expected range):

$$anomaly(k) = \begin{cases} yes & \text{If } \min(d_0^k, d_1^k) \gg \Delta/2 \\ no & \text{Otherwise} \end{cases} \tag{16}$$

If during the key establishment interval the frequency of alarms is larger than the designed false alarm rate (number of alarms that can be tolerated due to the presence of process and measurement noise), an attack is detected, and the key exchange is invalidated. From Fig. 5, it is clear that the percentage of alarms raised with the presence of $\eta \in [0,\ 0.1]$ increases rapidly, which vanishes any possibility for the attacker to reduce the key-bit difference (between what the attacker guesses and the agreed-upon key) while remaining undetectable.

V. Conclusions

We confirmed the possibility of designing a control theoretic key agreement scheme for stochastic CPSs, by leveraging the discrepancy in the model knowledge between the system's defender and the attacker. An unknown input observer is utilized to allow the controller to decode, from the sensor measurement, the bits encoded by the smart actuator in the control signal. The reliability of the encoding process is improved by leveraging an ECC scheme in which the controller accepts the bits chosen by the actuator if and only if the controller can make a very reliable decision about the codeword chosen by the actuator. The partial knowledge acquired by Eve is eliminated by the use of a standard privacy amplification technique. For future research, the effectiveness of the developed scheme can be further tested by implementing it on real-world testbeds.

References

[1] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[2] C.-S. Cho, W.-H. Chung, and S.-Y. Kuo, "Cyberphysical security and dependability analysis of digital control systems in nuclear power plants," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 3, pp. 356–369, 2015.

[3] C. Peng, H. Sun, M. Yang, and Y.-L. Wang, "A survey on security communication and control for smart grids under malicious cyber attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 8, pp. 1554–1569, 2019.

[4] D. Ding, Q.-L. Han, X. Ge, and J. Wang, "Secure state estimation and control of cyber-physical systems: A survey," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020.

[5] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annual Reviews in Control*, 2019.

[6] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 82–92, 2015.

[7] M. Ghaderi, K. Gheitasi, and W. Lucia, "A blended active detection strategy for false data injection attacks in cyber-physical systems," *IEEE Trans. on Control of Network Systems*, vol. 8, no. 1, pp. 168–176, 2020.

[8] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 2018.

[9] M. M. Mahmoud, J. Mišić, K. Akkaya, and X. Shen, "Investigating public-key certificate revocation in smart grid," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 490–503, 2015.

[10] C. E. Shannon, "Communication theory of secrecy systems," *The Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[11] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[12] A. D. Wyner, "The wire-tap channel," *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[13] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[14] C. A. Lara-Nino, A. Diaz-Perez, and M. Morales-Sandoval, "Key-establishment protocols for constrained cyber–physical systems," *Security in Cyber-Physical Systems: Studies in Systems, Decision and Control*, pp. 39–65, 2021.

[15] A. K. Sutrala, M. S. Obaidat, S. Saha, A. K. Das, M. Alazab, and Y. Park, "Authenticated key agreement scheme with user anonymity and untraceability for 5g-enabled softwarized industrial cyber-physical systems," *IEEE Trans. on Intelligent Transportation Systems*, 2021.

[16] Y. Zhang, Y. Xiang, and X. Huang, "A cross-layer key establishment model for wireless devices in cyber-physical systems," in *Proc. of 3rd ACM Workshop on Cyber-Physical System Security*, 2017, pp. 43–53.

[17] D. B. Rawat, T. White, M. S. Parwez, C. Bajracharya, and M. Song, "Evaluating secrecy outage of physical layer security in large-scale mimo wireless communications for cyber-physical systems," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1987–1993, 2017.

[18] H. V. Poor and R. F. Schaefer, "Wireless physical layer security," *Proc. of the National Academy of Sciences*, vol. 114, no. 1, pp. 19–26, 2017.

[19] H. Li, L. Lai, S. Djouadi, and X. Ma, "Key establishment via common state information in networked control systems," in *Proceedings of the 2011 American Control Conference*. IEEE, 2011, pp. 2234–2239.

[20] W. Lucia and A. Youssef, "Wyner wiretap-like encoding scheme for cyber-physical systems," *IET Cyber-Physical Systems: Theory Applications*, vol. 5, no. 4, pp. 359–365, 2020.

[21] A. Abdelwahab, W. Lucia, and A. Youssef, "Covert channels in cyber-physical systems," *IEEE Control Systems Letters*, vol. 5, no. 4, pp. 1273–1278, 2021.

[22] W. Lucia and A. Youssef, "Covert channels in stochastic cyber-physical systems," *IET Cyber-Physical Systems: Theory & Applications*, 2020.

[23] M. Darouach, M. Zasadzinski, A. B. Onana, and S. Nowakowski, "Kalman filtering with unknown inputs via optimal state estimation of singular systems," *International Journal of Systems Science*, vol. 26, no. 10, pp. 2015–2028, 1995.

[24] S. Z. Yong, M. Zhu, and E. Frazzoli, "A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems," *Automatica*, vol. 63, pp. 321–329, 2016.

[25] L. Ljung, "System identification," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–19, 1999.

[26] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM Journal on Computing*, vol. 17, no. 2, pp. 210–229, 1988.

[27] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, 1979.

[28] R. Lidl and H. Niederreiter, *Introduction to finite fields and their applications*. Cambridge university press, 1994.

[29] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[30] W. C. Huffman and V. Pless, *Fundamentals of error-correcting codes*. Cambridge university press, 2010.

[31] D. S. Karas, G. K. Karagiannidis, and R. Schober, "Neural network based phy-layer key exchange for wireless communications," in *22nd International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2011, pp. 1233–1238.

[32] K. H. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, 2000.