Towards a Theory of Free-Lunch Privacy in Cyber-Physical Systems

Ruoxi Jia, Roy Dong, Prashanth Ganesh, Shankar Sastry, Costas Spanos

Abstract— **Emerging cyber-physical systems (CPS) often require collecting end users' data to support data-informed decision making processes. There has been a long-standing argument as to the tradeoff between privacy and data utility. In this paper, we adopt a multiparametric programming approach to rigorously study conditions under which data utility has to be sacrificed to protect privacy and situations where free-lunch privacy can be achieved, i.e., data can be concealed without hurting the optimality of the decision making underlying the CPS. We formalize the concept of free-lunch privacy, and establish various results on its existence, geometry, as well as efficient computation methods. We propose the free-lunch privacy mechanism, which is a pragmatic mechanism that exploits free-lunch privacy if it exists with the constant guarantee of optimal usage of data. We study the resilience of this mechanism against attacks that attempt to infer the parameter of a user's data generating process. We close the paper by a case study on occupancy-adaptive smart home temperature control to demonstrate the efficacy of the mechanism.**

I. Introduction

One of the hallmarks of CPS, ranging from smart homes, smart transportation systems, smart energy systems, to smart cities, is that data collected from individuals or entities serve an indispensable part of decision making and control underlying the systems' operation. For instance, a household's occupancy is being sensed to inform the control over lighting and heating for greater energy efficiency; taxi drivers continually share their GPS data with a central dispatch solver in order to receive automated and optimized route suggestions. The individuals and entities that engage with these CPS infrastructures would naturally wish to preserve privacy of their data.

Different CPS operations require data with varying degrees of granularity, which can generally be categorized into *aggregate-level* and *individual-level*. In the former case, the CPS operation is contingent on the statistics extracted from aggregate data of a group of individuals. A typical example is smart grid load balancing, which utilizes the agglomeration of smart meter data to forecast future demand. In these types of operations, it can be assumed that there is a database that contains data records of relevant population and the operator queries the database to acquire the information of interest to system operation. There has been fruitful research of privacy preservation in this setup. Popular privacy metrics include differential privacy [1], [2] and k-anonymity [3], whose commonality is to hide an individual's private data in an aggregate of population, also known as "hiding in the crowd".

In contrast, there are also CPS applications, such as the aforementioned smart home control and taxi dispatch examples, where service is delivered based on individual data rather than aggregate information. The access to personal records is necessitated in this type of CPS due to its nature of personalized service provision. Privacy-preserving technologies developed under the "aggregate-level" setup cannot be directly applied here, as there is essentially "no crowd in which to hide". Several privacy metrics have been studied to protect data at the individual level as opposed to at the database level. One of the oldest ones is called local differential privacy [4], [5], which provides plausible deniability by randomizing individual records. Another widely studied approach is to use information-theoretic measures to model privacy loss [6], [7], and protect privacy by adding optimally designed noise such that the privacy loss is minimized under some utility constraints.

In this paper, we focus on privacy protection at an individual level in CPS applications. Our paper is originally inspired by the observations in smart home heating system control. It is found that many of the occupancy measurements are redundant in the sense that their values do not change the optimal control of the heating system. We formalize our observations from the smart home application, and propose the concept of *free-lunch privacy*. The idea is to identify the critical data which must be reported truthfully in order to enjoy the optimal service, versus the unimportant data which can be concealed or disturbed without sacrificing control performance. Free-lunch privacy exists at a data point if it can be replaced by a falsified value without harming the utility. Free-lunch privacy is a pragmatic privacy metric since its objective is to always guarantee the optimality of service provided by CPS while reducing private data release by exploiting the possible insensitivity of CPS operation to data measurements.

The contributions of this paper are as follows: Firstly, we develop a systematic approach to characterize the utility of data for control or a decision making process that can be characterized by an optimization problem. This approach can also be applied to information-theoretic privacy framework to rigorously study the tradeoff between privacy and control performance. We formalize the free-lunch privacy and study its existence and methods to compute it. Secondly, we propose a free-lunch privacy mechanism that processes original data into the one that contains minimum private information for realizing optimal control. It is a selective data disclosure procedure that materializes the Fair Information

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley ruoxijia@berkeley.edu, roydong@eecs.berkeley.edu, prashanthcganesh108@berkeley.edu, sastry@eecs.berkeley.edu, spanos@berkeley.edu

Principles [8], and can also be treated as an adaptive local differential privacy mechanism such that the reported data conforms with local differential privacy guarantee when it is not crucial for control or decision making. We also study the implication of free-lunch privacy mechanism for adversarial inference on the random process that generates an individual's data. Lastly, we present a case study to show the use of free-lunch privacy mechanism in heating, ventilation, and air conditioning (HVAC) control in smart buildings.

The paper is organized as follows. We describe the problem formulation in Section II, and introduce the properties of free-lunch privacy in Section III. The resilience of free-lunch privacy mechanism against statistical inference attacks is elaborated in Section IV. Section V demonstrates a case study on HVAC control. Section VI concludes the paper.

II. Problem Formulation

A. General setup

We consider a service provider that requires its end user to send personal information for delivering customized service. Let the user's data be $\theta \in \Theta$. The service is considered the result of a data-dependent decision making process, characterized by the following optimization problem

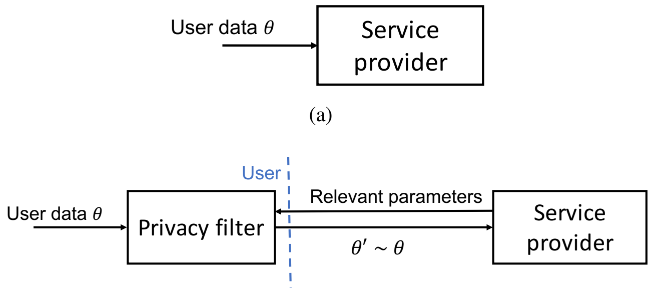
$$\begin{aligned} x^*(\theta) &= \arg\min_x J(x, \theta) \\ s.t. \quad & g(x, \theta) \leq 0 \end{aligned} \tag{1}$$

where $x \in \mathcal{X}$ is the decision variable, and the data θ can be treated as the *parameter* of the optimization problem. *User data* and *parameter* will be used interchangeably to indicate θ.

1) Example 1 (Occupancy controlled thermostat in smart home): The smart thermostat continually monitors home's occupancy via an occupancy sensor, and uses it as an input to a Model Predictive Controller (MPC) that is designed to minimize the energy cost while maintaining users' comfort [9]. Here, $\theta \in \{0, 1\}$, indicating home owner's presence/absence. x stands for the control actions such as supply air temperature. $J(x, \theta)$ is the objective function of MPC. $g(x, \theta)$ includes physical and comfort constraints in the MPC.

2) Example 2 (Taxi dispatch with real-time GPS data): Drivers' real-time GPS location data can be used for improving taxi dispatch efficiency. An optimal taxi dispatch problem is described in [10], where the objective $J(x, \theta)$ is to minimize the supply-demand mismatch and idle driving distance of all taxis, x includes dispatch order matrix and idling driving distance, θ is drivers' location information and considered to be privacy-sensitive, and $g(x, \theta)$ represents a set of operational constraints.

3) Example 3 (Integrated heat and electricity dispatch): Coordination between heat and electricity dispatch leads to a number of synergistic benefits including lower operational cost and more reliable energy supply. Often, in a large city there are many small district heating systems (DHSs) owned by different companies and only one electricity control center (ECC) managing the entire electricity system. To achieve combined heat and power, DHSs need to share with ECC the detailed information regarding the heat loads, which raises the concern of privacy [11]. In this example, θ includes information about the heat loads, x stands for the heat and electricity dispatch decision, and $J(x, \theta)$ and $g(x, \theta)$ represent the total operational cost and constraints, respectively.



(a)

(b)

Fig. 1: Diagram of (a) Traditional data sharing (b) Free-lunch privacy mechanism.

In this paper, we focus on the scenario where the required data for service provision is privacy-sensitive, as exemplified by the examples above. Our objective is to study the sensitivity of a service $x^*(\theta)$ to the variation of a data value θ. In particular, we are seeking an "equivalence set" that contains data points resulting in the same service as the true data. If the set exists, the true data can be replaced by an arbitrary point in its equivalence set without changing the service and thereby "free-lunch" privacy is achieved.

We will narrow down our scope to modeling the interaction between a single user and the service provider. In future work, we hope to extend the framework to multi-user scenario, where many interesting issues arise. For example, how does a system designer allocate privacy between users? There could be situations where only some of users can hide their parameters, and there needs to be a mechanism to decide which users get to experience "free-lunch privacy". Additionally, there is the potential for information leakage about other users through the sharing of equivalence sets.

We now formally state the definition of "equivalence set" and "free-lunch privacy".

Definition 1: For a given optimization problem in the form of (1), θ is said to be equivalent to θ', denoted by $\theta \sim \theta'$, if and only if $x^*(\theta) = x^*(\theta')$.

Definition 2 (Equivalence set): The equivalence set of $\theta \in \Theta$ is the set $[\theta] = \{\theta' \in \Theta : \theta' \sim \theta\}$

Proposition 1: The relation defined in Definition 1 is in fact an equivalence relation, i.e. it is reflexive, symmetric, and transitive. Furthermore, the collection of equivalent sets of Θ is a partition of Θ, i.e., every element of Θ belongs to one and only one equivalence set.

Definition 3 (Free-lunch privacy): The optimal decision making problem given by (1) is said to enjoy *free-lunch privacy* at θ, if $[\theta]$ is not singleton.

Proposition 1 allows us to devise a data sharing mechanism that exploits free-lunch privacy in an optimal decision making process, illustrated by Fig. 1b. Instead of directly requesting data from the user, the *free-lunch privacy mechanism* features a two-way communication as follows:

1) The service provider sends the user the equivalence partition of data space Θ, or the relevant information to compute the equivalence set of any given data
2) The user randomly selects a data point in the equivalence set of the true data and report it to the service provider

If the equivalence set of the true data is a singleton, then the user will end up reporting the true data; otherwise, the proposed mechanism can hide private data without impacting the optimal decision.

We can see that the degree of privacy protection offered by the free-lunch privacy mechanism depends on the existence of non-singleton equivalence sets. Next, we will focus on quadratic programming which has been extensively used in MPC, resource allocation, and financial applications, and study the geometry of the equivalence sets associated with the quadratic optimization problem.

B. Quadratic programming setup

Notations: If G is a matrix, then G_i denotes the ith row of G. In addition, if A is a index set, then G_A denotes the submatrix of the rows of G corresponding to the index set A. The dimension of a polyhedron $P = \{x \in \mathbb{R}^n | P^x x \leq P^c\}$ is the dimension of its affine hull and denoted by $\dim(P)$. If $\dim(P) = n$, then the polyhedron is said to be full-dimensional. The closure and interior of a set S is denoted by $\mathrm{cl}(S)$ and $\mathrm{int}(S)$, respectively.

We consider the following optimization with strictly convex quadratic objective and linear constraints,

$$\begin{aligned} J^*(\theta) = \min_x J(x, \theta) &= \frac{1}{2} x^T H x \\ \text{s.t.} \quad Gx &\leq w + S\theta \end{aligned} \tag{2}$$

where $x \in \mathbb{R}^s$, $\theta \in \mathbb{R}^n$, $G \in \mathbb{R}^{m \times s}$, $w \in \mathbb{R}^m$, and $S \in \mathbb{R}^{m \times n}$. $H \in \mathbb{R}^{s \times s}$ and $H \succ 0$. Note that the more general problem with $J(x, \theta) = \frac{1}{2} x^T H x + \theta^T F x$, where the objective and constraints are both dependent on the user data θ, can always be reformulated into the form of (2) by using the variable substitution $\tilde{x} = x + H^{-1} F^T \theta$. We denote by Θ^* the region of parameters such that (2) is feasible:

$$\Theta^* = \{\theta \in \mathbb{R}^n : \exists x \text{ satisfying } Gx \leq w + S\theta\} \tag{3}$$

Let $I = \{1, \cdots, m\}$ be the indices of constraints. In what follows, we define the key concepts for stating the main results of the paper.

Definition 4 (Optimal active set): Let x be a feasible point of (2) for a given θ. The active constraints are the constraints that satisfy $G_i x - w_i - S_i \theta = 0$. The indices of the constraints that are active at $x^*(\theta)$ is referred to as the *optimal active set*, denoted by $\mathcal{A}(\theta)$, i.e.,

$$\mathcal{A}(\theta) = \{i \in I : G_i x^*(\theta) - w_i - S_i \theta = 0\} \tag{4}$$

We also define as *weakly active constraint* an active constraint with an associated zero Lagrange multiplier, and as *strongly active constraint* an active constraint with a positive Lagrange multiplier. The *optimal inactive set* is similarly defined as

$$\mathcal{N}(\theta) = \{i \in I : G_i x^*(\theta) - w_i - S_i \theta < 0\} = I \setminus \mathcal{A}(\theta) \tag{5}$$

Definition 5 (Critical region): Given an index set $A \subseteq I$, the critical region CR_A associated with A is the set of parameters for which the optimal active set is equal to A, i.e.,

$$CR_A = \{\theta \in \Theta^* : \mathcal{A}(\theta) = A\} \tag{6}$$

Definition 6 (LICQ): We say that Linear Independence Constraint Qualification (LICQ) holds at θ if the set of constraints indexed by $\mathcal{A}(\theta)$ are linearly independent, i.e., $G_{\mathcal{A}(\theta)}$ has full row rank.

Theorem 1 ([12]): Consider the parametric quadratic programming in (2). The optimizer function $x^*(\theta) : \Theta^* \to \mathbb{R}^s$ is continuous and piecewise affine in the sense that there exists a finite set of full-dimensional polyhedral critical regions $\mathcal{CR} = \{CR_1, \cdots, CR_K\}$ such that $\Theta^* = \cup_{k=1}^K CR_k$, $\mathrm{int}(CR_i) \cap \mathrm{int}(CR_j) = \emptyset$ for all $i \neq j$ and $x^*(\theta)$ is affine inside each critical region CR_k, i.e., $x^*(\theta) = F_k \theta + g_k$, for all $k \in \{1, \cdots, K\}$.

Note that if the subscript of CR is an index, e.g., CR_i, then it will be used to denote the i-th critical region; if the subscript of CR is a set, e.g., CR_A, then it stands for the critical region whose optimal active set is A, i.e., $\mathcal{A}(\theta) = A, \forall \theta \in CR_A$.

Theorem 1 indicates that the parameter space can be partitioned into a collection of full-dimensional critical regions, and the optimizer function on each critical region is an affine function whose parameters can be completely determined given the parametric programming in (2).

The expression of critical regions and the optimizer function associated with each critical region can be obtained from the KKT conditions. The detailed derivation of F_k, g_k, and the characterization of CR_k can be found in [12].

III. Equivalence Set

A. Computation method

We start by addressing the problem of calculating the equivalence partition of the parameter space, with which the user can hide their original data while enjoying the same service as when reporting the truth. To do that, an algorithm is needed to compute the equivalence set of any given parameter.

Instead of exhaustively searching over the entire parameter space and checking the equivalence between every parameter and the given parameter, Theorem 1 allows us to search the parameter space in a region-by-region manner and directly solve for the equivalence set constrained to each critical region.

Theorem 2: Given any $\theta^* \in \Theta^*$, let CR_{k^*} be be critical region whose closure contains θ^*. Then,

$$[\theta^*] = \cup_{k=1}^K (CR_k \cap \mathcal{P}_k) \tag{7}$$

where

$$\mathcal{P}_k = \{\theta : F_k \theta = F_{k^*} \theta^* + g_{k^*} - g_k\} \tag{8}$$

Theorem 2 leads to Algorithm 1, which inspects every critical region in the parameter space, and calculates the intersection of each critical region with a linear subspace that achieves the same optimizer as the one at the true parameter θ^*.

Algorithm 1: Equivalence set computation

```
Input  : Parameter $\theta^*$, critical region $CR_k$ and
         associated optimizer function parametrized by
         $F_k$ and $g_k$ $(k = 1, \cdots, K)$
Output: Equivalence set $[\theta^*]$
Find $k^*$ such that $CR_{k^*}$ contains $\theta^*$;
$\mathcal{S} \leftarrow \emptyset$;
for $k = 1, \cdots, K$ do
    $\mathcal{P}_k \leftarrow \{\theta : F_k\theta = F_{k^*}\theta^* + g_{k^*} - g_k\}$;
    $\mathcal{S} \leftarrow \mathcal{S} \cup (CR_k \cap \mathcal{P}_k)$;
end
```

Next, we present some sufficient conditions that can be used to exclude the critical regions that do not contain any equivalent parameters to a given parameter without the need to explicitly compute the intersection.

Theorem 3: Assume that A and B are the optimal active sets associated with two full dimensional critical regions CR_A and CR_B and that LICQ holds on A and B, i.e., G_A and G_B both have full row rank. Moreover, assume that there are no constraints which are weakly active at the optimizer $x^*(\theta)$ for all θ in $\text{int}(CR_A)$ or $\text{int}(CR_B)$. Let $U = A \cup B$. If G_U has linear independent rows, then $x^*(\theta_1) \neq x^*(\theta_2)$ for all $\theta_1 \in \text{int}(CR_A)$ and $\theta_2 \in \text{int}(CR_B)$.

Proof: The KKT conditions for (2) are:

$$Hx^* + G^T u^* = 0, u^* \in \mathbb{R}^m \tag{9}$$

$$u_i^*(G_i x^* - w_i - S_i\theta) = 0 \tag{10}$$

$$Gx^* - w - S\theta \leq 0 \tag{11}$$

$$u_i^* \geq 0, \forall i \in \{1, \cdots, m\} \tag{12}$$

It follows that $x^* = -H^{-1}G^T u^*$. Since for inactive constraints $u_i^* = 0$, we have

$$x^* = -H^{-1}G_A^T u_A^*, \text{ on } CR_A \tag{13}$$

$$x^* = -H^{-1}G_B^T u_B^*, \text{ on } CR_B \tag{14}$$

We now prove the result by contradiction. Assume that there exist $\theta_1 \in \text{int}(CR_A)$ and $\theta_2 \in \text{int}(CR_B)$ such that $x^*(\theta_1) = x^*(\theta_2)$. Since H^{-1} is invertible and therefore a bijective mapping, we have $G_A^T u_A^* = G_B^T u_B^*$, i.e.,

$$\begin{aligned}\sum_{j \in A \cap B} u_{A,j}^* G_j^T + \sum_{j \in A \backslash B} u_{A,j}^* G_j^T \\ = \sum_{j \in A \cap B} u_{B,j}^* G_j^T + \sum_{j \in B \backslash A} u_{B,j}^* G_j^T\end{aligned} \tag{15}$$

which implies $u_{A,j}^* = 0$ for $j \in A \backslash B$ and $u_{B,j}^* = 0$ for $j \in B \backslash A$, which contradicts the assumption that there are no weakly active constraints in $\text{int}(CR_A)$ and $\text{int}(CR_B)$. ■

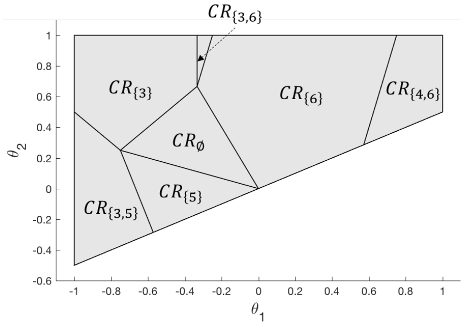


Fig. 2: Critical regions for Example 1. [12]

Corollary 1: Consider the same assumptions as in Theorem 3, if G_B can be obtained by adding or deleting rows from G_A, i.e., $A \subset B$ or $B \subset A$, then $x^*(\theta_1) \neq x^*(\theta_2)$, $\forall \theta_1 \in \text{int}(CR_A), \theta_2 \in \text{int}(CR_B)$.

Lemma 1 ([13], [14]): Let A be a given optimal set for some $\theta \in \Theta^*$. Assume that (1) LICQ holds for A, (2) there are no coinciding inequalities for facet of $\text{cl}(CR_A)$, and (3) there are no weakly active constraints at $x^*(\theta)$ for all $\theta \in \text{cl}(CR_A)$. Then, the optimal active set of any given adjacent critical region of CR_A only differs A in one element.

Corollary 2: For any two adjacent critical regions satisfying the assumptions in Lemma 1, denoted by CR_1 and CR_2, we have $x^*(\theta_1) \neq x^*(\theta_2)$, $\forall \theta_1 \in \text{int}(CR_1), \theta_2 \in \text{int}(CR_2)$.

Example 1: Consider the problem [12] where

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{16}$$

$$G^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{bmatrix} \tag{17}$$

$$S^T = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -2 \\ 1 & -1 & -1 & 1 & 3 & -1 \end{bmatrix} \tag{18}$$

$$w^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \tag{19}$$

and $-1 \leq \theta_1, \theta_2 \leq 1$. The critical regions and corresponding optimal active sets are depicted in Fig. 2. For instance, we are interested in computing the equivalence set of $\theta = [-0.4, 0]^T$, which resides in $CR_{\{6\}}$. Corollary 1 indicates that critical region $CR_\emptyset$, $CR_{\{3,6\}}$, $CR_{\{4,6\}}$ can be excluded from the search as their corresponding optimal active sets either include or are contained by $\{6\}$.

B. Existence of non-singleton equivalence set

We now consider the question: under what conditions is free-lunch privacy guaranteed to exist at a given parameter? According to the definition of free-lunch privacy, this is equivalent to examining when the equivalence set of a parameter is a non-singleton. We will first focus on the geometry of the equivalence set constrained to the critical set that contains the given parameter, because (1) this critical region is the only one that is always guaranteed to include equivalent points to the given parameter among all critical regions (because any parameter is equivalent to itself); (2) the existence of a non-singleton equivalence set on this critical region serve as a sufficient condition for the existence on the entire parameter space.

Definition 7 (Constrained equivalence set): For a given parameter θ^*, let the critical region containing θ^* be CR_{k^*}. The constrained equivalence set of θ^* is the set $[\theta^*]_c = \{\theta : \theta \in [\theta^*], \theta \in CR_{k^*}\}$.

Lemma 2: Given a parameter θ^*, assume that $\theta^* \in \mathrm{int}(CR_{k^*})$. Then, $\dim([\theta^*]_c) = \dim(\ker(F_{k^*}))$.

Proof: By Theorem 1, $[\theta^*]_c$ is the intersection of CR_{k^*} and a linear subspace characterized by

$$\{\theta : F_{k^*}\theta = F_{k^*}\theta^*\} = \theta^* + \ker(F_{k^*}) := L \qquad (20)$$

where $\ker(\cdot)$ denotes the kernel of a matrix, i.e., $\ker(F_{k^*}) = \{x : F_{k^*}x = 0\}$. It is clear that $\dim([\theta^*]_c) \leq \dim(\ker(F_{k^*}))$. We next prove $\dim([\theta^*]_c) \geq \dim(\ker(F_{k^*}))$ to obtain the theorem. By the assumption that $\theta^* \in \mathrm{int}(CR_{k^*})$, there exists a full-dimensional ball centered at θ^* with some radius r_1 that can be fitted in $\mathrm{int}(CR_{k^*})$. Let the ball be denoted by $B_1 = \{\theta \in \mathbb{R}^n : \|\theta - \theta^*\|_2 \leq r_1\}$. Because θ^* also belongs to L, there exists another ball B_2 centered at θ^* with radius $r_2 < r_1$ on L, i.e., $B_2 = \{\theta \in L : \|\theta - \theta^*\|_2 \leq r_2\}$. Since $B_2 \subset B_1$, any θ in B_2 also belongs to $(B_1 \cap B_2) \subset [\theta^*]_c$. This implies that a ball with dimension same as L can be fitted in $[\theta^*]_c$. Therefore, we can conclude that $\dim([\theta^*]_c) = \dim(L) = \dim(\ker(F_{k^*}))$. ■

Using Lemma 2, we can prove the following theorem that states the sufficient conditions for the existence of free-lunch privacy at a given parameter.

Theorem 4: Given θ^* and assume $\theta^* \in \mathrm{int}(CR_{k^*})$. Let A be the optimal active set at θ^*. It is ensured that $[\theta^*]$ is non-singleton if any of the following conditions is met:

- Case (1): No constraints are active at θ^*, i.e., $A = \emptyset$;
- Case (2): The number of optimization variable is less than the number of parameters, i.e., $s < n$;
- Case (3): The rank of G_A is less than the number of parameters, i.e., $\mathrm{rank}(G_A) < n$; particularly, if LICQ holds then this condition reduces to checking if the number of active constraints at θ^* is less than the number of parameters, i.e., $|A| < n$.

Proof: We consider the optimizer function in three different situations.

1): $A = \emptyset$. Since for inactive constraints $u^*_{I\backslash A} = 0$, we have $x^*(\theta) = -H^{-1}G^T u^* = -H^{-1}G^T u^*_{I\backslash A} = 0$. Therefore, free-lunch privacy holds on the entire critical region CR_A.

2): $A \neq \emptyset$ and LICQ holds. By [12],

$$F_{k^*} = H^{-1}G_A^T(G_A H^{-1} G_A^T)^{-1} S_A \qquad (21)$$

which follows that $\mathrm{rank}(F_{k^*}) \leq \min\{|A|, s, n\}$. In addition, by the rank-nullity theorem we have

$$\dim(\ker(F_{k^*})) = n - \mathrm{rank}(F_{k^*}) \qquad (22)$$

Therefore, if $s < n$ or $|A| < n$, then $\dim(\ker(F_{k^*})) \geq 1$ which implies free-lunch privacy at θ^* by Lemma 2.

3): $A \neq \emptyset$ and LICQ does not hold. Let $l = \mathrm{rank}(G_A)$. Again, using the results in [12] we have

$$F_{k^*} = H^{-1}G_{A,1}^T U_1^{-1} P \qquad (23)$$

where $G_{A,1} \in \mathbb{R}^{s\times l}$, $U_1 \in \mathbb{R}^{l\times l}$, $P \in \mathbb{R}^{l\times n}$. Therefore, we have $\mathrm{rank}(F_{k^*}) \leq \min\{s, l, n\}$. If $s < n$ or $l < n$, then $\dim(\ker(F_{k^*})) \geq 1$ is always ensured by the rank-nullity theorem.

Summarizing 1)-3), we prove the theorem. ■

Example 1 (continued): We use the problem described in (16)-(19) to demonstrate the utility of Theorem 4. We consider the constrained equivalence set of the Chebyshev center of each critical region. Let the Chebyshev center of critical region CR_A be denoted by θ^*_A. First, notice that there is no constraint active at $\theta^*_\emptyset$. By Case (1) in Theorem 4, free-lunch privacy always exists at $\theta^*_\emptyset$. This is validated by Fig 3d, where every point in the critical region is equivalent to $\theta^*_\emptyset$. Second, since $\theta^*_{\{3\}}$, $\theta^*_{\{5\}}$ and $\theta^*_{\{6\}}$ all have one active constraint, and the existence of free-lunch privacy at these points is ensured by Case (3) in Theorem 4, as illustrated by Fig. 3c, 3e and 3f, respectively. For $\theta^*_{\{3,5\}}$, $\theta^*_{\{3,6\}}$, and $\theta^*_{\{4,6\}}$, the number of active constraints are equal to the number of parameters and free-lunch privacy at these parameters are not guaranteed. As shown in Fig. 3b, 3g, 3h, the constrained equivalence set associated with each of these parameters collapses to a singleton.

IV. Implications for Statistical Estimation

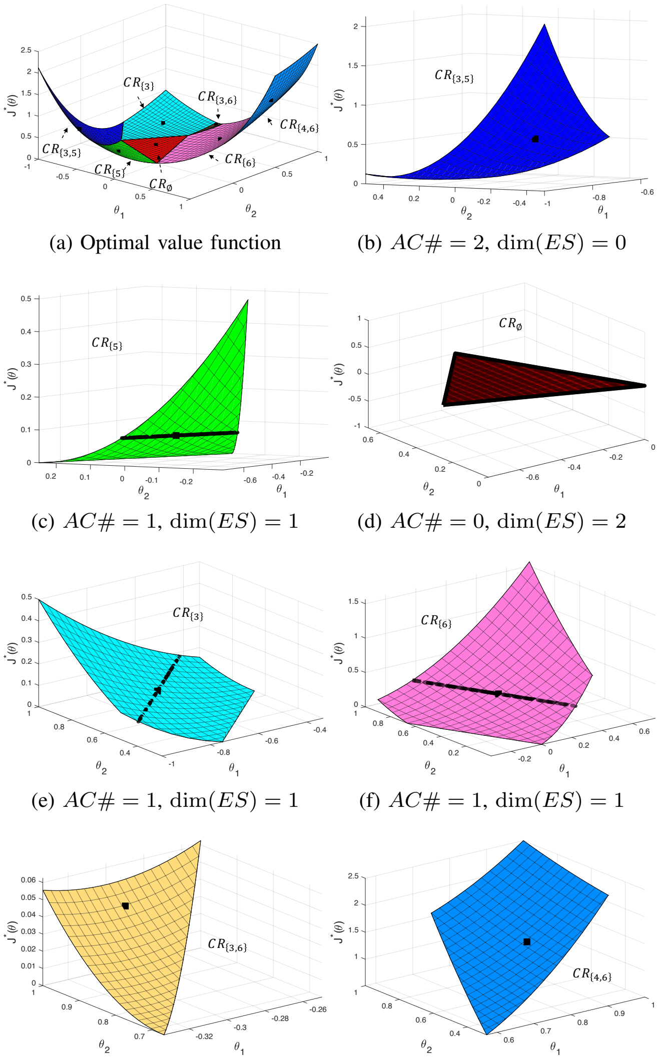
The analysis in the preceding section implies that free-lunch privacy may not always exist. If the user's data is shared sequentially, then there will be some data points which can be hidden without sacrificing the performance while sometimes the truthful data must be reported in order to maintain the optimality of a decision making process.

Assume that a user shares his/her data multiple times (continually or intermittently) to acquire the service. We consider two types of adversaries:

- A weak adversary, who can eavesdrop the communication link from the user to service provider and can intercept any data shared by the user;
- A strong adversary, who has access to not only the user data but the optimization problem that the service provider solves in order to offer the service.

Both adversaries are interested in learning the parameter of the underlying random process that generates the user's data. We further assume that users' data is exchanged via a simplified version of the free-lunch privacy mechanism described in II-A. The simplified mechanism only calculates the constrained equivalence set of a given data point in order to improve computational efficiency. More specifically, in the simplified mechanism the user randomly selects a data point in the constrained equivalence set of the truthful data and reports it to the service provider.

In this section, we will use a simple example to demonstrate that the free-lunch privacy mechanism can protect the process that generates a user's data against adversarial inference.



(a) Optimal value function

(b) $AC\# = 2$, $\dim(ES) = 0$

(c) $AC\# = 1$, $\dim(ES) = 1$

(d) $AC\# = 0$, $\dim(ES) = 2$

(e) $AC\# = 1$, $\dim(ES) = 1$

(f) $AC\# = 1$, $\dim(ES) = 1$

(g) $AC\# = 2$, $\dim(ES) = 0$

(h) $AC\# = 2$, $\dim(ES) = 0$

Fig. 3: (a) illustrates the optimal value function on each critical region; (b)-(h) demonstrates the constrained equivalence set of the chebyshev center of each critical region. The constrained equivalence set is shown in black line. $AC\#$ stands for the number of active constraints, and $\dim(ES)$ represents the dimension of the constrained equivalence set.

Assume the user's data is generated from an identically distributed and independent (i.i.d.) one-dimensional Gaussian source with mean μ and a known variance σ^2, i.e., $\theta_t \sim \mathcal{N}(\mu, \sigma^2)$, where θ_t represents t-th data sharing. The adversaries are interested in inferring μ from data observations under the simplified free-lunch privacy mechanism.

We first present a theorem that describes the geometry of constrained equivalence sets in the one-dimensional case, which helps us model the data reported under the privacy mechanism.

Lemma 3 ([11]): The optimal value function $J^*(\theta) : \Theta^* \to \mathbb{R}$ is a continuous, convex, and piecewise quadratic function.

Theorem 5: For the optimization problem in (2) with one-dimensional parameter space (i.e., $\theta \in \mathbb{R}$), if there exists a region where the optimal value function is constant, there is only one such region.

Proof: The proof simply follows from the convexity of $J^*(\theta)$. ■

Corollary 3: Consider $\theta \in \mathbb{R}$ for the optimization problem in (2). If there exists a region where the optimal value function is constant and the assumptions in Lemma 1 hold on the region, then all parameters in this region belong to the same critical region.

Proof: We use the contradiction technique for the proof. Assume that this constant region can be partitioned by multiple critical regions. We consider two neighboring ones and denote them by CR_i and CR_j, respectively. By Corollary 2, we know that for any $\theta_i \in \text{int}(CR_i)$ and $\theta_j \in \text{int}(CR_j)$, $x^*(\theta_i) \neq x^*(\theta_j)$. Using the similar technique to the proof of Theorem 3, we can also prove that $x^*(\theta_i) \neq -x^*(\theta_j)$. Therefore, $J^*(\theta_i) \neq J^*(\theta_j)$, which contradicts the fact that the optimal value function is constant on CR_i and CR_j. ■

Corollary 3 indicates that all non-singleton constrained equivalence sets correspond to the same set which can be characterized by a single closed interval. Therefore, we can model the data reported by the simplified free-lunch privacy mechanism as follows.

Let θ_t and $\tilde{\theta}_t$ denote the original and reported data, respectively. Let $[a, b]$ denote the unique constrained equivalence set. The generating process of $\tilde{\theta}_t$ can be described as:

- if $\theta_t \in [a, b]$, then

$$P(\tilde{\theta}_t = \theta) = \begin{cases} \frac{1}{a-b}, & \text{for } \theta \in [a, b] \\ 0, & \text{for } \theta \notin [a, b] \end{cases} \tag{24}$$

- if $\theta_t \notin [a, b]$, then $\tilde{\theta}_t = \theta_t$.

Note that when free-lunch privacy exists, i.e., $\theta_t \in [a, b]$, the reported data $\tilde{\theta}_t$ satisfy the local differential privacy [5] with $\epsilon = 0$ as $\frac{P(\tilde{\theta}_t | \theta_t = \theta')}{P(\tilde{\theta}_t | \theta_t = \theta)} = 1 \leq e^{\epsilon}$, $\forall \theta, \theta' \in [a, b]$.

We would like to compare the estimation of μ by some adversary with and without using the free-lunch privacy mechanism. We will assume that the weak adversary uses the sample mean to estimate μ, since, in the absence of more information, it is difficult for the adversary to design a better estimator. We will assume the strong adversary uses the maximum likelihood estimator (MLE) to estimate μ, as he has enough information to easily design this estimator.

Let $\hat{\mu}_{adv,data}$ denote the adversary's estimate, where $adv = \{w, s\}$ stands for the adversary type (weak/strong) and $data = \{o, r\}$ is the data observed by the adversary (original data/reported data via the free-lunch privacy mechanism).

Without using the free-lunch privacy mechanism, i.e., both strong and week adversaries have access to the original user data, the estimator used by both weak and strong adversary is

$$\tilde{\mu}_{adv,o} = \frac{\sum_{t=1}^{T} \theta_t}{T}, \qquad adv \in \{w, s\} \tag{25}$$

The bias and variance of the estimator are given by

$$\text{bias}[\tilde{\mu}_{adv,o}] = 0 \tag{26}$$

$$\text{var}[\tilde{\mu}_{adv,o}] = \frac{\sigma^2}{T} \tag{27}$$

Now we discuss the case where the simplified free-lunch privacy mechanism is adopted to sanitize the data. We first consider the weak adversary. Since the weak adversary only has access to the data measurements, there is no information that can help the adversary to improve the estimation. Therefore, we suppose the estimator used by the weak adversary to be

$$\tilde{\mu}_{w,r} = \frac{\sum_{t=1}^{T} \tilde{\theta}_t}{T} \tag{28}$$

Proposition 2: The systematic error of estimator, which we still call *bias*, is

$$\begin{aligned} &\text{bias}[\tilde{\mu}_{w,r}] = \mathbb{E}[\tilde{\mu}_{w,r}] - \mu \\ &= (\frac{a+b}{2} - \mu) P[\theta_t \in [a,b]] + \frac{\sigma}{\sqrt{2\pi}} (e^{-\frac{(b-\mu)^2}{2\sigma^2}} + e^{-\frac{(a-\mu)^2}{2\sigma^2}}) \end{aligned} \tag{29}$$

By Proposition 2, using the sanitized data of the free-lunch privacy mechanism the weak adversary's estimate of the mean of the user's data is always subject to some nonzero bias.

A strong adversary has access to the parameters of the optimization problem solved by the service provider and thereby it knows exactly how the reported data is generated. It can derive an MLE estimator based on its knowledge about the value of a and b.

Proposition 3: Let $A = \{i : \tilde{\theta}_i \notin [a,b]\}$. The MLE estimator for a strong adversary is

$$\tilde{\mu}_{s,r} = arg \max_{\mu} \log(W) \tag{30}$$

where

$$\log(W) = (T - |A|) \log P[\theta_t \in [a,b]] + \sum_{t \in A} \left(- \frac{(\tilde{\theta}_t - \mu)^2}{2\sigma^2} \right) \tag{31}$$

Example 2: Assume $\theta_t \sim \mathcal{N}(0,1)$, $a = -10$, $b = 0$ and $T = 1000$. We adopt numerical methods to compute the MLE estimator for the strong adversary, and compare it with the estimator used by the weak adversary. The simulation result is illustrated in Fig. 4. We can observe that the strong adversary's estimator enjoys lower bias than the weak adversary's estimator due to the additional knowledge about the optimization problem solved by the service provider.

V. Case Study: Occupancy-based HVAC Control

In this section, we demonstrate the use of the free-lunch privacy mechanism to minimize the release of private data in HVAC control. In smart buildings, the control of HVAC systems is adapted to room occupancy for the purpose of energy saving and thermal comfort. However, the occupancy data, especially in offices and households, contains rich information about space owners' habits and behaviors, and is therefore considered privacy-sensitive. We present a single-room temperature control example with MPC, and study how much occupancy information can be concealed without affecting the performance of HVAC control.

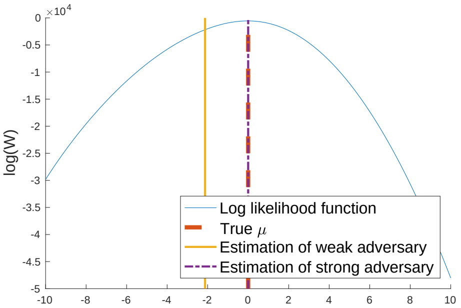


Fig. 4: Comparison of the strong and weak adversary's estimation of a user's data generating process under the free-lunch privacy mechanism. The user's data is generated from an i.i.d Gaussian process specified in Example 2.

TABLE I: Parameters used in the HVAC controller.

Param.	Meaning	Value & Units
Δt	Discretization step	$15min$
c_p	Thermal capacity of air	$1kJ/(kg \cdot K)$
M	Thermal capacity of the env.	$1000kJ/K$
c_{occ}	Thermal coefficient	$0.1kW$
$\dot{m}_z$	Supply air flow rate	$0.0382kg/s$
$T_{s,lb}$	Lower bound of supply air temperature	$5°C$
$T_{s,ub}$	Upper bound of supply air temperature	$40°C$
$T_{o,lb}$	Lower bound of comfort zone (occupied)	$21.5°C$
$T_{o,ub}$	Upper bound of comfort zone (occupied)	$27.5°C$
$T_{u,lb}$	Lower bound of comfort zone (unoccupied)	$18.5°C$
$T_{u,ub}$	Upper bound of comfort zone (unoccupied)	$36.5°C$

A. MPC model

Here, we briefly summarize the MPC model used for simulation with reference to the notations in Table I. For the detailed derivation, we refer the readers to [15], [9].

State dynamics. The continuous temperature dynamics of a zone can be derived from the law of conservation of energy,

$$M \frac{d}{dt} T = \dot{Q} + c_p \dot{m}_z (T_s - T) \tag{32}$$

which includes the zone temperature T, the thermal capacity of the zone M, the mass air flow to the zone $\dot{m}_z$, the supply air temperature T_s, and the heat capacity c_p. In this example, we consider regulating the room temperature by controlling the supply air temperature. Q represents the thermal load, which can be calculated by applying a thermal coefficient c_{occ} to the occupancy θ, i.e., $Q = \theta c_{occ}$. Here, we consider the occupancy is binary, i.e., $\theta \in \{0, 1\}$, indicating absence/presence.

Using the trapezoidal discretization, we obtain the following linear temperature dynamic model

$$\begin{aligned} (\frac{M}{\Delta t} + \frac{c_p \dot{m}_z}{2}) T(k+1) &= (\frac{M}{\Delta t} - \frac{c_p \dot{m}_z}{2}) T(k) \\ &+ c_p \dot{m}_z T_s(k) + \frac{c_{occ}(\theta(k) + \theta(k+1))}{2} \end{aligned} \tag{33}$$

Constraints. The system states and control inputs are subject to the following constraints:

C1: $T_{s,lb} \leq T_s(k) \leq T_{s,ub}$, representing the heating coil capacity;

C2: $T_{lb} \leq T(k) \leq T_{ub}$, delineating the comfort range. T_{lb} and T_{ub} take different values for different occupied status:

$$T_{lb} = \begin{cases} T_{o,lb}, & \text{if } \theta(k) = 1 \\ T_{u,lb}, & \text{if } \theta(k) = 0 \end{cases} \tag{34}$$

$$T_{ub} = \begin{cases} T_{o,ub}, & \text{if } \theta(k) = 1 \\ T_{u,ub}, & \text{if } \theta(k) = 0 \end{cases} \tag{35}$$

where $T_{o,lb}, T_{o,ub}$ represent the comfort range when a room is occupied. $T_{u,lb}, T_{u,ub}$ correspond to the lower and upper bound of room temperature when a room is not occupied. It is typically preferred that $[T_{o,lb}, T_{o,ub}] \subset [T_{u,lb}, T_{u,ub}]$ in order to save energy.

C3: $T(1) = T_{measurement}$, i.e., the measurement from temperature sensor is used for updating the initial state.

Cost function. Our objective is to minimize the energy consumption of the HVAC system, which is quantified as the l2-norm of the difference between the supply air temperature T_s and the temperature of the outside environment T_o. In addition, we would like to regulate the room temperature T around a desired temperature T_{des}. Let the MPC horizon be N, the cost function is given by

$$\min_{T \in \mathbb{R}^N, T_s \in \mathbb{R}^{N-1}} \|T_s - T_o\|_2^2 + \lambda \|T - T_{des}\|_2^2 \tag{36}$$

Furthermore, the occupancy is assumed to be constant in the optimization horizon. Note that the cost function is a quadratic function of the decision variables, and the constraints are affine in decision variables. The occupancy data (or parameter) appears as a linear term in the optimization constraints. Therefore, we can use the machinery developed in previous sections to analyze the free-lunch privacy in the occupancy-based HVAC control.

B. Simulation results

We apply the free-lunch privacy mechanism, which reports random bits or does not report anything when different occupancy status results in the same control action while reporting truthfully otherwise. The MPC horizon is fixed to be one hour. The value of other parameters are given in Table I. We assume temperature sensor measurements are given by adding a Gaussian random noise with standard deviation 0.1°C to the temperature model (33).

We simulate the HVAC system behavior for two days for different outside weather conditions and occupancy patterns. The results are illustrated in Fig. 5. Different occupancy patterns are simulated via the Markov chain with varying transition probabilities, i.e.,

$$P(\theta_{t+1}|\theta_t) = \begin{cases} q, & \text{if } \theta_{t+1} \neq \theta_t \\ 1-q, & \text{if } \theta_{t+1} = \theta_t \end{cases} \tag{37}$$

where q is referred to as *occupancy variability* in the figure. As shown in the figure, as outside ambient temperature becomes more extreme, the required number of truthful occupancy report increases. For instance, in the high-temperature scenario the change of occupancy status will dramatically change the control action. If there are no occupants, then the most energy-efficient way is to deactivate the temperature control and let the temperature drift; nevertheless, if the occupant is present in the space, the HVAC system is obligated to drive the temperature to the comfort range specified by the building code. In contrast, if the outside temperature does not deviate much from the comfort range then the occupancy measurements are actually not necessary for executing the optimal control. The free-lunch privacy mechanism differentiates the critical measurements from the "trivial" measurements that do not contribute to the control, and conceal the unimportant measurements for better privacy. As a comparison, the yellow plane demonstrates the case where the control is ex. The gap between the yellow and blue surface signals the amount of measurements that can be concealed as a free lunch. It is also interesting to notice that there is quite a bit of free-lunch privacy that can be exploited in HVAC control applications due to slow dynamics of building environment.

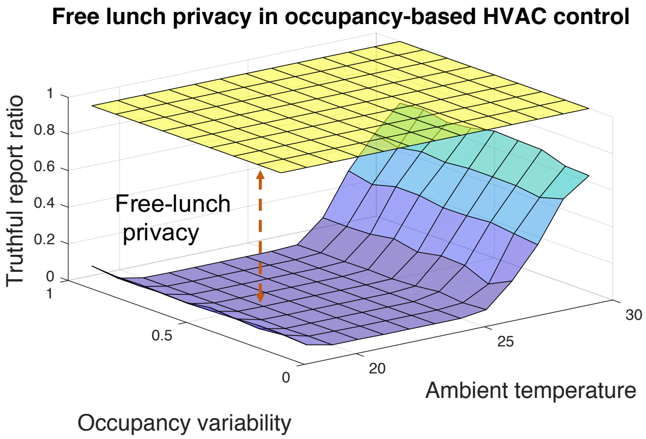


Fig. 5: Simulation of the proportion of occupancy measurements that must be reported truthfully in a typical occupancy-based HVAC control application for different weather conditions and occupancy patterns under the free-lunch privacy mechanism (blue surface). As a comparison, the case where occupancy data is reported without the free-lunch privacy mechanism is also shown here (yellow plane).

VI. Conclusion

In this paper, we present a framework to analyze free-lunch privacy in data-informed operations of CPS. Free-lunch privacy is the data ambiguity that can be allowed without affecting the optimal decision making. We present algorithms to efficiently compute free-lunch privacy given the optimization problem underlying the decision making process, as well as sufficient conditions for a quick check of the existence of free-lunch privacy. We also study the resilience of the free-lunch privacy mechanism against the adversaries that are interested in learning the process that generates a user's data. We demonstrate the use of the framework established in this paper to analyze the amount of free-lunch privacy in a smart building control application. It is shown through simulations that when the outside temperature is in a moderate range much of the occupancy measurements are redundant and therefore can be concealed to protect the user's privacy.

For future work, we plan to extend the analysis on quadratic programming presented in this paper to other nonlinear optimization problems. It is also interesting to apply the framework to some other CPS applications such as taxi dispatch and integrated energy planning.

The majority of work in privacy thus far has focused on how to allow consumers to "hide in the crowd" for large databases. In this paper, we consider ways in which a single user can modify their reported data without affecting their experienced quality of service *at all*. The free-lunch privacy framework ensures that the system will behave exactly as it would in the absence of any privacy-preserving mechanism, while still improving the privacy of users by reducing the amount of truthfully transmitted data. In this precise sense, our users can have their cake and eat it, too.

Acknowledgements

This research was funded in part by the Republic of Singapores National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program, FORCES (Foundations Of Resilient CybEr-physical Systems), National Science Foundation award numbers CNS-1238959, CNS-1238962, CNS-1239054, CNS-1239166, and the Republic of the Philippines Commission on Higher Education Philippine California Advanced Research Institutes (PCARI) award number IIID-2015-10.

References

[1] C. Dwork, "Differential privacy: A survey of results," in *International Conference on Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19.

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

[3] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[4] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[5] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*. IEEE, 2013, pp. 429–438.

[6] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1401–1408.

[7] L. Sankar, S. R. Rajagopalan, and S. Mohajer, "Smart meter privacy: A theoretical framework," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 837–846, 2013.

[8] F. T. Commission *et al.*, "Fair information practice principles," *last modified June*, vol. 25, 2007.

[9] R. Jia, R. Dong, S. S. Sastry, and C. J. Spanos, "Privacy-enhanced architecture for occupancy-based hvac control," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*. ACM, 2017, pp. 177–186.

[10] F. Miao, S. Han, S. Lin, J. A. Stankovic, D. Zhang, S. Munir, H. Huang, T. He, and G. J. Pappas, "Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 463–478, 2016.

[11] Z. Pan, Q. Guo, and H. Sun, "Feasible region method based integrated heat and electricity dispatch considering building thermal inertia," *Applied Energy*, vol. 192, pp. 395–407, 2017.

[12] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[13] J. SpjøTvold, E. C. Kerrigan, C. N. Jones, P. TøNdel, and T. A. Johansen, "On the facet-to-facet property of solutions to convex parametric quadratic programs," *Automatica*, vol. 42, no. 12, pp. 2209–2214, 2006.

[14] P. TøNdel, T. A. Johansen, and A. Bemporad, "An algorithm for multi-parametric quadratic programming and explicit mpc solutions," *Automatica*, vol. 39, no. 3, pp. 489–497, 2003.

[15] A. Beltran and A. E. Cerpa, "Optimal hvac building control with occupancy prediction," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014, pp. 168–171.