IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 November 2023, accepted 3 February 2024, date of publication 12 February 2024, date of current version 21 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365270

APPLIED RESEARCH

Adversarial Inference Control in Cyber-Physical Systems: A Bayesian Approach With Application to Smart Meters

RAMANA R. AVULA[1], TOBIAS J. OECHTERING[2], (Senior Member, IEEE), AND DANIEL MÅNSSON[3]

[1]Department of Electrification and Reliability, RISE Research Institutes of Sweden, 504 62 Borås, Sweden
[2]Division of Intelligent Systems, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
[3]Division of Electromagnetic Engineering and Fusion Science, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

Corresponding author: Tobias J. Oechtering (oech@kth.se)

ABSTRACT With the emergence of cyber-physical systems (CPSs) in utility systems like electricity, water, and gas networks, data collection has become more prevalent. While data collection in these systems has numerous advantages, it also raises concerns about privacy as it can potentially reveal sensitive information about users. To address this issue, we propose a Bayesian approach to control the adversarial inference and mitigate the physical-layer privacy problem in CPSs. Specifically, we develop a control strategy for the worst-case scenario where an adversary has perfect knowledge of the user's control strategy. For finite state-space problems, we derive the fixed-point Bellman's equation for an optimal stationary strategy and discuss a few practical approaches to solve it using optimization-based control design. Addressing the computational complexity, we propose a reinforcement learning approach based on the Actor-Critic architecture. To also support smart meter privacy research, we present a publicly accessible ''Co-LivEn'' dataset with comprehensive electrical measurements of appliances in a co-living household. Using this dataset, we benchmark the proposed reinforcement learning approach. The results demonstrate its effectiveness in reducing privacy leakage. Our work provides valuable insights and practical solutions for managing adversarial inference in cyber-physical systems, with a particular focus on enhancing privacy in smart meter applications.

INDEX TERMS Adversarial inference, Bayesian control, cyber-physical systems, deep reinforcement learning, privacy control, smart meters.

I. INTRODUCTION

A cyber-physical system (CPS) integrates physical components with computational and communication elements to enable real-time monitoring and control of physical systems. CPSs provide substantial advantages in utility systems such as electricity grids, water and gas supply networks, and transportation systems, including enhanced efficiency, stability, and automated network control. For instance, smart electric grids can use CPSs to monitor power usage and adjust supply and demand in real time, reducing energy waste and costs. However, the integration of CPSs in utility systems can also pose potential privacy risks as the usage patterns of resources can reveal sensitive private information about users to anyone with access to the data. For example, energy consumption data from smart meters (SMs) can be used to infer the types of household appliances [1] and their usage patterns, thereby disclosing sensitive private information about users, including presence or absence and the number of occupants, daily routines, and entertainment habits of occupants, medical equipment usage [2]. This information is susceptible to exploitation by malicious actors or unauthorized third parties for various purposes, such as targeted advertising or surveillance. The General Data Protection Regulation (GDPR) in Europe establishes stringent guidelines for handling data containing sensitive personal

The associate editor coordinating the review of this manuscript and approving it for publication was Junggab Son.

information. Specifically, the GDPR forbids processing data that could disclose such information without obtaining users' informed consent. For instance, this means that, when using SM data, one should not be able to infer appliance usage patterns that may reveal the religious beliefs of consumers without their explicit consent. Hence, it is crucial to develop privacy-enhancing methods for CPSs that safeguard users' privacy while still enabling their benefits.

Consider a hypothetical scenario where a third-party energy service provider is not only aware of the user's energy consumption patterns but also has perfect knowledge of the privacy-enhancing control strategy employed by the user. This situation presents a significant challenge for the user as the adversary is well-equipped to exploit any weaknesses in the control strategy, potentially leading to the exposure of private information about the user's habits, preferences, or lifestyle. In our previous work [3], we studied the problem of optimally controlling the sequential Bayesian hypothesis testing (SBHT) of an adversary who is unaware of the presence of a control system. In this work, we address an even stronger privacy question: *How can a user protect their privacy against an adversary who has perfect knowledge about the control strategy employed by the user?* By addressing this question, we aim to design conservative privacy control strategies against a worst-case adversary performing SBHT, which can serve as a benchmark.

A. RELATED WORKS

Addressing the privacy risks associated with smart meter data, several privacy-enhancing techniques have been proposed in the literature to protect sensitive user information without compromising the overall utility and benefits of smart meters. These techniques can be broadly classified into two approaches: Data Manipulation and Demand Shaping.

1) DATA MANIPULATION

Data manipulation techniques aim to protect user privacy by altering measured smart meter data before transmission. [4] presents a privacy-preserving smart metering approach using homomorphic encryption that allows computation of the aggregated energy consumption of a given set of users without accessing individual user data directly. In [5], the authors present a more efficient and scalable approach for data aggregation using Secure Multi-Party Computation (SMPC) cryptographic technique. The authors in [6] propose a privacy-preserving protocol using zero-knowledge proofs that enables billing with time-of-use tariffs without disclosing the actual consumption profile to the supplier. In [7], a simple and efficient method to preserve differential privacy is proposed by adding noise to the SM data in such a way that makes it difficult to learn anything about an individual, but still allows for accurate statistics to be computed. More recently, in [8] a data obfuscation method utilizing both differential privacy preserving data perturbation and a cryptographic noise distribution.

Data manipulation techniques, while useful, have certain drawbacks. First, altering the reported values may reduce their usefulness for grid management and load prediction, ultimately undermining the benefits of smart meters. Second, since these techniques do not address the issue at the physical layer level, adversaries with access to power lines could have potentially installed separate sensors, thereby circumventing the privacy protection offered by such methods.

2) DEMAND SHAPING

Demand shaping techniques physically alter the actual user energy demand from the grid in real-time to obfuscate sensitive information that can be inferred from SM data. This is achieved using energy storage systems (ESSs), flexible loads such as heating systems, and renewable energy sources. These physical layer techniques are highly effective in enhancing privacy since they limit information leakage even before data generation.

Several analytical techniques have been proposed in literature that quantify privacy using differential privacy measures [9], information-theoretic measures such as mutual information [10], [11], [12], [13], conditional entropy [14], and others, providing axiomatic guarantees on the maximum possible information leakage. Detection-theoretic privacy-enhancing techniques, on the other hand, offer *operational* privacy guarantees, such as protection against hypothesis tests [3], [15], [16], [17]. Related to controller-aware adversarial hypothesis testing, few attempts have been made in the literature to develop control policies to worsen adversarial detection performance. Li et al. [15] presented an optimal control strategy against a greedy and informed adversary conducting independent single-shot hypothesis tests. In [17], an informed adversary performing hypothesis tests on a static binary state is studied, and fundamental limits on achievable privacy are presented. In [16], the authors formulate a partially observable Markov decision process (POMDP) control problem, where the belief state of an informed adversary is optimally controlled over a given horizon, and the adversary is assumed to perform instantaneous hypothesis tests using only current observation. In another system setting, Liao et al. [13] proposed a privacy-enhancing mechanism to aid hypothesis testing while constrained by mutual information privacy measure, which differs from our work where hypothesis testing is used to model the adversary. In another related work, Salehkalaibar et al. [18] define a binary hypothesis state at the control level, where the controller is either in idle or privacy-enhancing state, and analyze hypothesis testing at the utility provider with access to some side information.

Heuristic-based computationally efficient techniques have also been proposed in literature. Notable approaches are the Best Effort Moderation approach [19] that aims to maintain a constant metered load by using a battery, and the Lazy stepping approach [20] that provides privacy by increasing the quantization error of the smart meter data by converting

the grid load into a step function using an arbitrary number of quantization levels. While these heuristic approaches are easier to implement in practice, their ability to provide formal privacy guarantees and comply with legal standards could be limited due to their reliance on simplistic and pre-defined rules, especially when faced with adversaries with knowledge about these rules.

B. CONTRIBUTIONS

In this paper, we address a strong physical layer privacy case by considering an adversary with complete knowledge of the user's control strategy and modeling the adversary's inferences using the SBHT inference model. Using the Markov decision process (MDP) framework, we measure privacy leakage in the physical layer using the Bayesian risk (adversarial reward) in the SBHT. For a finite state-space system, we derive a fixed-point equation for an optimal stationary strategy that minimizes the discounted aggregate value of the infinite-horizon Bayesian risk. The fixed-point equation is similar to Bellman's fixed-point equation of a continuous MDP with infinite state and action spaces and with a non-linear objective function that is impractical to solve exactly without making simplifying approximations. In this paper, we present a few practical approaches to solve it using optimization-based control design, highlighting their computational complexities.

Although exact optimal policies are theoretically computable using optimization-based approaches, they become intractable for high-dimensional state-space problems. To tackle the computational complexity, we introduce an actor-critic reinforcement learning (RL) algorithm named Adversarial Model-based Deterministic Policy Gradient (AMDPG). In an actor-critic RL, the actor is parameterized using a neural network, which can be used to generate continuous actions easily without the need for optimization procedures. The critic provides a low-variance performance knowledge of the actor by parameterizing the expected return of its actions [21]. This model-free approach allows us to handle the complex and dynamic nature of cyber-physical systems effectively where traditional MDP dynamic programming approaches are intractable. The proposed AMDPG algorithm is inspired by the Deep Deterministic Policy Gradient (DDPG) method [22]. A key difference between our proposed AMDPG and the DDPG algorithms can be observed in the noise generation process. In AMDPG, we not only integrate randomly generated noise into the actor output but also add noise obtained from the solution of an optimization problem, which computes an instantaneously optimal policy. This modification is intuitively expected to encourage more effective policy exploration and take into account long-term rewards by functioning near instantaneously optimal control during the learning process.

Furthermore, to facilitate smart meter privacy research, we introduce the publicly accessible "Co-LivEn" dataset, which contains detailed electrical measurements of appliances in a co-living household. This dataset provides a valuable resource for studying the privacy implications of NILM and the effectiveness of various privacy-enhancing techniques. Finally, we benchmark the proposed reinforcement approach using the presented household energy consumption data. The results show the effectiveness of the proposed control strategy in reducing the privacy leakage even in the worst-case scenario when the adversary is aware of the exact control strategy employed by the user.

The important contributions of the paper can be summarized as follows:

1) A novel and implementable RL approach to address a worst-case adversary with complete knowledge of the user's control strategy.
2) Derivation of a fixed-point equation for an optimal stationary strategy that minimizes the discounted aggregate value of the infinite-horizon Bayesian risk and practical approaches to solve it.
3) A publicly accessible energy consumption dataset that includes comprehensive electrical measurements of appliances in a co-living household.
4) Benchmarking of the proposed strategies using both synthetic and real data.

C. ORGANIZATION OF THE PAPER

The rest of the paper is organized as follows. In Section II, we present an overview of the system along with its modeling. In Section III, we present the preliminaries on the adversarial inference using SBHT framework. In Section IV, we formulate the optimal inference control problem. Subsequently, in Section V, we present several optimization-based approaches to achieve adversarial inference control. Using reinforcement learning, we present a novel control approach in Section VI. Furthermore, numerical studies using synthetic and real data are presented in Section VII and Section VIII. Lastly, we conclude the paper in Section IX.

In this paper, unless otherwise stated, we use capital letters to denote random variables, lowercase letters for their realizations, and calligraphic letters for their alphabets. We use $\mathbf{A}_{k:k+i}$ to denote the vector $[A_k, A_{k+1}, \ldots, A_{k+i}]^\intercal$, and $\mathcal{A}_{k:k+i}$ to denote the Cartesian product $\mathcal{A}_k \times \mathcal{A}_{k+1} \times \cdots \times \mathcal{A}_{k+i}$. The expectation operator is denoted by $\mathbb{E}[\cdot]$, and the matrix transpose operator by $(\cdot)^\intercal$. $P_A(a)$ denotes a probability distribution function, and $\mathbb{I}$ denotes an indicator function with $\mathbb{I}_a = 1$ if a is true, and 0 otherwise. $\mathbf{0}_n$ and $\mathbf{1}_n$ are n-dimensional vectors with all entries as zeros and ones, respectively. Δ_n denotes an $(n-1)$-dimensional simplex. In summations, if not otherwise specified, the domain of a variable is its complete alphabet. Throughout the paper, we use the term *policy* to refer to a map from some information set to some action at a certain time instance, and the term *strategy* to refer to a sequence of policies.

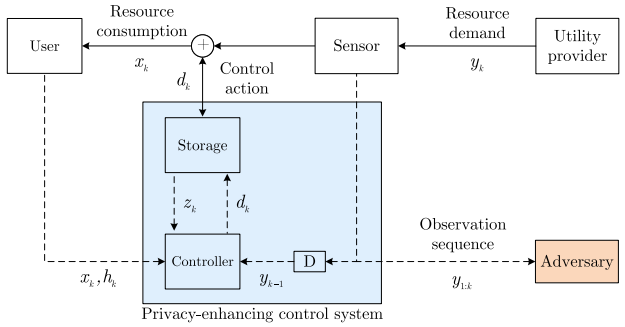

FIGURE 1. The proposed metering system that enables physical layer user privacy by altering the actual consumption using a storage. Here, the solid lines denote the physical resource flow and the dotted lines denote the information flow.

II. SYSTEM MODEL

We consider a privacy-concerned user consuming resource in a cyber-physical system as shown in Fig. 1. At the beginning of each slot $k \in \mathbb{N}$ in a discrete-time infinite-horizon $\mathbb{N} = \{1, 2, \dots\}$, the user's consumption is altered by a control strategy using a storage system. We restrict our analysis to discrete-time and discrete resource levels, designing a control strategy for systems with digital signal processing capabilities. We define random variables X_k and D_k on finite alphabets $\mathcal{X}$ and $\mathcal{D}$, respectively, where X_k represents the user's resource consumption, D_k represents the additional demand or usage of the stored resource specified by the control strategy, and $Y_k := X_k + D_k$ denotes the consumption measured by the sensor on a finite alphabet $\mathcal{Y} = \{x + d : x \in \mathcal{X}, d \in \mathcal{D}\}$.

We model the storage system's state transitions using a first-order Markov model characterized by the conditional distribution $P_{Z_{k+1}|Z_k,D_k}$, where Z_k represents the quantized value of the storage system state on a finite discrete alphabet $\mathcal{Z}$. We estimate the conditional distribution $P_{Z_{k+1}|Z_k,D_k}$ using Monte Carlo simulations and a sample-based density estimation approach [3]. In this work, we further simplify the storage system model by parametrizing the conditional distribution $P_{Z_{k+1}|Z_k,D_k}$ for each $(z_k, d_k) \in \mathcal{Z} \times \mathcal{D}$ using the beta distribution.

Moreover, to capture the sensitivity of user behavior, we introduce a privacy-sensitive state denoted by H_k, defined on a finite alphabet $\mathcal{H}$. This state, referred to as the *hypothesis state*, can represent various user activities such as cooking, taking a shower, and more, which are potentially of interest to an adversary seeking to infer personal information.

We model the dependency between the sequence of user demands and hypothesis states $[\mathbf{H}_{1:N}, \mathbf{X}_{1:N}]$ corresponding to the horizon $\mathbb{N}$ using a first-order time-homogeneous hidden Markov model (HMM) characterized by a set of parameters denoted as θ and is given by

$$\theta := \{P_{X_k|H_k}, P_{H_k|H_{k-1}}, P_{H_0} : k \in \mathbb{N}\},$$

where $P_{X_k|H_k}$ represents the observation probability of user demands, $P_{H_k|H_{k-1}}$ represents the transition probability of hypothesis states, and P_{H_0} represents the prior probability of hypothesis states.

As we will discuss in the subsequent analysis, the statistics of the ESS state Z_k are also relevant to the adversary when attempting to guess the hypothesis state. For analytical simplicity, we replace the state pair (Z_k, H_k) with a 1-dimensional random variable A_k. Let $\mathcal{A} := \{1, \dots, |\mathcal{Z}| \times |\mathcal{H}|\}$ denote the alphabet of A_k. We define $\mathbf{I}_k$ as the causal information available to the controller at the start of slot k, which is a discrete set defined on $\mathcal{I}_k = (\mathcal{X} \times \mathcal{Y} \times \mathcal{A})^{k-1}$ and includes $[\mathbf{X}_{1:k-1}, \mathbf{Y}_{1:k-1}, \mathbf{A}_{1:k-1}]$. Let $\mu_k \in \mathcal{U}_k$ denote a randomized *control policy* which represents the conditional distribution $P_{Y_k|X_k,H_k,\mathbf{I}_k}$, where $\mathcal{U}_k$ denotes the set of all randomized control policies.

III. ADVERSARIAL INFERENCE

Here we assume a strong adversary who knows the HMM parameter set θ, the storage system state transition probability $P_{Z_k|Z_{k-1},D_k}$, and the exact control strategy $\mu_{1:\infty}$ employed by the user. We model the adversary's inferences about the hypothesis state H_k in the infinite-horizon case using the SBHT framework. Let $\hat{H}_k$ denote the hypothesis state estimate of the adversary, which is defined on $\mathcal{H}$. We also define a randomized *detection policy* for the adversary, denoted by $\zeta_k \in \mathcal{C}_k$, which represents the conditional distribution $P_{\hat{H}_k|\mathbf{Y}_{1:k}}$, where $\mathcal{C}_k$ denotes the set of all randomized detection policies for the adversary.

In the SBHT framework [23], a reward is assigned to each possible test outcome denoted by $c(h, \hat{h})$ with $h, \hat{h} \in \mathcal{H}$. An optimal detection strategy is designed by maximizing the expected reward. The expected reward at time slot k, known as the *Bayesian reward*, is denoted by r_k and is given by

$$r_k = \mathbb{E}\big[c(H_k, \hat{H}_k)\big] = \textstyle\sum_{(h_k,\hat{h}_k)} c(h_k, \hat{h}_k) P_{H_k,\hat{H}_k}(h_k, \hat{h}_k). \quad (1)$$

where $P_{H_k,\hat{H}_k}$ is the joint distribution of the hypothesis state estimate $\hat{H}_k$ and the true hypothesis state H_k. Note that r_k represents the *Bayesian risk*, which is the average privacy cost for the user given a fixed function $c(h, \hat{h})$. In this work, we assume that $c(h, \hat{h}) \geq 0$ and that the reward for a correct guess is greater than that for an incorrect guess. As a result, the adversary seeks to maximize the average Bayesian reward, while the user aims to minimize it.

For any finite horizon $\mathcal{K}_N = \{1, 2, \dots, N\}$ of arbitrary length N, an optimal detection strategy of the adversary for any given control strategy $\mu_{1:N}$, denoted by $\zeta^*_{1:N}$, that maximizes the average Bayesian reward can be expressed as

$$\begin{aligned}\zeta^*_{1:N}(\mu_{1:N}) &= \underset{\zeta_{1:N} \in \mathcal{C}_{1:N}}{\operatorname{argmax}} \left[\frac{1}{N} \sum_{k=1}^{N} r_k(\zeta_k, \mu_{1:k})\right] \\ &= \underset{\zeta_{1:N} \in \mathcal{C}_{1:N}}{\operatorname{argmax}} \left[\mathbb{E}\Big[\sum_{k=1}^{N} r_{k|k}(\mathbf{Y}_{1:k}; \zeta_k, \mu_{1:k})\Big]\right], \quad (2)\end{aligned}$$

where $r_{k|k}$ denotes the *conditional Bayesian reward* of the adversary given the causal data $\mathbf{y}_{1:k} \in \mathcal{Y}^k$, expressed as

$$\begin{aligned} &r_{k|k}(\mathbf{y}_{1:k};\, \zeta_k,\, \mu_{1:k}) \\ &= \mathbb{E}\big[c(H_k, \hat{H}_k)\big|\mathbf{Y}_{1:k} = \mathbf{y}_{1:k}\big] \\ &= \textstyle\sum_{(h_k,\hat{h}_k)} c(h_k, \hat{h}_k)\, P_{H_k,\hat{H}_k|\mathbf{Y}_{1:k}}(h_k, \hat{h}_k|\mathbf{y}_{1:k}). \end{aligned} \tag{3}$$

Lemma 1: Let $\hat{\pi}_k$ denote the belief state of the adversary at time slot k, which represents the conditional probability vector $P_{A_k|\mathbf{Y}_{1:k}}$. For any given data $\mathbf{y}_{1:k}$ and control strategy $\mu_{1:k}$, the belief state of the adversary evolves according to the recursion:

$$\hat{\pi}_k(y_k, \bar{\mu}_k, \hat{\pi}_{k-1}) = \frac{\mathbf{M}_{\hat{\pi}}\left(y_k, \bar{\mu}_k\right) \cdot \hat{\pi}_{k-1}}{\mathbf{1}_{|\mathcal{A}|}^{\mathsf{T}} \cdot \mathbf{M}_{\hat{\pi}}\left(y_k, \bar{\mu}_k\right) \cdot \hat{\pi}_{k-1}}, \tag{4}$$

$$= \frac{\mathbf{M}_{\bar{\mu}}\left(y_k, \hat{\pi}_{k-1}\right) \cdot \bar{\mu}_k}{\mathbf{1}_{|\mathcal{A}|}^{\mathsf{T}} \cdot \mathbf{M}_{\bar{\mu}}\left(y_k, \hat{\pi}_{k-1}\right) \cdot \bar{\mu}_k}, \tag{5}$$

where $M_{\hat{\pi}}$ and $M_{\bar{\mu}}$ are $|\mathcal{A}| \times |\mathcal{A}|$ and $|\mathcal{A}| \times |\mathcal{W}|$ dimensional matrices respectively, and $\bar{\mu}_k$ denotes the control sub-policy, which represents the conditional probability $P_{W_k|\mathbf{Y}_{1:k-1}}$, and W_k denotes the conditional random variable $Y_k|X_k, H_k, A_{k-1}$ defined on the alphabet $\mathcal{W} := \{1, \ldots, |\mathcal{Y} \times \mathcal{X} \times \mathcal{H} \times \mathcal{A}|\}$.
The proof of Lemma 1 can be found in Appendix A. Note that the conditional Bayesian reward $r_{k|k}$ depends on the causal data $\mathbf{y}_{1:k-1}$ only through the sub-policy $\bar{\mu}_k$ and the belief state $\hat{\pi}_{k-1}$. Furthermore, for a fixed control strategy $\mu_{1:N}$, the sub-policy $\bar{\mu}_k$ and the belief state $\hat{\pi}_{k-1}$ do not depend on the detection strategy $\zeta_{1:N}$. Therefore, the optimization problem for the optimal SBHT adversarial detection strategy $\zeta_{1:N}^*$ in (2) can be decomposed into N linear programs:

$$\zeta_k^*(y_k, \bar{\mu}_k, \hat{\pi}_{k-1}) = \underset{\hat{h}_k \in \mathcal{H}}{\operatorname{argmax}}\Big[\mathbf{c}^{\mathsf{T}}(\hat{h}_k) \cdot \hat{\pi}_k(y_k, \bar{\mu}_k, \hat{\pi}_{k-1})\Big], \tag{6}$$

where $\mathbf{c}(\hat{h}_k)$ is a $|\mathcal{A}|$-dimensional vector with its elements defined as $[\mathbf{c}(\hat{h}_k)]_a = c(f_{\mathrm{H}}(a), \hat{h}_k)$ and $f_{\mathrm{H}} : \mathcal{A} \to \mathcal{H}$ is any deterministic function that maps the state A_k to its corresponding hypothesis state H_k. The adversarial detection policy given by (6) can be represented using a time-invariant and deterministic decision rule $\zeta^* : \Delta_{|\mathcal{A}|} \to \mathcal{H}$ given by

$$\zeta^*(\hat{\pi}_k) = \underset{\hat{h}_k \in \mathcal{H}}{\operatorname{argmax}}\Big[\mathbf{c}^{\mathsf{T}}(\hat{h}_k) \cdot \hat{\pi}_k\Big]. \tag{7}$$

Remark 1: In the computation of the belief state using (4) and (5), the current belief state $\hat{\pi}_k$ depends on the past observations $\mathbf{y}_{1:k-1}$ only through the sub-policy $\bar{\mu}_k$ and the previous belief state $\hat{\pi}_{k-1}$. However, the computation of the sub-policy $\bar{\mu}_k$ using (47) requires complete data $\mathbf{y}_{1:k-1}$ at each time slot k if the control policy μ_k is explicitly designed in the form $P_{Y_k|X_k,H_k,\mathbf{I}_k}$. As we will show in the coming section, using a control policy in this form with complete historical data does not help the user in minimizing the average Bayesian reward.

The information flow in the adversarial SBHT detection process is illustrated in Fig. 2. The optimization problem

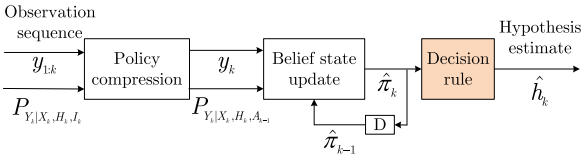

FIGURE 2. Information flow in the adversarial SBHT detection process with a control policy of the form $P_{Y_k|X_k,H_k,\mathbf{I}_k}$.

in (7) defining the adversarial decision rule ζ^* can also be represented using polyhedral *decision regions* in the simplex space $\Delta_{|\mathcal{A}|}$. For each hypothesis state $h \in \mathcal{H}$, the optimal adversarial decision region $\mathcal{R}_h$ is defined by

$$\mathcal{R}_h = \Big\{\hat{\pi}_k \in \Delta_{|\mathcal{A}|} : \big[\mathbf{c}^{\mathsf{T}}(h') - \mathbf{c}^{\mathsf{T}}(h)\big]\hat{\pi}_k < 0,\ \forall h' \in \mathcal{H} \setminus h\Big\}.$$

The set of all decision regions is denoted by $\mathcal{R} := \{\mathcal{R}_h : h \in \mathcal{H}\}$, and it satisfies $\cup_{h\in\mathcal{H}}\mathcal{R}_h = \Delta_{|\mathcal{A}|}$. Since the adversarial detection policy ζ^* given in (7) is time-invariant, the stationary strategy $\zeta_{1:\infty}^* := [\zeta^*, \zeta^*, \ldots]$ also maximizes both the average and discounted Bayesian rewards, denoted by $\bar{w}$ and w_ρ respectively. Specifically, for a given control strategy $\mu_{1:\infty}$, the average and discounted Bayesian rewards of the informed adversary can be expressed as

$$\bar{w}(\mu_{1:\infty}) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} r_k(\zeta^*, \mu_{1:k}), \tag{8}$$

$$w_\rho(\mu_{1:\infty}) = \sum_{k=1}^{\infty} \rho^{k-1}\, r_k(\zeta^*, \mu_{1:k}), \tag{9}$$

where $\rho \in [0, 1)$ is a discount factor.

IV. INFERENCE CONTROL PROBLEM FORMULATION

The SBHT adversarial inference poses a significant threat to the privacy of users in CPSs. To mitigate this threat, we aim to find a stationary control strategy for the infinite horizon $\mathbb{N}$ that minimizes the discounted Bayesian risk w_ρ. While the infinite horizon average Bayesian risk $\bar{w}$ is also a suitable optimization objective, it may not perform well in practice due to the dynamic nature of user demands. Therefore, we focus on minimizing w_ρ to derive an optimal control policy that ensures the privacy of the CPS.

Given any control strategy $\mu_{1:\infty}$ and any $\epsilon > 0$, for a bounded reward function $c(h, \hat{h})$ and a sufficiently large N, we have

$$\sum_{k=N+1}^{\infty} \rho^{k-1} r_k(\zeta^*, \mu_{1:k}) < \epsilon, \tag{10}$$

since $\rho^{k-1} \to 0$ as $k \to \infty$. Hence, we approximate the minimum discounted Bayesian risk, denoted by w_ρ^*,

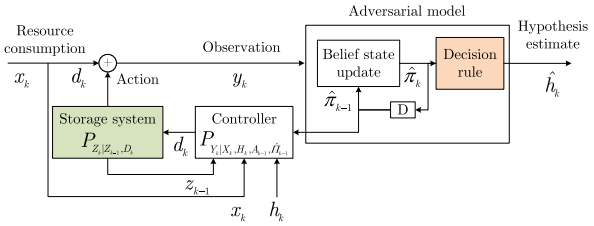

FIGURE 3. Information flow in the control system with the policy designed using the sufficient statistic $\hat{\pi}_k$.

as follows:

$$w_\rho^* \approx \min_{\mu_{1:N}} \left[\sum_{k=1}^{N} \rho^{k-1} r_k(\zeta^*, \mu_{1:k}) \right] = \min_{\mu_{1:N}} \left[\mathbb{E}\Big[\sum_{k=1}^{N} \rho^{k-1} r_{k|k^-}(\mathbf{Y}_{1:k-1}; \zeta^*, \mu_{1:k}) \Big] \right]. \quad (11)$$

Here, $r_{k|k^-}$ denotes the *conditional Bayesian risk* given any $\mathbf{y}_{1:k-1} \in \mathcal{Y}^{k-1}$, which is expressed as

$$r_{k|k^-}(\mathbf{y}_{1:k-1}; \zeta^*, \mu_{1:k}) = \mathbb{E}\big[c(H_k, \hat{H}_k) \big| \mathbf{Y}_{1:k-1} = \mathbf{y}_{1:k-1}\big] = \mathbf{c}^\intercal(\hat{h}_k^*) \cdot \mathbf{M}_{\hat{\pi}}\left(y_k, \bar{\mu}_k\right) \cdot \hat{\pi}_{k-1} \quad (12)$$

$$= \mathbf{c}^\intercal(\hat{h}_k^*) \cdot \mathbf{M}_{\bar{\mu}}\left(y_k, \hat{\pi}_{k-1}\right) \cdot \bar{\mu}_k, \quad (13)$$

where $\hat{h}_k^* = \zeta^*(\hat{\pi}_k)$ is the optimal hypothesis guess of the adversary given the belief state $\hat{\pi}_k$ which is updated according to (4) or (5) given $(y_k, \bar{\mu}_k, \hat{\pi}_{k-1})$. As mentioned in Remark 1, computing $\bar{\mu}_k$ when the control policy μ_k is explicitly designed in the form $P_{Y_k|X_k,H_k,\mathbf{I}_k}$ requires the complete data $\mathbf{y}_{1:k-1}$ at each k. In Theorem 1, we show that the belief state $\hat{\pi}_{k-1}$ is a sufficient statistic of the information vector i_k for finding an optimal control strategy that achieves the minimum achievable discounted Bayesian risk. Thus, using policies that rely on the entire data history $\mathbf{y}_{1:k-1}$ does not provide any performance improvement against the adversary.

Theorem 1: Let $\tilde{\mathcal{U}} = \{\tilde{\mu} : \Delta_{|\mathcal{A}|} \to \bar{\mathcal{U}}\}$ denote the set of all mappings from $\Delta_{|\mathcal{A}|}$ to $\bar{\mathcal{U}}$. For any discount factor $\rho \in (0, 1]$, at each $k \in \mathcal{K}_N$, there exists an optimal policy $\tilde{\mu}_k^* \in \tilde{\mathcal{U}}$ that achieves the minimum discounted Bayesian risk achievable by any policy $\mu_k \in \mathcal{U}_k$ and is given by

$$\tilde{\mu}_k^*(\hat{\pi}_{k-1}) = \underset{\bar{\mu}_k \in \bar{\mathcal{U}}}{argmin} \Big[\hat{\pi}_{k-1}^\intercal \big[\mathbf{M}_c(\bar{\zeta}_k^*) + \rho \mathbf{M}_v(\bar{\mathbf{v}}_{n-1}) \big] \bar{\mu}_k \Big], \quad (14)$$

$$v_n(\hat{\pi}_{k-1}) = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}} \Big[\hat{\pi}_{k-1}^\intercal \big[\mathbf{M}_c(\bar{\zeta}_k^*) + \rho \mathbf{M}_v(\bar{\mathbf{v}}_{n-1}) \big] \bar{\mu}_k \Big], \quad (15)$$

where $n = N - k + 1$ is the backward iteration index starting from $k = N$ for some arbitrarily large N. Here, v_n, known as the value function, is the aggregate value of discounted conditional Bayesian risk due to optimal strategy $\tilde{\mu}_{k:N}^*$.

The proof of Theorem 1 can be found in Appendix B. Fig. 3 illustrates the information flow in the control system with the policy $\tilde{\mu}_k \in \tilde{\mathcal{U}}$ designed using the sufficient statistic $\hat{\pi}_k$.

Corollary 1: For any discount factor $\rho \in (0, 1)$, there exists a unique fixed point value function $v^* : \Delta_{|\mathcal{A}|} \to \mathbb{R}_+$ to which the Bellman's recursion in (15) converges. Consequently, the optimal stationary policy $\tilde{\mu}^* \in \tilde{\mathcal{U}}$ that achieves the minimum discounted Bayesian risk w_ρ^* is the solution to the fixed point equation:

$$v^*(\hat{\pi}_{k-1}) = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}} \Big[\hat{\pi}_{k-1}^\intercal \big[\mathbf{M}_c(\bar{\zeta}_k^*) + \rho \mathbf{M}_v(\bar{\mathbf{v}}^*) \big] \bar{\mu}_k \Big] = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}} \Big[\max_{\bar{\zeta} \in \mathcal{H}^{|\mathcal{Y}|}} \big[\hat{\pi}_{k-1}^\intercal \mathbf{M}_c(\bar{\zeta}) \bar{\mu}_k \big] + \rho \hat{\pi}_{k-1}^\intercal \mathbf{M}_v(\bar{\mathbf{v}}^*) \bar{\mu}_k \Big], \quad (16)$$

where $[\bar{\mathbf{v}}^*]_y = v^*\big(\hat{\pi}_k(y, \bar{\mu}_k, \hat{\pi}_{k-1})\big)$.

Proof: Due to the monotonicity and contraction properties of the Bellman's recursion [24], the Banach's fixed point theorem [25] implies that there exists a unique fixed point $v^* \in \mathcal{V}$ to which the Bellman's recursion in (15) converges.□

Remark 2: The optimal SBHT control problem of minimizing the infinite-horizon discounted Bayesian risk can be formulated as a POMDP problem with continuous state $\hat{\pi}_{k-1}$, continuous action $\bar{\mu}_k$, and a piecewise-linear convex cost function: $\max_{\bar{\zeta}} \big[\hat{\pi}_{k-1}^\intercal \cdot \mathbf{M}_c(\bar{\zeta}) \cdot \bar{\mu}_k \big]$.

V. OPTIMIZATION-BASED CONTROL

In this section, we discuss several practical optimization-based approaches for solving the inference control problem (16) introduced in Section IV. These approaches take advantage of the problem's structure and apply various constraints on the discount factor, control policy space, and belief state space to simplify the optimization problem. These methods are particularly useful for scenarios where the state and action spaces of the system model are small. We also provide a discussion on the computational complexity of each approach.

A. INSTANTANEOUSLY-OPTIMAL CONTROL

An empirical upper bound on the minimum discounted Bayesian risk w_ρ^* can be obtained by using the instantaneously-optimal control policy with $\rho = 0$. The objective function in (16) becomes piecewise-linear with respect to $(\hat{\pi}_{k-1}, \bar{\mu}_k)$ when $\rho = 0$, enabling efficient computation of an exact instantaneously-optimal policy. The minimum instantaneous risk, denoted by $r_{k|k^-}^*$ is given by

$$r_{k|k^-}^*(\hat{\pi}_{k-1}) = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}} \Big[\max_{\bar{\zeta} \in \mathcal{H}^{|\mathcal{Y}|}} \big[\hat{\pi}_{k-1}^\intercal \mathbf{M}_c(\bar{\zeta}) \bar{\mu}_k \big] \Big] = \min_{\substack{\bar{\zeta} \in \mathcal{H}^{|\mathcal{Y}|}, \\ \bar{\mu}_k \in \bar{\mathcal{U}}_r(\bar{\zeta}, \hat{\pi}_{k-1})}} \Big[\hat{\pi}_{k-1}^\intercal \mathbf{M}_c\big(\bar{\zeta}\big) \bar{\mu}_k \Big], \quad (17)$$

where $\bar{\mathcal{U}}_r(\bar{\zeta}, \hat{\pi}_{k-1})$ denotes the set of all policies $\bar{\mu}_k \in \bar{\mathcal{U}}$ that satisfy the set of constraints imposed by the decision regions $\mathcal{R}$ corresponding to the decision vector $\bar{\zeta}$ and the belief state

$\hat{\pi}_{k-1}$, given by

$$\bar{\mathcal{U}}_r(\bar{\zeta}, \hat{\pi}_{k-1}) = \Big\{\bar{\mu}_k \in \bar{\mathcal{U}} : \hat{\pi}_k(y, \bar{\mu}_k, \hat{\pi}_{k-1}) \in \mathcal{R}_h, h = [\bar{\zeta}\,]_y, y \in \mathcal{Y}\Big\}. \tag{18}$$

Note that $\bar{\mathcal{U}}_r(\bar{\zeta}, \hat{\pi}_{k-1})$ is a polyhedron in $\mathbb{R}_+^{|\mathcal{W}|}$ since the belief state $\hat{\pi}_{k-1}$ evolves to $\hat{\pi}_k$ following the linear-fractional transformation given in (4) and the adversarial decision regions $\mathcal{R}$ are also polyhedrons in the belief space $\Delta_{|\mathcal{A}|}$. As a result, the exact instantaneously-optimal policy can be obtained in real-time by solving a linear program for the observed belief state $\hat{\pi}_{k-1}$.

Remark 3: Computing the exact instantaneously-optimal policy requires solving a piecewise minimum over the set $\mathcal{H}^{|\mathcal{Y}|}$ as given in (17). Therefore, the worst-case time complexity of the instantaneously-optimal control policy is $\mathcal{O}(\mathcal{H}^{|\mathcal{Y}|})$, which may grow exponentially with the size of the observation space $|\mathcal{Y}|$.

B. OPTIMAL CONTROL WITH FINITE SUB-POLICY SPACE

Here, we present an approach to solve the SBHT control problem by restricting the feasible space of the control policies in (15) to a finite set of control sub-policies, denoted by $\bar{\mathcal{U}}_F$. For example, $\bar{\mathcal{U}}_F$ can be chosen to be the finite set of all degenerate sub-policies. With a finite control sub-policy space, for each $\hat{\pi}_{k-1} \in \Delta_{|\mathcal{A}|}$, the Bellman's equation in (15) can be rewritten as

$$v_n(\hat{\pi}_{k-1}) = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}_F} \Big[\hat{\pi}_{k-1}^{\intercal} \cdot \gamma_k(\bar{\mu}_k, \hat{\pi}_{k-1}, v_{n-1})\Big], \tag{19}$$

where $\gamma_k(\bar{\mu}_k, \hat{\pi}_{k-1}, v_{n-1})$ is a vector in $\mathbb{R}_+^{|\mathcal{A}|}$ with its elements given by

$$\big[\gamma_k(\bar{\mu}_k, \hat{\pi}_{k-1}, v_{n-1})\big]_{a_{k-1}} = \sum_{(h_k, a_k, y_k)} P_{H_k, A_k, Y_k|A_{k-1}} \times \big[c(h_k, \zeta^*(\hat{\pi}_k(y_k, \bar{\mu}_k, \hat{\pi}_{k-1})) + \rho v_{n-1}(\hat{\pi}_k(y_k, \bar{\mu}_k, \hat{\pi}_{k-1}))\big].$$

The set of all hyperplanes in $\mathbb{R}_+^{|\mathcal{A}|}$ that define the boundaries of the decision regions $\mathcal{R}$ is denoted by $\bar{\mathcal{B}}$, and is given by

$$\bar{\mathcal{B}} = \Big\{\big[\mathbf{c}^{\intercal}(\hat{h}') - \mathbf{c}^{\intercal}(\hat{h})\big]\mathbf{b} = 0 : (h, h') \in \mathcal{H}^2, h \neq h'\Big\}. \tag{20}$$

Let $\bar{\mathcal{B}}^0 = \bar{\mathcal{B}}$ and for $n \geq 1$, $\bar{\mathcal{B}}^n$ denotes the set of all hyperplanes in $\mathbb{R}_+^{|\mathcal{A}|}$ given by

$$\bar{\mathcal{B}}^n = \Big\{\beta_i^{\intercal}\mathbf{M}_{\hat{\pi}}(y, \bar{\mu})\mathbf{b} = 0 : \{\beta_i^{\intercal}\tilde{\mathbf{b}} = 0\} \in \bar{\mathcal{B}}^{n-1}, i \in [1, |\bar{\mathcal{B}}^{n-1}|], y \in \mathcal{Y}, \bar{\mu} \in \bar{\mathcal{U}}_F\Big\}. \tag{21}$$

Since the control sub-policy space $\bar{\mathcal{U}}_F$ is finite, by initializing with $v_{N+1}(\hat{\pi}_N) = 0$, we can solve the optimization problem in (19) at each $k \leq N$ by recursively partitioning $\Delta_{|\mathcal{A}|}$ into a finite set of polyhedral partitions using all hyperplanes in $\bar{\mathcal{B}}^n \cup \bar{\mathcal{B}}$. These resulting polyhedral regions are called *Markov partitions*, similar to those in a POMDP control problem [26]. Within each Markov partition, the adversarial inference $\hat{h}_k^* = \zeta^*(\hat{\pi}_k)$ and the vector γ_k are constant with respect to the belief state $\hat{\pi}_{k-1}$. These Markov partitions along with corresponding γ_k vectors completely characterize the value function v_k in the Bellman's recursion (19).

In a POMDP control problem with a linear cost function, the value function v_k can be characterized without the need to compute Markov partitions at each iteration. Instead, it can be completely characterized by computing all possible γ_k vectors in the belief space $\Delta_{|\mathcal{A}|}$ [24, §7.5.1]. Remarkably, this result also holds true for a SBHT control problem, even when the cost function is piecewise-linear.

Proposition 1: Let $\mathcal{F}$ denote the set of all Markov partitions obtained by partitioning the unit simplex $\Delta_{|\mathcal{A}|}$ using all the hyperplanes in $\bar{\mathcal{B}}^1 \cup \bar{\mathcal{B}}$. Within each partition $\mathcal{F}_i \in \mathcal{F}$, the value function v_k is piecewise-linear and concave with respect to $\hat{\pi}_{k-1} \in \mathcal{F}_i$. That is,

$$v_n(\hat{\pi}_{k-1}) = \min_{\substack{\bar{\mu} \in \bar{\mathcal{U}}_F, \\ \gamma \in \Gamma_i(\bar{\mu}, n)}} \Big[\hat{\pi}_{k-1}^{\intercal} \cdot \gamma\Big], \tag{22}$$

where $n = N - k + 1$ denotes the backward iteration index starting from $k = N$ for some arbitrarily large N, and $\Gamma_i(\bar{\mu}, n)$ is a finite set of vectors given by

$$\Gamma_i(\bar{\mu}, n) = \oplus_{y \in \mathcal{Y}} \Big\{\frac{\mathbf{M}_c(\bar{\zeta}^*)\bar{\mu}}{|\mathcal{Y}|} + \mathbf{M}_{\hat{\pi}}^{\intercal}(y, \bar{\mu})\tilde{\gamma} : \tilde{\gamma} \in \Gamma_i^+(\bar{\mu}, n-1)\Big\}, \tag{23}$$

$$\Gamma_i^+(\bar{\mu}, n-1) = \cup_{\bar{\mu}' \in \bar{\mathcal{U}}_F} \Big\{\Gamma_j(\bar{\mu}', n-1) : \mathcal{F}_j \cap \mathrm{T}\big(\mathcal{F}_i, \mathbf{M}_{\hat{\pi}}(y, \bar{\mu})\big) \neq \emptyset,\ \forall j \in [1, |\mathcal{F}|]\Big\}, \tag{24}$$

initialized with

$$\Gamma_j(\bar{\mu}', 0) = \mathbf{0}_{|\mathcal{A}|},\ \forall j \in [1, |\mathcal{F}|],\ \bar{\mu}' \in \bar{\mathcal{U}}_F. \tag{25}$$

Here, $[\bar{\zeta}^*]_y = \zeta^*(\hat{\pi}_k(y, \bar{\mu}, \hat{\pi}_{k-1}))$ and $\mathrm{T}\big(\mathcal{F}_i, \mathbf{M}_{\hat{\pi}}(y, \bar{\mu})\big)$ denotes the affine transformation of the polyhedron $\mathcal{F}_i$ w.r.t. the belief transformation matrix $\mathbf{M}_{\hat{\pi}}(y, \bar{\mu})$ defined in (4); and $\oplus$ denotes the cross-sum operation, which is the pairwise addition of vectors from two sets.

The finite set $\Gamma_i(\bar{\mu}, n)$ in (23) is constructed backward recursively by taking a cross-sum of $|\mathcal{Y}|$ sets, where each set corresponding to a control action $y \in \mathcal{Y}$ contains all possible vectors $\frac{\mathbf{M}_c(\bar{\zeta}^*)\bar{\mu}}{|\mathcal{Y}|} + \mathbf{M}_{\hat{\pi}}^{\intercal}(y, \bar{\mu})\tilde{\gamma}$, as a result of Bellman's dynamic programming. Here, $\tilde{\gamma}$ belongs to the set $\Gamma_j(\bar{\mu}', n-1)$ corresponding to each $\bar{\mu}' \in \bar{\mathcal{U}}_F$ and each partition $\mathcal{F}_j \in \mathcal{F}$ that can be reached from $\mathcal{F}_i \in \mathcal{F}$ using an affine transformation using $\mathbf{M}_{\hat{\pi}}(y, \bar{\mu})$. The Prop. 1 can be shown using an induction technique similar to the proof of [24, Theorem 7.4.1] for a POMDP control problem. The construction of the Markov partitions set $\mathcal{F}$ using all the hyperplanes in $\bar{\mathcal{B}}^1 \cup \bar{\mathcal{B}}$ ensures that the adversarial decision vector $\bar{\zeta}^*$ and consequently, the instantaneous reward $r_{k|k-}^*$ in (17) are constants w.r.t. $\hat{\pi}_{k-1} \in \mathcal{F}_i$ at each iteration n. Furthermore, as the belief state $\hat{\pi}_k$ evolves according to

the linear fractional transformation in (4), the value function in (19) becomes piecewise-linear with respect to $\hat{\pi}_{k-1}$ within each Markov partition $\mathcal{F}_i$.

Remark 4: Although Prop. 1 shows that it is theoretically possible to compute the exact optimal stationary policy over a finite control sub-policy space $\bar{\mathcal{U}}_F$ using (22), the worst-case space complexity of this approach is exponential with respect to $|\bar{\mathcal{U}}_F|$, $|\mathcal{A}|$, $|\mathcal{Y}|$, and $|\mathcal{H}|$, and double exponential with respect to n, since $|\bar{\mathcal{B}}^1| = \mathcal{O}(|\mathcal{A}| \times |\mathcal{Y}| \times |\bar{\mathcal{U}}_F| \times |\mathcal{H}|)$, $|\mathcal{F}| = \mathcal{O}(\delta^{|\bar{\mathcal{B}}^1|})$, where $1 < \delta < 2$, and $|\Gamma_i(\bar{\mu}, n)| = \mathcal{O}\big((|\bar{\mathcal{U}}_F| \times |\mathcal{F}|)^{|\mathcal{Y}|^n}\big)$. Due to the high computational complexity of the approach, computing the optimal policy may not be feasible even for low-dimensional problems.

C. SUB-OPTIMAL CONTROL WITH FINITE SUB-POLICY SPACE

As mentioned in Section V-B, computing the exact optimal stationary policy over a finite control sub-policy space is intractable, even for small state-space problems. To overcome this issue, we propose a sub-optimal approach based on the method proposed by Lovejoy [27] for POMDP control problems with a linear cost function with respect to the belief state. Since the cost function of the SBHT control problem is piecewise-linear, we use a similar approach to find an upper bound on the minimum discounted Bayesian risk w^*_ρ empirically.

The key idea in this approach is to retain only a subset of the γ vectors in the set $\Gamma_i(\bar{\mu}, n)$ at each iteration, denoted by $\bar{\Gamma}_i(\bar{\mu}, n)$, thereby avoiding the double-exponential growth of the γ vectors. Given a set $\Gamma_i(\bar{\mu}, n)$ computed using (23), we first choose a finite set of arbitrary belief states within the corresponding partition $\mathcal{F}_i$, denoted by $\bar{\mathcal{F}}_i$. We then construct $\bar{\Gamma}_i(\bar{\mu}, n)$ for each $\bar{\mu} \in \bar{\mathcal{U}}_F$ as

$$\bar{\Gamma}_i(\bar{\mu}, n) = \left\{ \underset{\gamma \in \Gamma_i(\bar{\mu}, n)}{\operatorname{argmin}} \left[\hat{\pi}^\intercal \gamma\right] : \hat{\pi} \in \bar{\mathcal{F}}_i \right\} \tag{26}$$

We then iterate (23) using $\bar{\Gamma}_i(\bar{\mu}, n)$ instead of $\Gamma_i(\bar{\mu}, n)$ until convergence of the vectors in each set $\bar{\Gamma}_i(\bar{\mu}, n)$ up to some finite precision. This approach gives a sub-optimal stationary policy that yields an upper bound to w^*_ρ with a fixed space complexity of $\mathcal{O}\big((|\bar{\mathcal{U}}_F| \times |\bar{\mathcal{F}}|)^{|\mathcal{Y}|}\big)$ at each iteration, where $|\bar{\mathcal{F}}| = \sum_i |\bar{\mathcal{F}}_i|$ is the number of all belief states we choose within the simplex $\Delta_{|\mathcal{A}|}$.

D. SUB-OPTIMAL CONTROL WITH DISCRETE BELIEF SPACE

Here, we consider the SBHT control problem with a restricted belief state space that is discretized with some precision $\epsilon > 0$ for the probability measure. Let $\bar{\Delta}_{|\mathcal{A}|} \subset \Delta_{|\mathcal{A}|}$ denote the resulting finite discrete space, and let $\xi_i \in \bar{\Delta}_{|\mathcal{A}|}$ for $1 \leq i \leq m$ represent the discrete belief states. We use the nearest-neighbour (NN) classification boundaries in $\bar{\Delta}_{|\mathcal{A}|}$ to define the *Voronoi region* $\mathcal{N}_i$ of each ξ_i, which is a polyhedron in $\Delta_{|\mathcal{A}|}$ given by

$$\mathcal{N}_i = \left\{ \hat{\pi} \in \Delta_{|\mathcal{A}|} : \left\| \xi_i - \hat{\pi} \right\| \leq \left\| \xi_j - \hat{\pi} \right\|, \forall \xi_j \in \bar{\Delta}_{|\mathcal{A}|} \backslash \xi_i \right\}. \tag{27}$$

Fig. 4 illustrates the approximation of belief space in $\mathbb{R}^3_+$ to a discrete belief space obtained using 0.25 as the precision for the probability measure.

Let $\bar{\mathbf{g}}_k$ be a $|\mathcal{Y}|$-dimensional vector representing the indices of belief state $\hat{\pi}_k$ in $\bar{\Delta}_{|\mathcal{A}|}$. We denote by $\bar{\mathcal{U}}_b(\bar{\mathbf{g}}_k, \xi_i)$ the set of all control policies $\bar{\mu}_k \in \bar{\mathcal{U}}$ that satisfy the set of linear constraints:

$$\left\{ \frac{M_{\bar{\mu}}(y, \xi_i) \cdot \bar{\mu}_k}{\mathbf{1}^\intercal_{|\mathcal{A}|} \cdot M_{\bar{\mu}}(y, \xi_i) \cdot \bar{\mu}_k} \in \mathcal{N}_{[\bar{\mathbf{g}}_k]_y} : y \in \mathcal{Y} \right\}. \tag{28}$$

Then, the Bellman's equation in (15) for each $\hat{\pi}_{k-1} = \xi_i \in \bar{\Delta}_{|\mathcal{A}|}$ can be rewritten as a linear programming problem:

$$v_n(\xi_i) = \min_{\substack{\bar{\mathbf{g}}_k \in \bar{\Delta}^{|\mathcal{Y}|}_{|\mathcal{A}|}, \\ \bar{\mu}_k \in \bar{\mathcal{U}}_b(\bar{\mathbf{g}}_k, \xi_i)}} \left[\alpha_k^\intercal(\bar{\mathbf{g}}_k, \xi_i, v_{n-1}) \cdot \bar{\mu}_k \right]. \tag{29}$$

Here, $\alpha_k(\bar{\mathbf{g}}_k, \xi_i, v_{n-1})$ is a vector in $\mathbb{R}^{|\mathcal{W}|}_+$ with elements:

$$\begin{aligned} &\left[\alpha_k(\bar{\mathbf{g}}_k, \xi_i, v_{n-1})\right]_{w_k} \\ &\quad = \sum_{(a_k, y_k, x_k, h_k, a_{k-1})} P_{H_k|A_k} \\ &\qquad \times P_{W_k|Y_k, X_k, H_k, A_{k-1}} P_{X_k, A_k|Y_k, A_{k-1}} [\xi_i]_{a_{k-1}} \\ &\qquad \times \left[c(h_k, \zeta^*(\xi_{[\bar{\mathbf{g}}_k]_{y_k}})) + \rho v_{n-1}(\xi_{[\bar{\mathbf{g}}_k]_{y_k}}) \right]. \end{aligned} \tag{30}$$

Remark 5: Due to the linear fractional constraint in (28), the optimal solution $\bar{\mu}^*_k(\hat{\pi}_{k-1})$ in this case may not necessarily be a non-randomized control sub-policy.

Remark 6: The cardinality of the belief space $\bar{\Delta}_{|\mathcal{A}|}$ with precision ϵ is $\mathcal{O}(|\mathcal{H} \times \mathcal{Z}|^{(1/\epsilon)})$. Since (29) requires computing the piecewise minimum over $\bar{\Delta}_{|\mathcal{A}|}$, finding the optimal stationary policy has a worst-case time complexity of $\mathcal{O}(|\bar{\Delta}_{|\mathcal{A}|}|^{|\mathcal{Y}|}) = \mathcal{O}((|\mathcal{H} \times \mathcal{Z}|^{(1/\epsilon)})^{|\mathcal{Y}|})$. Thus, the computation time may grow double exponentially with respect to the cardinality of the observation space $|\mathcal{Y}|$.

Note that approximating $\hat{\pi}_k$ to the nearest discrete belief state $\xi_j \in \bar{\Delta}_{|\mathcal{A}|}$ introduces an approximation error in the value function v_n at each iteration n. This error can propagate through the recursive iterations and may lead to a sub-optimal policy against an adversary that uses a more precise belief state $\hat{\pi}_k$. Therefore, the trade-off between the precision of the belief state space and the computational cost needs to be carefully considered when using this approach.

VI. REINFORCEMENT LEARNING-BASED CONTROL

Although the SBHT inference control problem (16) can be solved approximately using the approaches presented in Section V, due to their complexity, they are only computationally tractable for low-dimensional problems. To address this challenge, we present a reinforcement learning-based control approach based on the Actor-Critic architecture [28]. Our approach, called Adversarial Model-based Deterministic

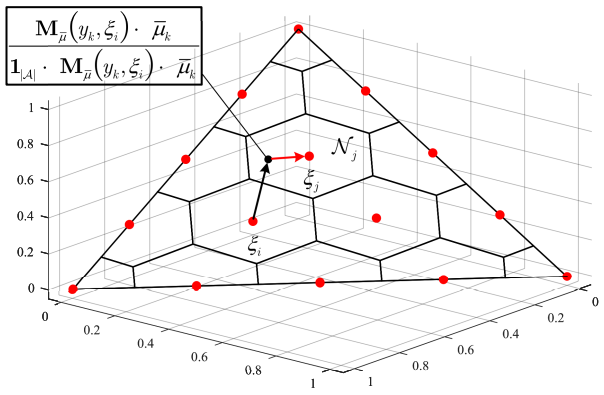

FIGURE 4. Illustration of belief space discretization in $\mathbb{R}^3_+$.

Policy Gradient (AMDPG), is inspired by the Deep Deterministic Policy Gradient (DDPG) algorithm [22] and enables tractable policy computation even in high-dimensional problems. The AMDPG algorithm is presented in Alg. 1.

The goal of the AMDPG algorithm is to learn a deterministic policy that maps the current belief state to an optimal control sub-policy based on a critic's evaluation of the quality of the action. Like the DDPG algorithm, the critic evaluates the actor by estimating the state-action value function, denoted by Q_μ, corresponding to any stationary policy μ, which is expressed as

$$\begin{aligned} Q_\mu(\hat{\pi}_{k-1},\bar{\mu}_k) &= \mathbb{E}_\mu\left[\sum_{k=1}^{\infty}\rho^{k-1}\,c(H_k,\hat{H}_k)\Big|\hat{\pi}_{k-1},\bar{\mu}_k\right] \\ &= \mathbb{E}_\mu\left[r_{k|k^-}(\hat{\pi}_{k-1},\bar{\mu}_k)+\rho Q_\mu(\hat{\pi}_k,\bar{\mu}_{k+1})\right], \end{aligned} \tag{31}$$

The optimal state-action value function $Q_{\tilde{\mu}^*}$ and the state value function v^* in (16) for the optimal stationary policy $\tilde{\mu}^*$ are related through the stationary distribution of the belief states, denoted by $P_{\hat{\Pi}_\infty}(\hat{\pi})$, as:

$$w_\rho(\tilde{\mu}^*)=\int_{\hat{\pi}}P_{\hat{\Pi}_\infty}(\hat{\pi})v^*(\hat{\pi})=\int_{\hat{\pi}}P_{\hat{\Pi}_\infty}(\hat{\pi})Q_{\tilde{\mu}^*}(\hat{\pi},\tilde{\mu}^*(\hat{\pi})). \tag{32}$$

In the AMDPG algorithm, the actor network selects a control sub-policy at each time step by observing the belief state $\hat{\pi}_{k-1}$. The resulting belief state transition and adversarial Bayesian reward $r_{k|k} = \mathbf{c}^\intercal(\hat{h}_k)\hat{\pi}_k$ are stored in a replay buffer. By using random sample batches, the critic network estimates the expected reward, and the actor network updates its parameters by minimizing the expected reward using the gradient descent algorithm. To map the actor network outputs to the control sub-policy action space $\mathcal{W}$, the AMDPG algorithm employs an additive-logistic transformation denoted by $\mathcal{L}$. For $\eta\in\mathbb{R}^{n-1}$, $\mathcal{L}:\mathbb{R}^{n-1}\to\Delta_n$

Algorithm 1 Adversarial Model-Based Deterministic Policy Gradient (AMDPG)

Require: Replay buffer $\mathcal{B}$, critic network Q, actor network ψ, exploration noise ϵ_1,ϵ_2, batch size M, learning rates χ_Q and χ_ψ, network parameter η_Q and η_ψ.

Initialize $Q(\hat{\pi},\bar{\mu}|\eta_Q)$, $\mu(\hat{\pi}|\eta_\psi)$.
Initialize replay buffer $\mathcal{B}$.
for episode = 1 to T_{episodes} **do**
  Initialize state $\hat{\pi}_0$.
  for $k=1$ to N **do**
    $\bar{\mu}_k^{\#}\leftarrow\psi(\hat{\pi}_{k-1}|\eta_\psi)$
    Sample rand randomly from (0, 1).
    if rand $<\epsilon_1$ **then**
      Compute $P_{Y_k|Y_{1:k-1}}$ for $\bar{\mu}_k^{\#}$ using (36).
      Choose top $\mathcal{Y}_{exp}\subset\mathcal{Y}$ based on $P_{Y_k|Y_{1:k-1}}$.
      Solve (35) to get $\hat{\mu}_k^*(\bar{\phi}_a,\hat{\pi}_{k-1}),\forall\bar{\phi}_a\in\Phi_a$.
      Solve (38) to get $\hat{\mu}_k^\dagger$.
      $\bar{\mu}_k\leftarrow\hat{\mu}_k^\dagger$.
    else if rand $<\epsilon_1+\epsilon_2$ **then**
      $\bar{\mu}_k\leftarrow\mathcal{L}\left(\bar{\mu}_k^{\#}+\mathcal{N}\right)$
    else
      $\bar{\mu}_k\leftarrow\mathcal{L}\left(\bar{\mu}_k^{\#}\right)$
    end if
    Execute action $\bar{\mu}_k$ and next state $\hat{\pi}_k$
    Compute expected reward $r_{k|k}=\mathbf{c}^\intercal(\hat{h}_k)\hat{\pi}_k$.
    Store experience $(\hat{\pi}_{k-1},\bar{\mu}_k,r_k,\hat{\pi}_k)$ in $\mathcal{B}$.
    Sample random batch from $\mathcal{B}$ of size B.
    Compute $\nabla_{\eta_Q}Q$, the critic gradient using the loss:

$$\frac{1}{B}\sum_{i=1}^{B}(\omega_i-Q(\hat{\pi}_{i-1},\bar{\mu}_i|\eta_Q)^2,$$

    where $\omega_i=r_i+\rho Q(\hat{\pi}_i,\psi(\hat{\pi}_i|\eta_\psi)|\eta_Q)$.
    Compute $\nabla_{\eta_\psi}\psi$, the sample policy gradient [22]:

$$\mathbb{E}_{\text{batch}}[\nabla_{\eta_\psi}Q(\hat{\pi},\psi(\hat{\pi}|\eta_\psi)|\eta_Q)],$$

    using the batch samples.
    Update the actor and critic network parameters:

$$\eta_Q\leftarrow\eta_Q+\chi_Q\nabla_{\eta_Q}Q,$$
$$\eta_\psi\leftarrow\eta_\psi+\chi_\psi\nabla_{\eta_\psi}\psi.$$

  end for
end for

is given by

$$\mathcal{L}(\eta)=\left[\frac{e^{\eta_1}}{1+\sum_{i=1}^{n-1}e^{\eta_i}},\ldots,\frac{e^{\eta_{n-1}}}{1+\sum_{i=1}^{n-1}e^{\eta_i}},\frac{1}{1+\sum_{i=1}^{n-1}e^{\eta_i}}\right]^\intercal, \tag{33}$$

and its unique inverse transformation $\mathcal{L}^{-1}$ for $\kappa\in\Delta_n$ is

$$\mathcal{L}^{-1}(\kappa)=\left[\log\left(\frac{\kappa_1}{\kappa_n}\right),\ldots,\log\left(\frac{\kappa_{n-1}}{\kappa_n}\right)\right]^\intercal. \tag{34}$$

Because of the piecewise-linear structure of the reward as shown in (17), the AMDPG algorithm aims to enhance exploration by occasionally selecting an action based on a pool of solutions to the optimization problems that minimize the instantaneous rewards, that is:

$$\hat{\mu}_k^*(\bar{\phi}_a, \hat{\pi}_{k-1}) = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}_a(\bar{\phi}_a, \hat{\pi}_{k-1})} \left[\hat{\pi}_{k-1}^{\mathsf{T}} \mathbf{M}_c\big(\bar{\xi}(\bar{\phi}_a)\big) \bar{\mu}_k \right], \quad (35)$$

where $\bar{\phi}_a \in \Phi_a \subseteq \mathcal{A}^{|\mathcal{Y}_{exp}|}$, $\mathcal{Y}_{exp} \subseteq \mathcal{Y}$ contains only the top e elements from $\mathcal{Y}$, which are chosen according to the likelihood probability of observation y_k given the actor network output $\bar{\mu}_k^{\#}$, given by

$$P_{Y_k|\mathbf{Y}_{1:k-1}}(y|\mathbf{y}_{1:k-1}) = \mathbf{1}_{|\mathcal{A}|}^{\mathsf{T}} \cdot \mathbf{M}_{\bar{\mu}}\big(y, \hat{\pi}_{k-1}\big) \cdot \mathcal{L}(\bar{\mu}_k^{\#}), \quad (36)$$

the constraint set $\tilde{\mathcal{U}}_a$, similar to (18) as:

$$\tilde{\mathcal{U}}_a(\bar{\phi}_a, \hat{\pi}_{k-1}) = \Big\{ \bar{\mu} \in \bar{\mathcal{U}} : \hat{\pi}_k(y, \bar{\mu}, \hat{\pi}_{k-1}) \in \mathcal{N}_a, \; a = [\bar{\phi}_a\,]_y, y \in \mathcal{Y}_{exp} \Big\}, \quad (37)$$

and $\mathcal{N}_a$ is the *Voronoi region* of a simplex vertex $a \in \mathcal{A}$ given in (27). Then, the exploratory action $\hat{\mu}_k^{\dagger}$ is given by

$$\hat{\mu}_k^{\dagger}(\hat{\pi}_{k-1}) = \min_{\hat{\mu}_k^*(\bar{\phi}_a, \hat{\pi}_{k-1}), \forall \bar{\phi}_a \in \Phi_a} \Big[\mathbb{E}_{\mu} \big[r_{k|k^-}(\hat{\pi}_{k-1}, \hat{\mu}_k^*) + \rho Q_{\mu}(\hat{\pi}_k, \bar{\mu}_{k+1}) \big] \Big]. \quad (38)$$

Further, the set of potential vectors in Φ_a can be condensed by randomly selecting a pre-defined finite number of vectors from the $\mathcal{A}^{\mathcal{Y}_{exp}}$ space. This allows for a more manageable computation process to generate adversarial model-based noise. Then, to balance exploration and exploitation, the agent selects the exploratory action $\hat{\mu}_k^{\dagger}$ with probability ϵ_1, $\mathcal{L}(\bar{\mu}_k^{\#} + \mathcal{N})$ with probability ϵ_2, and the network output $\mathcal{L}(\bar{\mu}_k^{\#})$ with probability $1 - \epsilon_1 - \epsilon_2$, where $\mathcal{N}$ is any randomly generated noise.

The presented model-free RL approach allows us to handle the dynamic nature of complex CPSs effectively where traditional MDP dynamic programming approaches are insufficient. Note that the computational steps consist of solving a few linear programs and computing gradients of actor and critic. Due to the usage of a replay buffer, depending on the computational power of the agent, these computational steps can be spread across a few or several time steps of the controller. Moreover, a noteworthy feature of this approach is its runtime adaptability, as the policy can be learned and adjusted dynamically during system operation. This enables real-world deployability, even in situations where system dynamics may evolve or are not entirely known a priori. Overall, the AMDPG algorithm provides a computationally tractable approach for solving the SBHT inference control problem even in high-dimensional cases. Our simulation results in the next sections demonstrate that it achieves competitive performance compared to other approaches.

VII. NUMERICAL STUDY WITH SYNTHETIC DATA

In this section, we present a numerical study using synthetic data to evaluate the effectiveness of the proposed approaches for the SBHT inference control problem. We consider a simple system with binary state-spaces of $\mathcal{H}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. The system has three control actions $\mathcal{D}$, a horizon length of $N = 96$. We generate synthetic data using different HMMs with the same prior and observation probabilities, but varying transition probability with parameters $\lambda_0, \lambda_1 \in \{0.2, 0.4, 0.6, 0.8\}$. The HMM model parameters are set for all $k \in \mathbb{N}$ as follows:

$$P_{H_0} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \quad P_{X_k|H_k} = \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix},$$

$$P_{H_k|H_{k-1}} = \begin{bmatrix} \lambda_0 & 1-\lambda_1 \\ 1-\lambda_0 & \lambda_1 \end{bmatrix}.$$

To investigate the attainable privacy levels in relation to various system dynamics, we develop and assess the performance of the proposed control methods for different HMM transition probabilities. We model the state transitions of the storage system using the conditional distribution $P_{Z_k|Z_{k-1},D_k}$ with the following elements:

$$P_{Z_k|Z_{k-1},D_k}(1|1,0) = P_{Z_k|Z_{k-1},D_k}(2|2,0) = 1,$$
$$P_{Z_k|Z_{k-1},D_k}(1|1,1) = P_{Z_k|Z_{k-1},D_k}(2|2,-1) = 0.05,$$
$$P_{Z_k|Z_{k-1},D_k}(2|1,1) = P_{Z_k|Z_{k-1},D_k}(1|2,-1) = 0.95.$$

We select a cost function that protects less frequently occurring hypothesis states since they are more informative. Specifically, we set $c(h_i, h_i) = 1/P_{H_\infty}(h_i)$, where P_{H_∞} denotes the stationary probability of the HMM; and $c(h_i, h_j) = 0, \forall h_i \neq h_j$.

Before evaluating the privacy control approaches presented in the previous sections, we first implement an approach based on our previous work [3], where we design a policy that minimizes the discounted Bayesian reward of an adversary who is unaware of the control system's existence. We also discretize the belief states with a precision of 0.2 for the probability measure, as discussed in Section V-D, and use a discount factor of $\rho = 0.9$. We then evaluate this approach against an adversary with complete knowledge of the control system. Fig. 5 shows the average Bayesian reward corresponding to both the aware adversary (AA) and the unaware adversary (UA). This demonstrates that when the control system is designed for a weaker adversarial case, a stronger adversary can improve its detection performance with knowledge of the implemented control strategy. This result highlights the potential necessity to employ privacy control measures against the most extreme adversaries, ensuring that the system remains secure even in the face of worst-case scenarios.

As noted in Remark 4 in Section V-B, computing an exact optimal policy for even a simple binary system can be highly computationally complex. To illustrate this, we computed the γ vectors for an HMM with $(\lambda_0, \lambda_1) = (0.2, 0.2)$ using a finite control sub-policies space $\bar{\mathcal{U}}_F$ consisting

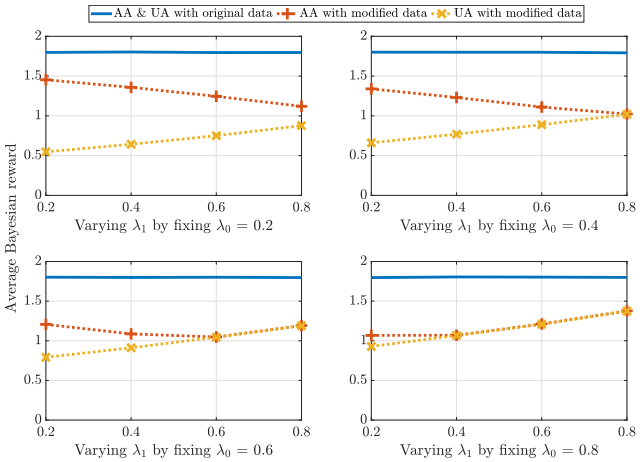

FIGURE 5. Comparison of average Bayesian rewards of aware adversary (AA) and unaware adversary (UA) when the control system is designed to minimize the Bayesian reward of the UA.

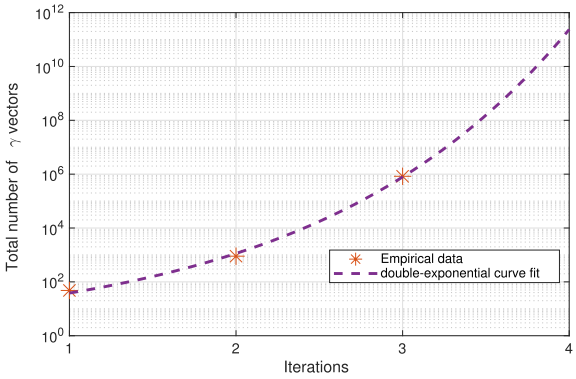

FIGURE 6. Double-exponential growth of γ vectors for a binary-state problem using the exact optimal approach in Section V-B.

of degenerate sub-policies and with belief transformation matrices $\mathbf{M}_{\hat{\pi}}(y_k, \bar{\mu})$ such that $\det(\mathbf{M}_{\hat{\pi}}(y, \bar{\mu})) > 0.01 \forall y \in \mathcal{Y}$. This results in $|\bar{\mathcal{U}}_F| = 4, |\bar{\mathcal{B}}^1| = 8, |\mathcal{F}| = 12$. Fig. 6 shows the double-exponential growth of the total number of γ vectors in all the sets $\Gamma_i(\bar{\mu}, n)$ for up to three iterations. Due to the impracticality of this approach, stemming from its excessive computational complexity, we exclude it from the following numerical study.

To design the sub-optimal policy with a finite control sub-policy space, as presented in Section V-C, we select the control sub-policy space $\bar{\mathcal{U}}_F$ to be the set of degenerate policies $\bar{\mu}$ with transformation matrices $\mathbf{M}_{\hat{\pi}}(y, \bar{\mu})$ such that $\det(\mathbf{M}_{\hat{\pi}}(y, \bar{\mu})) > 10^{-5}$. To reduce computational complexity, we only choose 12 degenerate policies $\bar{\mu}$ with the highest $\min_y \left[\det(\mathbf{M}_{\hat{\pi}}(y, \bar{\mu}))\right]$. To design the sub-optimal policy with a discretized belief space, as discussed in Section V-D, we use a precision of $\epsilon = 0.2$ for the probability measure, resulting in $|\bar{\Delta}_{|\mathcal{A}|}| = 56$. Here, we design two such sub-optimal policies by discretizing belief states of an aware adversary (AA) and an unaware adversary (UA).

In addition, we simulate the Best Effort Moderation (BEM) approach [19] where the controller aims to maintain a constant metered load by charging or discharging the battery based on previous load y_{k-1}, current battery state z_k and current consumption x_k. We also simulate a differential privacy (DP) mechanism with a Laplacian noise distribution given by

$$f_L(x) = \frac{exp(-|x|/b)}{2b}, \tag{39}$$

where $b = x_{\max}/\epsilon$, and $\epsilon > 0$ is a parameter which denotes level of the privacy guarantee the user desires. The lower the ϵ, the higher is the privacy due to more added noise.

Furthermore, to design the AMDPG control policy, we use an exploration probability of $\epsilon_1 = 0.03$, where the adversarial model-based noise is generated by solving an instantaneously optimal policy with relaxed constraints using $|\mathcal{Y}_{exp}| = 1$. In addition, a uniformly distributed random noise in $[0, 0.05]$ is used to generate noisy action with an exploration probability of $\epsilon_2 = 0.03$. The actor and critic neural networks are designed with 170 and 279 learnable parameters, respectively. We train the actor and critic for 2000 episodes, each containing 96 time slots. We set the discount factor ρ to 0.9 for all other policies to expedite convergence, but for AMDPG, we set it to 0.99.

The performance of the designed control policies is evaluated using an aware adversary in Monte Carlo simulations comprising 2000 episodes. Fig. 7 shows the average Bayesian reward of the aware adversary under different control policies. The results indicate that the sub-optimal policies obtained by restricting either the control sub-policy space or the belief state-space perform poorly against an informed adversary. Additionally, in this binary state system, both the instantaneously-optimal and the AMDPG control policies yield the lowest Bayesian reward among the evaluated control policies. Furthermore, we evaluate the control policies using the adversarial precision metric given by the formula:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{40}$$

Precision is useful when dealing with cases where one state is significantly more frequent than the other. In such cases, accuracy can be misleading, and precision provides a better understanding of how well the model performs for the less-frequent state. As illustrated in Fig. 8, the precision of the aware adversary follows a similar pattern as the average Bayesian reward across various HMM parameter settings, represented by λ_0 and λ_1. This result highlights the effectiveness of our proposed Bayesian approach in mitigating the precision of adversarial inference, thereby enhancing user privacy within the system.

VIII. NUMERICAL STUDY WITH REAL DATA

In this section, we present an experimental study with real data to evaluate the effectiveness of the proposed AMDPG control policy for the SBHT inference control problem. We first describe the Co-LivEn dataset, which was collected

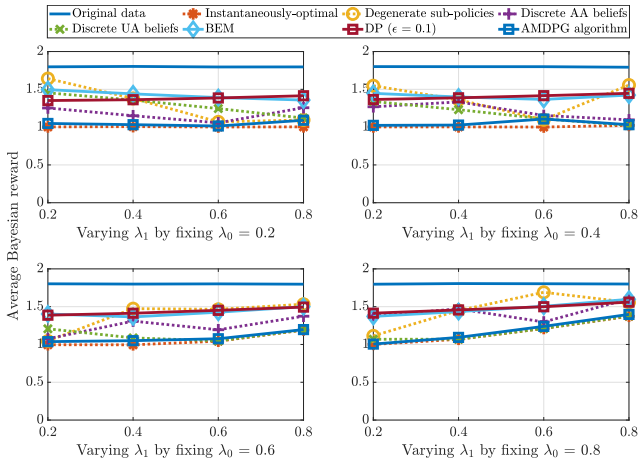

FIGURE 7. Comparison of the average Bayesian rewards of the aware adversary when using different inference control approaches.

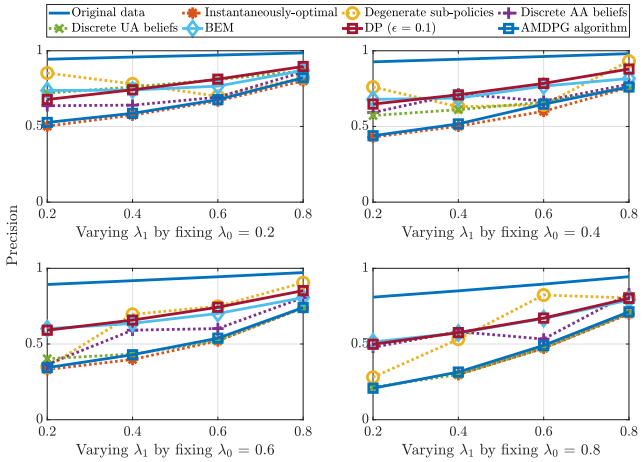

FIGURE 8. Comparison of the precision (true positive rate) of the aware adversary when using different inference control approaches.

from a multi-occupancy household with energy consumption data for a variety of appliances. Next, we evaluate the proposed AMDPG control policy on this dataset and compare its performance with that of a control policy designed against an unaware adversary.

A. CO-LIVEN DATASET

The Co-LivEn dataset used in this study is available as a public repository at https://zenodo.org/record/6480220. The dataset contains detailed electricity measurements of various appliances in a collective living (co-living) student household at KTH Live-in-Lab in Stockholm, Sweden. The household comprises four single rooms with attached bathrooms, a shared kitchen, and a common living room. The measurements include root-mean-square (RMS) voltage, RMS current, real power, and power factor of each appliance in the household. The data was collected with a sampling rate of 1 second and over a period of 277 days between August 28, 2020, and May 31, 2021. This energy dataset is unique and comes from a Nordic country, providing insights into the energy consumption patterns of students living in a shared household throughout different seasons.

The data was collected using off-the-shelf smart plugs that were connected to the sockets for each appliance. The smart plugs were equipped with Wi-Fi modules that transmitted the data wirelessly to a local server. The server stored the data in its raw format, which was then pre-processed to eliminate missing data, outliers, and noise. The dataset contains detailed electrical measurements of 32 unique appliances as shown in Table 1. A detailed visualization of the appliance usage data over a single day can be seen in Fig. 9, which is derived from the Co-LivEn dataset. It is important to note that this figure does not depict all appliances, as those with low power consumption are difficult to visualize and have been excluded for clarity. To facilitate access to the data, it has been made available in two formats. The first is a compressed file called "appliance_csv.zip," which contains the data in plain CSV file format. The second is a compressed file called "appliance_mat.zip," which contains the data in MATLAB file format. Both files are publicly accessible and can be downloaded from the repository.

The dataset is organized into folders according to the location of the appliances, such as the common living room, kitchen, and each individual room. Each location folder contains folders for each appliance, and within each appliance folder, there is a separate file for each day of data collection. This structure allows for easy navigation and selection of specific appliances and time periods of interest. The high resolution and wide range of appliances present in the co-living household energy dataset make it a valuable resource for evaluating the effectiveness of proposed control policies in a real-world setting.

B. EVALUATION OF THE AMDPG CONTROL POLICY

In this section, we present an experimental study with real data to evaluate the effectiveness of the proposed AMDPG control policy for the SBHT inference control problem. We performed numerical simulations using the Co-LivEn Dataset. Specifically, we consider a scenario where users aim to conceal their cooking activities during the daytime to prevent potential disclosure of their presence at home. To model the system, we combine the consumption of high-power consuming kitchen appliances, such as the stove and oven, and define a hypothesis state with two possible outcomes, representing whether at least one of them is on or all of them are off. The consumption from all other appliances is assumed to be independent noise. We used the first 60% of the dataset to train the HMM parameters (using the FHMM approach [29]) and the remaining 40% for evaluating the designed control policy. In the simulations, we used 5-minute time slots between 10:00-14:00 each day, resulting in a horizon length of 48. We discretized the mean power consumption data in each time slot using a $400W$ quantization and set $|\mathcal{X}| = |\mathcal{Y}| = 5$. Additionally, we consider a 48V-

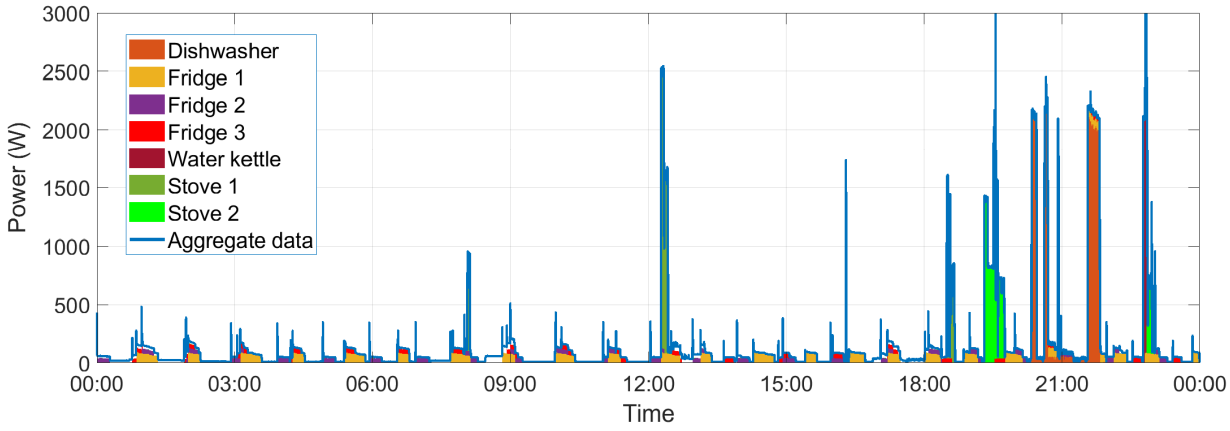

FIGURE 9. Visualization of appliance usage data over a single day, obtained from the Co-LivEn dataset.

TABLE 1. Appliance types by location.

Location	Appliance Type	Count
Common Living Room	Smart Speaker	1
	Study Lamp	1
	Vacuum Cleaner	1
	Total	3
Kitchen	Coffee Machine	1
	Dishwasher	1
	Fridge	3
	Kettle	2
	Oven	2
	Smart Speaker	1
	Stove	2
	Toaster	1
	Total	13
Room 1	Floor Lamp	1
	Hair Dryer	1
	Laptop	2
	Phone Charger	2
	Total	6
Room 2	Floor Lamp	1
	Laptop	1
	Phone Charger	1
	Total	3
Room 3	Floor Lamp	1
	Laptop	1
	Phone Charger	1
	Total	3
Room 4	Floor Lamp	1
	Hair Dryer	1
	Laptop	1
	Phone Charger	1
	Total	4
Combined Total		**32**

30Ah battery with $|\mathcal{Z}| = 35$. As a result, we have the belief state dimension as $|\mathcal{A}| = 70$ and the control sub-policy dimension as $|\mathcal{W}| = 3500$. To reduce the computational complexity, we generated parameters of a three-circuit energy storage model [3] at a higher discretization value than the storage state Z_k using Monte Carlo simulations and used them to estimate state transitions at each discrete state Z_k that fall within the high-level state. The estimated storage state transition probability $P_{Z_k|Z_{k-1},D_k}$ was used to simulate a battery in both the reinforcement learning and evaluation phases. Further, to design the AMDPG control policy, we use a discount factor of $\rho = 0.99$, an exploration probability of $\epsilon_1 = 0.03$, a random noise probability of $\epsilon_2 = 0.03$, and actor and critic neural networks with 8×10^6 and 19.5×10^6 learnable parameters, respectively. We train the actor and critic for 15000 episodes, each containing 48 time slots. In addition, we also implement the approach based on our previous work [3], where we design a policy that minimizes the discounted Bayesian reward of an adversary who is unaware of the control system's existence, using discretized belief states with a precision of 0.2. The performance of the designed control policies is evaluated using an aware adversary in Monte Carlo simulations comprising 2000 episodes, which are generated by randomly picking each episode from the available 111 episodes (40%) reserved for evaluation from the dataset. In addition, we simulate the BEM approach [19] and a differential privacy mechanism with a Laplacian noise distribution for $\epsilon \in \{0.1, 1, 10\}$.

Table 2 shows the average Bayesian reward and precision of the aware adversary when using the designed control policies. It was observed that, with original data, the adversarial precision to identify the cooking (stove and oven) state of household is 0.6. That is, when adversary makes a guess that someone is using stove or oven at the household, it is accurate around 60% of the time. By using the proposed AMDPG control policy, the precision is reduced to 0.29, which is a 52% reduction compared to the original data, demonstrating its effectiveness in reducing privacy risk. BEM and differential privacy (with $\epsilon = 0.1$) approaches also perform reasonably well in this case, with each achieving a precision of 0.4 and 0.33. In this case, although these heuristic approaches perform relatively well compared to the original data, they are not guaranteed to work as well in other cases as they operate based on pre-defined rules. In addition, we observe that when using a control policy designed against an unaware adversary, the aware adversarial precision actually increases to 0.95. This result further emphasizes the importance of employing a control policy against a worst-case adversary.

TABLE 2. Comparison of the aware adversarial performance using AMDPG control policy and the policy designed by discretizing the unaware adversarial belief state.

Control policy	Avg. Bayesian reward	Adversarial precision
Original data	1.62	0.60
Discrete UA beliefs	1.82	0.95
BEM	1.22	0.40
DP ($\epsilon = 0.1$)	1.25	0.33
DP ($\epsilon = 1$)	1.30	0.35
DP ($\epsilon = 10$)	1.58	0.59
AMDPG	1.11	0.29

In this study, we evaluated the effectiveness of the proposed control strategies against a privacy scenario related to hiding cooking patterns. Other interesting potential scenarios could be related to hiding occupancy patterns, electric vehicle ownership, usage patterns of entertainment devices such as TV, stereo etc. The MATLAB code used for computing the control policies is publicly available at https://github.com/r2avula/AdversarialInferenceControl. In this work, we use YALMIP [30], MPT3 [31], and Gurobi [32] for mathematical modeling and optimization.

IX. CONCLUSION

In this paper, we presented a Bayesian approach to control adversarial inference and address the physical-layer privacy problem in CPSs. We considered a worst-case privacy scenario, assuming an adversary with complete knowledge of the user's control strategy and modeling the adversary's inferences using SBHT. We employed the MDP framework to quantify privacy leakage in the physical layer by calculating the Bayesian risk (adversarial reward) in the SBHT.

For finite state-space problems, we derived the fixed-point Bellman's equation for an optimal stationary strategy and proposed practical optimization-based control design approaches to solve it. While these optimization-based methods can produce finite or infinite horizon optimal policies by discretizing either the belief state or sub-policy space, they are not computationally tractable for high-dimensional problems. However, they can serve as useful benchmarks for smaller-scale, toy problems. To tackle the computational complexity of exact optimal policies for high-dimensional state-space problems, we introduced the Adversarial Model-based Deterministic Policy Gradient (AMDPG) RL algorithm, providing a more practical solution for protecting privacy against adversaries with perfect knowledge of the user's control strategy in complex systems.

The numerical simulations with a toy problem demonstrate that a stronger adversary can enhance their detection performance when the control system is designed to counter weaker adversaries by acquiring knowledge of the implemented control strategy. We also found that the achievable privacy is dependent on the HMM transition probabilities, implying that some HMM systems inherently possess higher risks than others. In a binary state-space system, both the instantaneously optimal and proposed AMDPG strategies achieve the minimum Bayesian risk compared to other evaluated strategies.

Additionally, we presented the Co-LivEn dataset, a publicly available energy consumption dataset containing comprehensive electrical measurements of appliances in a co-living household. Using this dataset, we benchmarked the proposed AMDPG strategy and compared it with a control strategy designed for a controller-unaware adversary. Notably, the AMDPG control policy significantly reduced the aware adversary's precision compared to the original data, indicating its effectiveness in mitigating privacy risks. The results reveal that when using a control policy designed against an unaware adversary, not only does it fail to achieve the primary objective of minimizing adversarial performance, but it inadvertently assists the aware adversary in improving their performance relative to the original data. This further emphasizes the importance of implementing a control policy against a worst-case adversary.

In conclusion, the proposed Bayesian privacy control approach and the RL-based policy design can help mitigate privacy risks and limit information leakage in CPSs. The Co-LivEn dataset supports smart meter privacy research by offering real-world data for benchmarking and comparison of privacy-enhancing techniques. Overall, this work contributes to the advancement of privacy-enhancing techniques for CPSs, enabling the full realization of the benefits these systems provide while safeguarding user privacy.

APPENDIX A
PROOF OF LEMMA 1

Since the adversarial guess $\hat{H}_k$ follows the detection policy $P_{\hat{H}_k|\mathbf{Y}_{1:k}}$, the joint probability of $(H_k, \hat{H}_k)$ in (3) becomes

$$P_{H_k,\hat{H}_k|\mathbf{Y}_{1:k}} = P_{\hat{H}_k|\mathbf{Y}_{1:k}} P_{H_k|\mathbf{Y}_{1:k}}. \tag{41}$$

Further, using Bayes' rule, we have

$$P_{H_k|\mathbf{Y}_{1:k}} = \frac{P_{H_k,Y_k|\mathbf{Y}_{1:k-1}}}{P_{Y_k|\mathbf{Y}_{1:k-1}}}. \tag{42}$$

To compute $P_{H_k,Y_k|\mathbf{Y}_{1:k-1}}$, we use the law of total probability and the conditional independence structure of the model to obtain:

$$\begin{aligned} &P_{H_k,Y_k|\mathbf{Y}_{1:k-1}} \\ &= \textstyle\sum_{(x_k,z_{k-1},h_{k-1})} P_{X_k,H_k|H_{k-1}} \\ &\quad \times P_{Y_k|X_k,H_k,Z_{k-1},H_{k-1},\mathbf{Y}_{1:k-1}} P_{Z_{k-1},H_{k-1}|\mathbf{Y}_{1:k-1}}, \end{aligned} \tag{43}$$

Similarly, we can express $P_{Z_k,H_k|\mathbf{Y}_{1:k}}$ as:

$$P_{Z_k,H_k|\mathbf{Y}_{1:k}} = \frac{P_{Z_k,H_k,Y_k|\mathbf{Y}_{1:k-1}}}{P_{Y_k|\mathbf{Y}_{1:k-1}}}, \tag{44}$$

$$P_{Y_k|\mathbf{Y}_{1:k-1}} = \textstyle\sum_{(z_k,h_k)} P_{Z_k,H_k,Y_k|\mathbf{Y}_{1:k-1}}, \tag{45}$$

$$\begin{aligned} P_{Z_k,H_k,Y_k|\mathbf{Y}_{1:k-1}} &= \textstyle\sum_{(x_k,z_{k-1},h_{k-1})} P_{Z_k|Z_{k-1},D_k} \\ &\quad \times P_{X_k,H_k|H_{k-1}} P_{Y_k|X_k,H_k,Z_{k-1},H_{k-1},\mathbf{Y}_{1:k-1}} \\ &\quad P_{Z_{k-1},H_{k-1}|\mathbf{Y}_{1:k-1}}. \end{aligned} \tag{46}$$

Further, given a control policy $P_{Y_k|X_k,H_k,\mathbf{I}_k}$, we have

$$P_{Y_k|X_k,H_k,Z_{k-1},H_{k-1},\mathbf{Y}_{1:k-1}} = \sum_{(x_{1:k-1},a_{1:k-2})} P_{X_{k-1}|H_{k-1}} \times P_{Y_k|X_k,H_k,\mathbf{I}_k} \prod_{j=1}^{k-2} P_{Z_j|Z_{j-1},D_j} P_{X_j,H_j|H_{j-1}}. \quad (47)$$

The lemma follows directly from (44) - (46), where $M_{\hat{\pi}} \in \mathbb{R}_+^{|\mathcal{A}|\times|\mathcal{A}|}$ and $M_{\bar{\mu}} \in \mathbb{R}_+^{|\mathcal{A}|\times|\mathcal{W}|}$ are matrices whose elements are obtained by reformulating (46) into a matrix equation in terms of $\hat{\pi}_{k-1}$ and $\bar{\mu}_k$.

APPENDIX B
PROOF OF THEOREM 1

To obtain an optimal control strategy $\mu^*_{1:N}$ associated with the finite-horizon optimization problem in (11), we use Bellman's dynamic programming equation [24]:

$$v_k(\mathbf{y}_{1:k-1}) = \min_{\mu_k \in \mathcal{U}_k} \Big[r_{k|k^-}(\mathbf{y}_{1:k-1}; \zeta^*_k, \mu_{1:k}) + \rho \mathbb{E}\Big[v_{k+1}(\mathbf{Y}_{1:k}) | \mathbf{Y}_{1:k-1} = \mathbf{y}_{1:k-1} \Big] \Big], \quad (48)$$

where v_k, known as the *value function*, is the aggregate value of discounted conditional Bayesian risk from k to N due to optimal strategy $\mu^*_{k:N}$.

For a given observation sequence $\mathbf{y}_{1:k-1}$, the objective function in (48) can be expressed using (12) and (46) as:

$$r_{k|k^-}(\mathbf{y}_{1:k-1}; \zeta^*_k, \mu_{1:k}) + \rho \mathbb{E}\Big[v_{k+1}(\mathbf{Y}_{1:k}) | \mathbf{Y}_{1:k-1} = \mathbf{y}_{1:k-1} \Big] = \sum_{(\hat{h}_k, y_k, a_k)} P_{\hat{H}_k|Y_{1:k}} P_{A_k,Y_k|\mathbf{Y}_{1:k-1}} \times \Big[c(f_{\mathrm{H}}(a_k), \hat{h}_k) + \rho v_{k+1}(\mathbf{y}_{1:k}) \Big] \quad (49)$$

$$= \hat{\pi}^{\mathsf{T}}_{k-1} \mathbf{M}_c(\bar{\zeta}^*_k) \bar{\mu}_k + \rho \hat{\pi}^{\mathsf{T}}_{k-1} \mathbf{M}_v(\bar{\mathbf{v}}_{k+1}) \bar{\mu}_k, \quad (50)$$

where $\bar{\zeta}^*_k$ and $\bar{\mathbf{v}}_{k+1}$ are $|\mathcal{Y}|$ dimensional vectors with elements as $[\bar{\zeta}^*_k]_{y_k} = \zeta^*(\hat{\pi}_k(y_k, \bar{\mu}_k, \hat{\pi}_{k-1}))$, $[\bar{\mathbf{v}}_{k+1}]_{y_k} = v_{k+1}(\mathbf{y}_{1:k})$; $\mathbf{M}_c$, $\mathbf{M}_v$ are $|\mathcal{A}| \times |\mathcal{W}|$ dimensional matrices whose elements are given by (46) and (49). From (48) and (50), we note that the value function v_k depends on $\mathbf{y}_{1:k-1}$ only through the variables $(\hat{\pi}_{k-1}, \bar{\mu}_k, \bar{\zeta}^*_k, \bar{\mathbf{v}}_{k+1})$. Therefore, if the optimization routine (48) is initialized using $\bar{\mathbf{v}}_{N+1}$ that only depends on $\mathbf{y}_{1:N}$ through $\hat{\pi}_N$, then at each $k \in \mathcal{K}_N$ and for any given $\mathbf{y}_{1:k-1} \in \mathcal{Y}^k$, there exists a policy $\tilde{\mu}_k \in \tilde{\mathcal{U}}$ that results in the same value v_k as an optimal policy $\mu^*_k \in \mathcal{U}_k$. Therefore, the belief state $\hat{\pi}_{k-1}$ forms a sufficient statistic of $\mathbf{y}_{1:k-1}$ to compute an optimal policy $\tilde{\mu}^*_k$ that achieves the minimum discounted Bayesian risk achievable by any $\mu_k \in \mathcal{U}_k$.

Let $n = N - k + 1$ denote the backward iteration index starting from $k = N$ for some arbitrarily large N. Using (50), the Bellman's recursive equation in (48) can be expressed in terms of the sufficient statistic $\hat{\pi}_{k-1}$ as:

$$v_n(\hat{\pi}_{k-1}) = \min_{\bar{\mu}_k \in \bar{\mathcal{U}}} \Big[\hat{\pi}^{\mathsf{T}}_{k-1} \big[\mathbf{M}_c(\bar{\zeta}^*_k) + \rho \mathbf{M}_v(\bar{\mathbf{v}}_{n-1}) \big] \bar{\mu}_k \Big], \quad (51)$$

where $[\bar{\mathbf{v}}_{n-1}]_y = v_{n-1}\big(\hat{\pi}_k(y, \bar{\mu}_k, \hat{\pi}_{k-1})\big)$.

ACKNOWLEDGMENT

The authors would like to acknowledge the valuable support and assistance provided by the KTH Live-in Lab Team in the collection and processing of the energy consumption data, which enabled the creation of the Co-LivEn dataset used in this study. Their contributions are greatly appreciated. They also extend their appreciation to the residents of the Live-in Lab for their cooperation in the data collection process. They also acknowledge that this article includes content that has been adapted from the Ramana R. Avula's Ph.D. thesis [33], submitted to the KTH Royal Institute of Technology, in May 2023. Ramana R. Avula was with the Division of Intelligent Systems, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden, where the majority of the work was done.

REFERENCES

[1] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, Dec. 2012.

[2] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sens. Syst. Energy-Efficiency Build.*, 2010, pp. 61–66.

[3] R. R. Avula, J.-X. Chin, T. J. Oechtering, G. Hug, and D. Månsson, "Design framework for privacy-aware demand-side management with realistic energy storage model," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3503–3513, Jul. 2021.

[4] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Proc. 6th Int. Int. Workshop Secur. Trust Manage.*, Athens, Greece. Cham, Switzerland: Springer, 2010, pp. 226–238.

[5] M. A. Mustafa, S. Cleemput, A. Aly, and A. Abidin, "A secure and privacy-preserving protocol for smart metering operational data collection," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6481–6490, Nov. 2019.

[6] M. Jawurek, M. Johns, and F. Kerschbaum, "Plug-in privacy for smart metering billing," in *Proc. Int. Symp. Privacy Enhancing Technol. Symp.* Cham, Switzerland: Springer, 2011, pp. 192–210.

[7] G. Ács and C. Castelluccia, "I have a DREAM! (differentially private smart metering)," in *Information Hiding*, vol. 6958. Berlin, Germany: Springer, 2011, pp. 118–132.

[8] H.-Y. Tran, J. Hu, and H. R. Pota, "Smart meter data obfuscation with a hybrid privacy-preserving data publishing scheme without a trusted third party," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16080–16095, Sep. 2022.

[9] M. Backes and S. Meiser, "Differentially private smart metering with battery recharging," in *Data Privacy Management and Autonomous Spontaneous Security*. Berlin, Germany: Springer, 2014, pp. 194–212.

[10] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[11] J. Gómez-Vilardebó and D. Gündüz, "Smart meter privacy for multiple users in the presence of an alternative energy source," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 1, pp. 132–141, Jan. 2015.

[12] G. Giaconi, D. Gündüz, and H. V. Poor, "Smart meter privacy with renewable energy and an energy storage device," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 129–142, Jan. 2018.

[13] J. Liao, L. Sankar, V. Y. F. Tan, and F. du Pin Calmon, "Hypothesis testing under mutual information privacy constraints in the high privacy regime," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1058–1071, Apr. 2018.

[14] J. Yao and P. Venkitasubramaniam, "On the privacy-cost tradeoff of an in-home power storage mechanism," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2013, pp. 115–122.

[15] Z. Li and T. J. Oechtering, "Privacy-aware distributed Bayesian detection," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1345–1357, Oct. 2015.

[16] Z. Li, T. J. Oechtering, and M. Skoglund, "Privacy-preserving energy flow control in smart grids," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2194–2198.

[17] Z. Li, T. J. Oechtering, and D. Gündüz, "Privacy against a hypothesis testing adversary," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1567–1581, Jun. 2019.

[18] S. Salehkalaibar, F. Aminifar, and M. Shahidehpour, "Hypothesis testing for privacy of smart meters with side information," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2059–2067, Mar. 2019.
[19] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Oct. 2010, pp. 232–237.
[20] W. Yang, N. Li, Y. Qi, W. Qardaji, S. McLaughlin, and P. McDaniel, "Minimizing private data disclosures in the smart grid," in *Proc. ACM Conf. Comput. Commun. Secur.* Raleigh, NC, USA: ACM, Oct. 2012, pp. 415–427.
[21] I. Grondman, L. Busoniu, G. A. D. Lopes, and R. Babuska, "A survey of actor-critic reinforcement learning: Standard and natural policy gradients," *IEEE Trans. Syst., Man, Cybern., C Appl. Rev.*, vol. 42, no. 6, pp. 1291–1307, Nov. 2012.
[22] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.
[23] P. K. Varshney, *Distributed Detection and Data Fusion*. Berlin, Germany: Springer, 2012.
[24] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[25] V. Pata, *Fixed Point Theorems and Applications*, vol. 116. Berlin, Germany: Springer, 2019.
[26] E. J. Sondik, "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs," *Oper. Res.*, vol. 26, no. 2, pp. 282–304, 1978.
[27] W. S. Lovejoy, "Computationally feasible bounds for partially observed Markov decision processes," *Oper. Res.*, vol. 39, no. 1, pp. 162–175, Feb. 1991.
[28] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[29] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. 15th Int. Conf. Artif. Intell. Statist.*, 2012, pp. 1472–1482.
[30] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Conf. Robot. Autom.*, Taipei, Taiwan, Sep. 2004, pp. 284–289.
[31] M. Herceg, M. Kvasnica, C. N. Jones, and M. Morari, "Multi-parametric toolbox 3.0," in *Proc. Eur. Control Conf. (ECC)*, Zürich, Switzerland, Jul. 2013, pp. 502–510. [Online]. Available: http://control.ee.ethz.ch/~mpt
[32] Gurobi Optimization, LLC. (2023). *Gurobi Optimizer Reference Manual*. [Online]. Available: https://www.gurobi.com
[33] R. R. Avula, "Towards realistic smart meter privacy against Bayesian inference," Ph.D. dissertation, KTH Roy. Inst. Technol., Stockholm, Sweden, 2023.

RAMANA R. AVULA received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology, Madras, India, in 2015, and the Ph.D. degree in electrical engineering from the KTH Royal Institute of Technology, Sweden, in 2023. From 2015 to 2017, he was a Senior Engineer with Robert Bosch Engineering and Business Solutions, India. Currently, he is a Researcher with the Department of Electrification and Reliability, RISE Research Institutes of Sweden. His research interests include statistical signal processing and cybersecurity.

TOBIAS J. OECHTERING (Senior Member, IEEE) received the Dipl.-Ing. degree in electrical engineering and information technology from RWTH Aachen University, Germany, in 2002, and the Dr.-Ing. degree in electrical engineering from Technische Universität Berlin, Germany, in 2007. In 2008, he joined the KTH Royal Institute of Technology, Stockholm, Sweden, and has been a Professor, since 2018. His research interests include communication and information theory and physical layer privacy and security, including in particular smart meter privacy, statistical signal processing, and communication for networked control. In 2009, he received the "Förderpreis 2009" from the Vodafone Foundation. He has served on numerous technical program committees for IEEE sponsored conferences. He was the General Co-Chair of IEEE ITW 2019. He has been a Senior Editor of IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, since May 2020. Previously, he served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, in June 2016; and IEEE COMMUNICATIONS LETTERS, from 2012 to 2015.

DANIEL MÅNSSON received the M.Sc. and Ph.D. degrees in engineering physics from Uppsala University, Uppsala, Sweden, in 2003 and 2008, respectively, and the Docent degree in electrical engineering, in 2016. He is a Professor with the KTH Royal Institute of Technology, Stockholm, Sweden, working in smart electricity grids and power system components. His current research is mainly focused on different aspects of energy storage systems and their implementation and ability to deliver services to prosumers and the grids.