Secure Estimation Under Causative Attacks

Saurabh Sihag, *Student Member, IEEE*, and Ali Tajer, *Senior Member, IEEE*

***Abstract*— This paper considers the problem of secure parameter estimation when an estimation algorithm is prone to *causative* attacks. Causative attacks, in general, target decision-making algorithms (e.g., inference or learning algorithm) to alter their decisions in specific scenarios (e.g., distort parameter estimates for specific ranges of the parameter of interest). Such attacks influence the decisions via tampering with the mechanisms through which an algorithm acquires the statistical model of the population about which it aims to form a decision. Such attacks are viable, for instance, by contaminating the historical or training data, or by compromising an expert who provides the statistical model. In the presence of causative attacks, inference algorithms operate under a distorted statistical model for the data samples. This paper introduces a notion of secure parameter estimation and formalizes a framework under which secure estimation can be formulated and analyzed. The central premise underlying the secure estimation framework is that forming secure estimates introduces a new dimension to the estimation objective, pertaining to detecting attacks and isolating the true model. Since detection and isolation decisions themselves are imperfect, their inclusion induces an inherent coupling between the desired secure estimation objective and the auxiliary detection and isolation decisions that need to be formed in conjunction with the estimates. This paper establishes the fundamental interplay among these decisions, and characterizes the general decision rules in closed-forms for any desired estimation cost function. Furthermore, to circumvent the computational complexity associated with growing parameter dimension or attack complexity, a scalable estimation algorithm is provided, which is shown to enjoy certain optimality guarantees. Finally, the theory developed is applied to secure parameter estimation in sensor networks.**

***Index Terms*— Estimation, security, tradeoffs.**

I. Introduction

A. Motivation

STATISTICAL inference offers mechanisms for deducing the statistical properties of a population based on the data sampled from the population. Inference problems, broadly, focus on discerning the statistical model of the population or forming estimates about an unknown parameter that specifies the statistical model of the population.

The sampled data can be corrupted or compromised due to a variety of reasons such as failures in data acquisition systems or adversarial attacks. In such cases, anomaly detection constitutes a major class of inference problems the objective of which is raising alarms when the data patterns deviate significantly from their expected patterns. Effective detection of anomalies hinges on having reliable rules that can distinguish normal and abnormal patterns in the data. These rules, for instance, can be specified by an expert or by leveraging the historical data, depending on the context of the application.

While detecting anomalies in data patterns is studied extensively, the vulnerability of the *inference algorithms* to being compromised is far less-investigated. The nature of security vulnerabilities that inference algorithms are exposed to is fundamentally different from those that the data is exposed to. To highlight the distinction, note that for inferring a latent aspect of a population from sampled data (i) the inference algorithms (rules) are designed based on the data model, and (ii) the decisions (algorithms outputs) are determined based on the data (algorithm inputs). In this context, when the data (algorithm input) is compromised, the algorithm remains intact, borrowing its design from the assumed statistical model. In such situations, a measure for countering the compromised data generally involves winnowing out the compromised data samples and forming decisions based on the filtered data. In contrast, an attack on the algorithm can be exerted by providing the algorithm with an incorrect statistical model for the data. This is viable by, for instance, contaminating the historical data or by confusing the expert that produces a model, which are critical for furnishing the true model for the statistical model of the data. Therefore, when the data is compromised, an inference algorithm produces decisions based on an un-compromised known model for the data, while the data that it receives and processes is compromised. On the other hand, when the model is compromised, an inference algorithm functions based on an incorrect model for the data, in which case even un-compromised data produces unreliable decisions.

The aforementioned security vulnerabilities for the inference algorithms can be capitalized on by adversaries in order to force an inference algorithm to deviate from its optimal structure and produce decisions in ways that serve an adversary's purposes. Such attacks on decision algorithms are often referred to as **causative attacks**, through which an adversary aims to (i) make the inference algorithms oblivious to specific attacks, or (ii) degrade the performance of the inference algorithm in the presence of such an attack [1].

While the notion of secure decision-making in adjacent domains (e.g., machine learning and data mining) is heavily investigated in recent years, the fundamental limits of secure statistical inference are not well-investigated, and all the

Manuscript received December 26, 2018; revised January 9, 2020; accepted March 12, 2020. Date of publication April 6, 2020; date of current version July 14, 2020. This research was supported in part by the U. S. National Science Foundation CAREER award ECCS-1554482, and grants DMS-1737976 and ECCS-1933107. *(Corresponding author: Ali Tajer.)*

The authors are with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: tajer@ecse.rpi.edu).

Communicated by R. Balan, Associate Editor for Statistical Learning.

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

Digital Object Identifier 10.1109/TIT.2020.2985956

limited existing studies remain rather ad-hoc. In this paper, we provide a framework for secure parameter estimation under the potential presence of *causative* attacks. We establish the fundamental tradeoffs involved in decision-making under causative attacks and characterize the optimal decision rules for securely estimating the parameters and concurrently detecting the presence of the attackers. Furthermore, we provide a scalable algorithm for addressing settings in which the dimension or the size of the attacks grow, and provide optimality guarantees on the performance of this algorithm. A summary of the content and the contributions is provided in Section I-B, and the relevant literature on secure statistical inference is reviewed in Section I-C.

B. Overview and Contributions

Consider the canonical parameter estimation problem in which we have a collection of probability distributions $\{P_X : X \in \mathcal{X}\}$ defined over a common measurable space. The objective is to estimate X, which lies in a known set $\mathcal{X} \subseteq \mathbb{R}^p$, from data samples $\mathbf{Y} \triangleq [Y_1, \ldots, Y_n]$, where the sample Y_r is distributed according to P_X and lies in a known set $\mathcal{Y} \subseteq \mathbb{R}^m$. We denote the probability density functions (pdfs) that the statistician *assumes* about the underlying distributions of X and Y_r by π and $f(\cdot \mid X)$, respectively, i.e.,

$$Y_r \sim f(\cdot \mid X), \quad \text{with} \quad X \sim \pi. \tag{1}$$

For clarity in notations, we will assume that the pdfs do not have any non-zero probability masses over lower-dimensional manifolds. The objective of the statistician is formalizing a reliable estimator

$$\hat{X}(\mathbf{Y}) : \mathcal{Y}^n \mapsto \mathcal{X}. \tag{2}$$

Causative Attacks: In an adversarial environment, a malicious attacker might launch a ***causative*** attack to influence (degrade) the quality of $\hat{X}(\mathbf{Y})$. The purpose of such an attack is to compromise the *process that underlies acquiring the statistical models*. We emphasize that such an attack is different from those that aim to compromise the *data*, e.g., false data injection attacks that aim to distort the data samples $\mathbf{Y}$. Consequently, the effect of a causative attack is misleading the statistician about the true model $f(\cdot \mid X)$ that it assumes about the data. Such attacks are possible by compromising the historical (or training) data that is used for defining a model for the data. Depending on the specificity and the extent of a causative attack, e.g., the fraction of the historical or training data that is compromised, the true model $f(\cdot \mid X)$ can deviate to alternative forms, the space of which we denote by $\mathcal{F}$. The attack can affect the statistical distribution of any number of the m coordinates of $\mathbf{Y}$. There are two major aspects to selecting $\mathcal{F}$ as a viable model space.

- An attack is effective if the compromised model is sufficiently distinct from the model assumed by the statistician. Hence, even though in general $\mathcal{F}$ can be any representation of possible kernels $f(\cdot \mid X)$ mapping $\mathcal{Y}$ to $\mathbb{R}^m$, only a subset of such mappings suffices to describe the set of effective attacks.
- There exists a tradeoff between the complexity of the model space and its expressiveness. Specifically, if it is overly expressive, it can represent the possible compromised models with a more refined accuracy at the expense of having more complex inferential rules.

The specifics of the attack model will be discussed in Section II. Next, we provide a definition that is central to the proposed secure estimation framework.

(q, β)***-Security:*** The potential presence of an adversary introduces a new dimension to the estimation problem in (2). Specifically, the stochastic model of the data can be altered by an attack and detecting whether the data model is compromised becomes an additional inference task. Hence, designing an optimal estimation rule strongly hinges on successfully isolating the true model. Hence, there exists an inherent coupling between the original estimation problem of interest and the introduced auxiliary problem (i.e., detecting the presence of an attacker and isolating the true model). Based on this observation, in an adversarial setting, there exists uncertainty about the true model, based on which the quality of the estimator is expected to degrade with respect to an attack-free setting. We are interested in establishing the fundamental interplay between the quality of discerning the true model and the degradation level in the estimation quality. To establish this interplay, we say that an estimator is (q, β)***-secure*** if its estimation cost is weaker than that of the attack-free setting by a factor $q \in [1, \infty)$, while missing at most $\beta \in (0, 1]$ fraction of the attacks.[1]

In this paper, we pursue three intertwined objectives. First, we characterize the fundamental tradeoffs between q and β and delineate the associated tradeoff curve. Secondly, we characterize the inference rules in closed-forms and provide a secure estimation algorithm that achieves the optimal tradeoffs for any desired point on the tradeoff curve. Finally, to circumvent the computational complexity as the the dimension of the data, (p) grows, or the complexity of the attacks scales up (e.g., the number of coordinates compromised grows), we provide a scalable algorithm that has low computational complexity with guaranteed optimality in the asymptote of large data dimension p.

Note that in general, the specificity of the models available to the statistician has different natures in the attack-free and the attacked instances. In the attack-free setting, the model can either be fully specified, specified up to a level of uncertainty, or completely unspecified. Our model under the attack-free setting belongs to the category of fully specified models, which is apt for the scenarios when there is ample training or historical data available to learn the model. On the other hand, a partly specified model or an unspecified model are studied broadly under the domains of robust inference and non-parametric inference, respectively, which are different in scope from the parametric estimation problem studied in this paper. In the attacked setting, the specificity of the attack is characterized by the space $\mathcal{F}$. Similar to the attack-free scenarios, the space $\mathcal{F}$ may or may not be fully specified in

[1] Estimation costs and the associated estimation degradation factor q will be defined in Section II-C.

general. However, for an attack to be effective, $\mathcal{F}$ should be constrained. This follows from the fact that for the attacks to not be easily detectable, the attack models in $\mathcal{F}$ should not be too distinct from the attack-free model. At the same time, for the attacks to be successful in their objectives, the attack models in $\mathcal{F}$ must not be too similar to the attack-free model. Clearly, determining a reasonable choice for $\mathcal{F}$ hinges on the context.

C. Related Studies

The problem of secure inference is studied primarily in the context of sensor networks. The study in [2], in particular, considers parameter estimation in a two-sensor network in which one sensor is known to be secured, and one sensor is vulnerable to attacks. The objective is forming an estimate based on the mean-squared error criterion, for which a heuristic detection-driven estimator is designed. According to this design, first a decision is formed about whether the unsecured sensor is attacked. If it is deemed to be attacked, then the estimator will rely only on the secured sensor, and otherwise, it uses the data from both sensors. Unlike in [2], we consider a model with arbitrary size, assume that all data coordinates are vulnerable to the attack, and characterize the optimal decision structure, which turns out to be different from being a detection-driven design studied in [2]. Through a case study, we will also show a rather significant improvement in the estimation quality when using the optimal rules, compared to the rules specified in [2].

The adversarial setting defined in this paper is also similar to the widely-studied Byzantine attack models in sensor networks, in which the data generated by the compromised sensors are modified arbitrarily by the adversaries in order to degrade or the inference quality. An overview of the impact of Byzantine attacks on inference quality in sensor networks and relevant mitigation strategies are discussed in [3]. Detection-driven estimation strategies (i.e., when attack detection precedes the estimation routine) when the effects of the Byzantine attacks characterized by randomly flipping of the information bits generated by the sensors are discussed in [4]–[7]. Furthermore, attack-resilient target localization strategies are investigated in [4] and [8], where it is assumed that the attacker adopts a fixed strategy for maximum disruption to the inference. In these studies, an attacker may deviate from the worst-case strategy of incurring the maximum damage in order to launch a less powerful but sustainable attack, which may not be detected perfectly. Finally, strategies for isolating the compromised nodes in sensor networks are investigated in [9]–[11]. The emphasis of these studies is primarily focused on detecting attacks, or isolating the attacked sensors, which is different from the focus of our paper on parameter estimation.

The problem of secure estimation in linear *dynamical* systems has been studied extensively in the recent years [12]–[18]. The studies that are more relevant to the scope of this paper include [13], [17], and [18], which focus on the robust estimation of the states in dynamic systems. Specifically, a coding-theoretic interplay between the number of sensors compromised and the guarantees on perfect state recovery is characterized in [13], a Kalman filter-based approach for identifying the most reliable set of sensors to make an inference from is investigated in [17], and designing estimators that are robust against dynamical model uncertainty is studied in [18]. The degradation in estimation performance in a dynamical system consisting of a single sensor network is studied from the adversary's perspective in [19], where the bounds on the degradation in estimation performance with degrees of stealthiness of the attacker are characterized.

All the aforementioned studies that involve secure estimation, irrespectively of their focus or objective, conform in their design principle. They decouple the estimation decisions from all other decisions involved (e.g., attack detection or attacked sensor isolation), and produce either detection-driven estimators or estimation-driven detection routines. In the detection-driven estimation routines, an initial decision regarding the presence of an adversary (e.g., based on Neyman-Pearson theory) is followed by an optimal estimator based on the detection decision (e.g., Bayesian estimation). Such approaches to estimation implicitly assume that the detection decision has been perfect. Similarly, in an estimation-driven approach, the unknown parameter is first estimated, and then a detection decision is made (e.g., the generalized likelihood ratio test). Such approaches achieve optimality only asymptotically, i.e., when having an infinite number of samples. The premise that decoupling such intertwined estimation and detection problems into independent estimation and detection routines is sub-optimal is well-investigated [20]–[23].

Secure estimation is also related to robust estimation [24]–[29]. While sharing some assumptions (e.g., data model uncertainty), these two problems pursue two different inference objectives. Specifically, besides the estimation objective, both problems also face resolving uncertainties about the data model. Their main distinction is that they resolve model uncertainties differently, and this leads to significant differences in the formulation of the problems and the designs of the optimal decision rules. Specifically, in robust estimation, the emphasis is on forming the most reliable estimates and resolving model uncertainties is an intermediate task. Resolving model uncertainties can be carried by a wide range of approaches, spanning from averaging out the effect of the model to isolating or estimating the model. Irrespectively of how this intermediate task is performed, the ultimate interest of robust estimation is optimizing the estimation quality, and it generally does not have any regards for the quality of the decisions pertinent to resolving model uncertainty, i.e., the quality of that decision is not part of the design, and it will be dictated by the decision rules optimized for producing the best estimates. These robust methods are also often investigated under *detection-driven* estimation approaches. In contrast, in secure estimation, we are interested in the qualities of both decisions: estimating the desired parameter and detecting the unknown model. Hence, unlike robust estimation, we face *combined estimation and detection* decisions. The natural coupling between the two inference tasks is reflected in how the problem is formulated

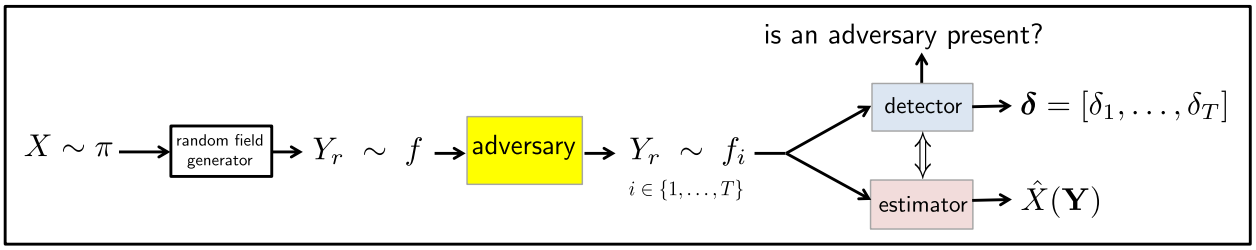

Fig. 1. The effect of the adversary on the data model, and the inferential decisions involved.

(in (17)), which requires that the decisions are optimized jointly. We remark that the secure estimation approach subsumes a relevant class of robust estimation problems (i.e., minimax robust estimation with model a discrete model uncertainty space) as its special case.

II. Data Model and Definitions

A. Attack Model

Our focus is on the canonical estimation problem in (2). The objective is to form an optimal estimate $\hat{X}(\mathbf{Y})$ (under the general cost functions specified later) in the potential presence of a causative attack. Under the attack-free setting, the data is assumed to be generated according to the known distribution

$$Y_r \sim f(\cdot \mid X), \text{ with } X \sim \pi, \quad \text{for } r \in \{1, \ldots n\}. \tag{3}$$

In an adversarial setting, an adversary, depending on its strength and preference, can launch an attack that can compromise the underlying process that the statistician uses for acquiring $f(\cdot \mid X)$. An attack will be carried out for the ultimate purpose of degrading the estimation quality of X. We assume that the adversary can corrupt the data model of *up to* $K \in \{1, \ldots, m\}$ coordinates of $\mathbf{Y}$. Hence, for a given K, there exists $T = \sum_{i=1}^{K} \binom{m}{i}$ number of attack scenarios under which the compromised data models are distinct. Define $\mathcal{S} \triangleq \{S_1, \ldots, S_T\}$ as the set of all possible combinations of attack scenarios, where $S_i \subseteq \{1, \ldots, m\}$ describes the set of coordinates the models of which are compromised under scenario $i \in \{1, \ldots, T\}$.

Under the attack scenario $i \in \{1, \ldots, T\}$, the joint distribution of Y_r deviates from f and changes to a model in the space $\mathcal{F}_i$. As discussed earlier, there exists a tradeoff between the expressiveness of this space and the complexity of the ensuing inferential rules. Specifically, a larger space $\mathcal{F}_i$ can distinguish different attack strategies with a more accurate resolution at the expense of high complexity in the analysis and the resulting decision rules. Furthermore, the model can be effective if it encompasses sufficiently distinct models. Throughout the analysis of the paper, we assume that $\mathcal{F}_i \triangleq \{f_i(\cdot \mid X)\}$, i.e., $\mathcal{F}_i$ consists of one alternative distribution. Based on this model, when the data models in the coordinates contained in S_i are compromised, the joint distribution changes from $f(\cdot \mid X)$ to $f_i(\cdot \mid X)$. It is noteworthy that the assumption that $\mathcal{F}_i$ is a singleton is for the convenience in analysis, and all the results presented can be readily generalized to any arbitrary space with countable elements. Specifically, our analysis for characterizing the optimal decision rules depends on the characteristics of the T statistical models $\{f_1, \ldots, f_T\}$. If we have multiple models per attack scenario, the only difference is that the total number of models T increases. This can potentially render characterizing the models $\{f_1, \ldots, f_T\}$ more complicated, but once these models are specified, the analysis remains intact.

Different attack scenarios might occur with different likelihoods. For instance, compromising one coordinate is easier than compromising two, and it might turn out to be more likely. To distinguish such likelihoods we adopt a Bayesian framework in which we define ϵ_0 as the prior probability of having an attack-free scenario and define ϵ_i as the prior probability of the event that the attacker compromises the model under the coordinates specified by S_i. A block diagram of the attack model and the inferential goals to be characterized, which are discussed in the remainder of this section, is depicted in Fig. 1. Finally, we define the marginal pdf of the data at coordinate $l \in \{1, \ldots, m\}$ under the attack-free setting and when the coordinate is compromised by g_l^0 and g_l^1, respectively.

B. Compound Decisions

The estimation objective constantly faces the uncertainty about whether an adversary exists. Furthermore, when one is deemed to exist, there is additional uncertainty pertaining to the number and the identity of the coordinates in which the data is being compromised. Hence, forming the estimate $\hat{X}(\mathbf{Y})$ is inherently entwined with discerning the true model of the data. Decoupling the decisions for isolating the model and estimating the parameter under the isolated model does not generally render optimal performance. In fact, there exist extensive studies on formalizing and analyzing such compound decisions, which generally aim to decouple the inferential decisions. In [21], it is shown that the generalized likelihood ratio test (GLRT) is not always optimal, and necessary and sufficient conditions for its asymptotic optimality are provided. Moreover, note that GLRT utilizes maximum likelihood estimates of unknown parameters in its decision rule and thus, it is primarily focused on the detection performance. In [20], the problem of signal detection and estimation in noise is investigated under the Bayes criterion. In [22] and [23], non-asymptotic frameworks for optimal joint detection and estimation are developed. Specifically, in [22], a binary hypothesis testing problem is investigated in which one hypothesis is composite and consists of an unknown

parameter to be estimated. In [23], the theory in [22] is extended to a composite binary hypothesis testing problem in which both hypotheses correspond to composite models. We use similar principles as established in [22] and [23] to formulate the problem of secure estimation that incorporates the estimation objective with the decision on the true model. Therefore, unlike the strategies that decouple the detection and estimation routines, our secure estimation framework is characterized by optimal estimation rules and detection rules that are formed in parallel.

C. Decision Cost Functions

1) Attack Detection Costs: The possibility of having multiple alternatives to the attack-free model renders the model detection problem as the following $(T+1)-$composite hypothesis testing problem.

$$\begin{aligned} \mathsf{H}_0 \ &: \mathbf{Y} \sim f(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \mathsf{H}_i \ &: \mathbf{Y} \sim f_i(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X), \end{aligned} \tag{4}$$

for $i \in \{1, \ldots, T\}$, where H_0 is the hypothesis corresponding to the attack-free setting, and H_i is the hypothesis corresponding to an attack launched at the coordinates in $S_i \in \mathcal{S}$. Throughout the rest of the paper we denote the attack-free data model by $f_0(\cdot \mid X)$, i.e., $f_0(\cdot \mid X) = f(\cdot \mid X)$. We define $\mathsf{D} \in \{\mathsf{H}_0, \ldots, \mathsf{H}_T\}$ as the decision on the hypothesis testing problem in (4), and $\mathsf{T} \in \{\mathsf{H}_0, \ldots, \mathsf{H}_T\}$ as the true hypothesis. We adopt a general *randomized* test $\boldsymbol{\delta}(\mathbf{Y}) \triangleq [\delta_0(\mathbf{Y}), \ldots, \delta_T(\mathbf{Y})]$ for discerning the true hypothesis, where $\delta_i(\mathbf{Y}) \in [0, 1]$ denotes the probability of deciding in favor of H_i. Clearly,

$$\sum_{i=0}^{T} \delta_i(\mathbf{Y}) = 1. \tag{5}$$

Hence, the likelihood of deciding in favor of H_j while the true model is H_i is given by

$$\mathbb{P}(\mathsf{D}=\mathsf{H}_j \,|\, \mathsf{T}=\mathsf{H}_i) = \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{6}$$

We define P_{md} as the aggregate probability of incorrectly identifying the true model under the presence of compromised coordinates, i.e.,

$$\begin{aligned} \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) &\triangleq \mathbb{P}(\mathsf{D} \neq \mathsf{T} \mid \mathsf{T} \neq \mathsf{H}_0) \\ &= \frac{1}{\mathbb{P}(\mathsf{T} \neq \mathsf{H}_0)} \sum_{i=1}^{T} \mathbb{P}(\mathsf{D} \neq \mathsf{H}_i \mid \mathsf{T} = \mathsf{H}_i) \mathbb{P}(\mathsf{T} = \mathsf{H}_i) \end{aligned} \tag{7}$$

$$= \sum_{i=1}^{T} \frac{\epsilon_i}{1-\epsilon_0} \cdot \mathbb{P}(\mathsf{D} \neq \mathsf{H}_i \mid \mathsf{T} = \mathsf{H}_i). \tag{8}$$

Furthermore, we define P_{fa} as the aggregate probability of erroneously declaring that a set of coordinates are compromised, while operating in an attack-free scenario. We have

$$\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta}) \triangleq \mathbb{P}(\mathsf{D} \neq \mathsf{H}_0 \mid \mathsf{T} = \mathsf{H}_0) = \sum_{i=1}^{T} \mathbb{P}(\mathsf{D}=\mathsf{H}_i \,|\, \mathsf{T}=\mathsf{H}_0). \tag{9}$$

2) Secure Estimation Costs: Next, we define two estimation cost functions for capturing the fidelity of the estimate $\hat{X}(\mathbf{Y})$ that we aim to form for X. For this purpose, we adopt a generic and *non-negative* cost function $\mathsf{C}(X, U(\mathbf{Y}))$ to quantify the discrepancy between the ground truth X and a generic estimator $U(\mathbf{Y})$.

Due to having distinct data models under different attack models, we consider having possibly distinct estimators under different models. We denote the estimate of X under model H_i by $\hat{X}_i(\mathbf{Y})$, and accordingly, we define

$$\hat{\mathbf{X}}(\mathbf{Y}) \triangleq [\hat{X}_0(\mathbf{Y}), \ldots, \hat{X}_T(\mathbf{Y})]. \tag{10}$$

Considering such distinct estimators, the estimation cost $\mathsf{C}(X, \hat{X}_i(\mathbf{Y}))$ is relevant only if the decision is H_i. Hence, for any generic estimator $U_i(\mathbf{Y})$ of X under model H_i, we define the *decision-specific average cost function* for $i \in \{0, \ldots, T\}$ as

$$J_i(\delta_i, U_i(\mathbf{Y})) \triangleq \mathbb{E}_i[\mathsf{C}(X, U_i(\mathbf{Y})) \mid \mathsf{D} = \mathsf{H}_i], \tag{11}$$

where the conditional expectation is with respect to X and $\mathbf{Y}$. Accordingly, we define an aggregate average estimation cost according to

$$J(\boldsymbol{\delta}, \mathbf{U}) \triangleq \max_{i \in \{0,\ldots,T\}} J_i(\delta_i, U_i(\mathbf{Y})), \tag{12}$$

where we have defined $\mathbf{U} \triangleq [U_0(\mathbf{Y}), \ldots, U_T(\mathbf{Y})]$. Finally, corresponding to the attack-free scenario, in which the only possible data model is the assumed model f, corresponding to any generic estimator $V(\mathbf{Y})$ we define the average estimation cost according to

$$J_0(V) = \mathbb{E}[\mathsf{C}(X, V(\mathbf{Y}))], \tag{13}$$

where the expectation is with respect to X and $\mathbf{Y}$ under model f. It is noteworthy that J_0 defined in (13) is fundamentally different from $J(\boldsymbol{\delta}, \mathbf{U})$ defined in (12), since the former is the estimation cost when there is no alternative to f (i.e., the attack-free scenario), while the latter is the estimation cost in an adversarial setting in which we have decided that the attacker has compromised the data. Clearly, this decision is never perfect and can be inaccurate with a non-zero probability. The role of $J_0(V)$ in our analysis is furnishing a baseline for the estimation quality in order to assess the impact of the potential presence of an adversary on the estimation quality.

Remark 1: We note that another possible choice for the estimation cost $J(\boldsymbol{\delta}, \mathbf{U})$ can be a weighted sum of the individual cost functions $J_i(\delta_i, U_i(\mathbf{Y}))$. Adopting such a cost function leads to a classical problem on combined estimation and detection, for which characterizing closed-form optimal decision rules is still an open problem. Asymptotically optimal decision rules, in the asymptote of large data size, are investigated in [20] and [30].

III. Secure Parameter Estimation

In this section, we formalize the secure estimation problem. The core premise underlying the notion of secure estimation presented is that there exists an inherent interplay between

the quality of estimating X and the quality of isolating the true model governing the data. Specifically, perfect detection of an adversary's attack model is impossible. At the same time, the estimation quality strongly relies on the successful isolation of the true data model. Lack of a perfect decision about the data model is expected to degrade the estimation quality compared to the attack-free scenario. To quantify such an interplay as well as the degradation in estimation quality with respect to the attack-free scenario, we provide the following definition.

Definition 1 (Estimation Degradation Factor): For a given estimator V in the attack-free scenario, and a secure estimation procedure specified by the detection and estimation rules $(\boldsymbol{\delta}, \mathbf{U})$ in the adversarial scenario, we define the estimation degradation factor (EDF) as

$$q(\boldsymbol{\delta}, \mathbf{U}, V) \triangleq \frac{J(\boldsymbol{\delta}, \mathbf{U})}{J_0(V)}. \tag{14}$$

Based on Definition 1, next we define the performance region, which encompasses all the pairs of decision qualities $q(\boldsymbol{\delta}, \mathbf{U}, V)$ and $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$ over the space of all possible decision rules $(\boldsymbol{\delta}, \mathbf{U}, V)$.

Definition 2 (Performance Region): We define the performance region as the region of all simultaneously achievable estimation quality $q(\boldsymbol{\delta}, \mathbf{U}, V)$ and detection performance $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$.

Remark 2: We remark that, in principle, the EDF $q(\boldsymbol{\delta}, \mathbf{U}, V)$ lies in the range $[0, +\infty)$, i.e., it can fall below 1. When $q(\boldsymbol{\delta}, \mathbf{U}, V) \in [0, 1)$, it indicates that the adversary is in fact *improving* the estimate. While theoretically viable, in reality, for an adversary to be able to launch such attacks we need to impose often unrealistic assumptions. For instance, consider the widely studied scenario of bad data injection attacks, according to which the attack-free data $\mathbf{Y}$ is related to the unknown parameter X according to

$$\mathbf{Y} = h(X) + \mathbf{N},$$

where h accounts for the sensing process, and $\mathbf{N}$ is a random variable accounting for the additive noise. When the adversary is active, the data model changes to

$$\mathbf{Y} = h(X) + \mathbf{N} + \mathbf{Z},$$

where $\mathbf{Z}$ represents the adversary's injection. For the attack model to have a better estimation performance than the nominal model, the adversary must counter or even nullify the noise. That means that the attacker should be aware of the the instantaneous realization of the noise term $\mathbf{N}$. In this paper, we focus on the attacks that adhere to the commonly studied characteristics of the adversarial behavior, where the objective of the adversary is to degrade the performance of statistical inference with respect to that in the attack-free model. Therefore, $q(\boldsymbol{\delta}, \mathbf{U}, V)$ measures the degradation in estimation quality in comparison to the attack-free scenario and we assume $q(\boldsymbol{\delta}, \mathbf{U}, V) \in [1, \infty)$.

By leveraging the characteristics of the performance region, next we define the notion of (q, β)-security, which is instrumental in defining the secure estimation problem of interest. For this purpose, note that EDF normalizes the estimation cost in the adversarial setting by that of the attack-free scenario. The two estimation cost functions involved in $q(\boldsymbol{\delta}, \mathbf{U}, V)$ can be computed independently, and as a result, determining their attendant decision rules can be carried out independently. For this purpose, we define V^* as the optimal decision rule under the attack-free setting, and J_0^* as the corresponding estimation cost, i.e.,

$$V^* \triangleq \arg\min_V J_0(V), \qquad \text{and} \qquad J_0^* \triangleq \min_V J_0(V). \tag{15}$$

Definition 3 ((q, β)-Security): An estimation procedure specified by $(\boldsymbol{\delta}, \mathbf{U}, V^*)$ for the adversarial scenario is said to be (q, β)-secure if the decision rules $(\boldsymbol{\delta}, \mathbf{U})$ yield the minimal EDF among all the decision rules corresponding to which the average rate of missing the attacks does not exceed $\beta \in (0, 1]$, i.e.,

$$q \triangleq \min_{\boldsymbol{\delta}, \mathbf{U}} q(\boldsymbol{\delta}, \mathbf{U}, V^*), \qquad \text{s.t.} \qquad \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) \leq \beta. \tag{16}$$

The performance region, and its boundary, which specifies the interplay between q and β, are illustrated in Fig. 2. Based on these definitions, we aim to characterize:

1) The region of all simultaneously achievable values of $q(\boldsymbol{\delta}, \mathbf{U}, V^*)$ and $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$, which is illustrated by the dashed region in Fig. 2.
2) The (q, β)-secure decision rules $(\boldsymbol{\delta}, \mathbf{U}, V^*)$ that solve (16), and specify the boundary of the performance region, which is illustrated by a solid line as the boundary of the performance region in Fig. 2.

By noting that $q(\boldsymbol{\delta}, \mathbf{U}, V^*) = \frac{J(\boldsymbol{\delta}, \mathbf{U})}{J_0^*}$, where J_0^* is a constant, the performance region and the (q, β)-secure decision rules are found as the solutions to

$$\mathcal{Q}(\beta) \triangleq \begin{cases} \min_{\boldsymbol{\delta}, \mathbf{U}} & J(\boldsymbol{\delta}, \mathbf{U}) \\ \text{s.t.} & \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) \leq \beta \end{cases}. \tag{17}$$

Solving $\mathcal{Q}(\beta)$ ensures that the likelihood of missing an attack is confined below β. However, it is insensitive to the rate of the false alarms that the decision rules generate. In case the statistician wishes to also control the rate of false alarms, that is the rate of erroneously declaring an attack while there is no attack, we can further extend the notion of (q, β)-security as follows.

Definition 4: An estimation procedure is (q, α, β)-secure if it is (q, β)-secure and the likelihood of false alarms does not exceed $\alpha \in (0, 1]$.

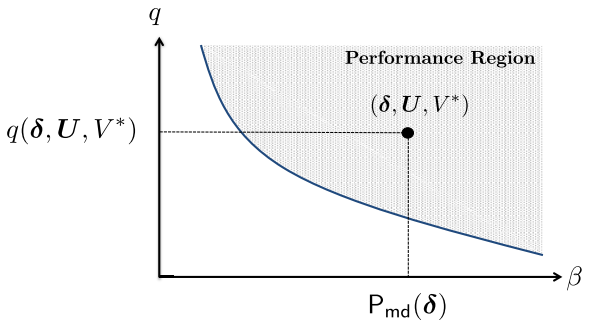

Fig. 2. Performance region.

The optimal decision rules that yield (q, α, β)-secure decisions can be found as the solution to

$$\mathcal{P}(\alpha,\beta) = \begin{cases} \min_{\boldsymbol{\delta},\mathbf{U}} & J(\boldsymbol{\delta},\mathbf{U}) \\ \text{s.t.} & \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) \leq \beta \\ & \mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta}) \leq \alpha \end{cases}. \tag{18}$$

Remark 3: It can be easily verified that $\mathcal{Q}(\beta) = \mathcal{P}(1,\beta)$.

Remark 4 (Feasibility): The probabilities $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$ and $\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta})$ cannot be made arbitrarily small simultaneously. Specifically, from the Neyman-Pearson theory [31], it can be readily verified that for any given α, there exists a value $\beta^*(\alpha)$, which specifies the smallest feasible value for β. Throughout the paper we assume that the pair (α,β) in (18) are selected such that $\mathcal{P}(\alpha,\beta)$ has a feasible solution.

We characterize the optimal solution to problems $\mathcal{P}(\alpha,\beta)$ and $\mathcal{Q}(\beta)$ in closed-forms in Section IV. Close scrutiny of the optimal decision rules indicates that the complexity of the rules grows exponentially with the dimension of X, and the number of coordinates that an adversary can compromise. We address the scalability issue in Section V. Specifically, we provide alternative low-complexity decision rules and show that despite their simple structures, they satisfy asymptotic optimality guarantees.

Remark 5 (Tradeoff): Note that the inherent tradeoff between q and β arises from how the combined inference objectives are formulated in (18). Such interplay often does not exist in other approaches that face both estimation and detection decisions. For instance, in the *detection-driven estimation* approaches, we first form a detection decision. Once the decision is made, it is treated as a perfectly correct detection decision, based on which, subsequently, the estimation decisions are formed. In such approaches, the problem is decoupled into a detection problem, the objective of which is classifying the model, followed by an optimal estimation on the decided model. In these strategies, the estimation performance is guided by the performance of the detection rules. Hence, as the quality of detection decisions improve, the quality of the estimates improves as well, indicating that the two decision qualities improve simultaneously. In a sharp contrast, in our approach we form the detection and estimation decisions in parallel in order to achieve a jointly optimal performance. Since we do not have a pre-specified order in making the two decisions, we are not observing that simultaneous improvements of the decision qualities.

IV. Secure Parameter Estimation: Optimal Decision Rules

In this section, we characterize an optimal solution to the more general problem $\mathcal{P}(\alpha,\beta)$, i.e., the estimators $\{\hat{X}_i(\mathbf{Y}) : i \in \{0,\ldots,T\}\}$ and the detectors $\{\delta_i(\mathbf{Y}) : i \in \{0,\ldots,T\}\}$. We will also specify how these decision rules can be simplified to characterize the solution to the problem $\mathcal{Q}(\beta) = \mathcal{P}(1,\beta)$. In order to proceed, we start by providing the expansion of the decision error probability terms $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$ and $\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta})$ in terms of the data models and decision rules. By noting (6) and leveraging (7), we have

$$\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) = \sum_{i=1}^{T} \frac{\epsilon_i}{1-\epsilon_0} \sum_{\substack{j=0 \\ j\neq i}}^{T} \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{19}$$

Similarly, by noting (6) and based on (9), we have

$$\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta}) = \sum_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{20}$$

By using the expansions in (19) and (20), the problem of interest in (18) can be equivalently cast as

$$\mathcal{P}(\alpha,\beta) = \begin{cases} \min\limits_{(\boldsymbol{\delta},\boldsymbol{U})} & J(\boldsymbol{\delta},\mathbf{U}) \\ \text{s.t.} & \sum\limits_{i=1}^{T} \frac{\epsilon_i}{1-\epsilon_0} \sum\limits_{\substack{j=0 \\ j\neq i}}^{T} \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \beta \\ & \sum\limits_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \alpha \end{cases}. \tag{21}$$

The roles of the estimators $\{U_i(\mathbf{Y}) : i \in \{0,\ldots,T\}\}$ appear only in the utility function $J(\boldsymbol{\delta},\mathbf{U})$. This allows for decoupling the optimization problem $\mathcal{P}(\alpha,\beta)$ into two sub-problems, as formalized in Theorem 1.

Theorem 1: The optimal secure estimators of X under different models, i.e., $\hat{\mathbf{X}} = [\hat{X}_0,\ldots,\hat{X}_T]$ is the solution to

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{U}} J(\boldsymbol{\delta},\mathbf{U}). \tag{22}$$

Furthermore, the solution of $\mathcal{P}(\alpha,\beta)$, and subsequently the design of the attack detectors, can be found by equivalently solving

$$\mathcal{P}(\alpha,\beta) = \begin{cases} \min\limits_{\boldsymbol{\delta}} & J(\boldsymbol{\delta},\hat{\mathbf{X}}) \\ \text{s.t.} & \sum\limits_{i=1}^{T} \frac{\epsilon_i}{1-\epsilon_0} \sum\limits_{j=0,j\neq i}^{T} \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \beta \\ & \sum\limits_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \alpha \end{cases}. \tag{23}$$

By leveraging the property that this theorem establishes for the optimal estimator in (22), and also taking into account the decoupled structure of the problem $\mathcal{P}(\alpha,\beta)$ in (23), in the following theorem we provide optimal designs for the secure estimators. Interestingly, it is shown that the optimal estimator under each model can be specified by optimizing a relevant cost function defined exclusively for that model.

Theorem 2 ((q,α,β)-Secure Estimators): For the optimal secure estimators $\hat{\mathbf{X}}$ we have

1) The minimizer of the estimation cost $J_i(\delta_i, U_i(\mathbf{Y}))$ is given by

$$U_i^*(\mathbf{Y}) \triangleq \arg \inf_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}), \tag{24}$$

in which $\mathsf{C}_{\mathrm{p},i}(U(\mathbf{Y}) \mid \mathbf{Y})$ is the average posterior cost function defined as

$$\mathsf{C}_{\mathrm{p},i}(U(\mathbf{Y}) \mid \mathbf{Y}) \triangleq \mathbb{E}_i\left[\mathsf{C}(X, U(\mathbf{Y})) \mid \mathbf{Y}\right], \tag{25}$$

where the conditional expectation in (25) is with respect to X under the model H_i.

2) The optimal estimator $\hat{\mathbf{X}} = [\hat{X}_0, \ldots, \hat{X}_T]$, specified in (22), is given by

$$\hat{X}_i(\mathbf{Y}) = U_i^*(\mathbf{Y}). \tag{26}$$

3) The cost function $J(\boldsymbol{\delta}, \hat{\mathbf{X}})$ is given by

$$J(\boldsymbol{\delta}, \hat{\mathbf{X}}) = \max_i \left\{ \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}} \right\}, \tag{27}$$

where we have defined

$$\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) \triangleq \inf_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}). \tag{28}$$

Proof: See Appendix A. ∎

For illustration purposes, in the next corollary, we provide the closed-forms of these decision rules when the distributions $\{f_i(\cdot \mid X) : i \in \{0, \ldots, T\}\}$ are all Gaussian.

Corollary 1 [(q, α, β)-Secure Estimators in Gaussian Models]: When the data models are Gaussian such that

$$f_i(\cdot \mid X) \sim \mathcal{N}(\theta_i, X), \qquad \text{for} \quad \theta_i \in \mathbb{R}\,, \tag{29}$$

where the mean values $\{\theta_i : i \in \{0, \ldots, T\}\}$ are distinct, and

$$X \sim \mathcal{X}^{-1}(\zeta, \phi), \tag{30}$$

where $\mathcal{X}^{-1}(\zeta, \phi)$ denotes the inverse chi-squared distribution with parameters ζ and ϕ, such that $\zeta + n > 4$, and the cost $\mathsf{C}(X, U(\mathbf{Y}))$ is

$$\mathsf{C}(X, U(\mathbf{Y})) = \|X - U(Y)\|^2, \tag{31}$$

for the optimal secure estimators $\hat{\mathbf{X}}$, we have:

1) The minimizer of the estimation cost $J(\delta_i, U_i(\mathbf{Y}))$, i.e., the estimation cost function under model H_i, is given by

$$U_i^*(\mathbf{Y}) = \frac{\zeta\phi + \sum_{r=1}^n \|Y_r - \theta_i\|_2^2}{\zeta + n - 2}. \tag{32}$$

2) The optimal estimator $\hat{\mathbf{X}} = [\hat{X}_0, \ldots, \hat{X}_T]$, specified in (22), is given by

$$\hat{X}_i(\mathbf{Y}) = U_i^*(\mathbf{Y}). \tag{33}$$

3) The cost function $J(\boldsymbol{\delta}, \hat{\mathbf{X}})$ is given by

$$J(\boldsymbol{\delta}, \hat{\mathbf{X}}) = \max_i \left\{ \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}} \right\}, \tag{34}$$

where we have

$$\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) = \frac{2\left(\zeta\phi + \sum_{r=1}^n \|Y_r - \theta_1\|^2\right)^2}{(\zeta_i + n - 2)^2(\zeta + n - 4)}. \tag{35}$$

Next, given the optimal estimators $\hat{\mathbf{X}}$, we characterize the optimal detection rules. The main observation is that even though we started by considering general randomized decision rules, these rules in their optimal forms reduce to deterministic ones. Furthermore, the decisions rules depend on the estimation costs that are computed based on the optimal estimation costs. These estimation costs make the decisions coupled. In order to proceed, we first show that problem $\mathcal{P}(\alpha, \beta)$ in (23) can be solved by leveraging the result of the following theorem, which specifies an auxiliary convex problem in a variational form.

Theorem 3: For any arbitrary $u \in \mathbb{R}_+$, we have $\mathcal{P}(\alpha, \beta) \leq u$ if and only if $\mathcal{R}(\alpha, \beta, u) \leq 0$, where we have defined

$$\mathcal{R}(\alpha, \beta, u) \triangleq \begin{cases} \min_{\boldsymbol{\delta}} & \eta \\ \text{s.t.} & \displaystyle\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) [\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u] \, \mathrm{d}\mathbf{Y} \leq \eta, \quad \forall i \\ & \displaystyle\sum_{i=1}^T \frac{\epsilon_i}{1 - \epsilon_0} \sum_{\substack{j=0 \\ j \neq i}}^T \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \beta + \eta\,. \\ & \displaystyle\sum_{i=1}^T \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \alpha + \eta \end{cases} \tag{36}$$

Furthermore, $\mathcal{R}(\alpha, \beta, u)$ is convex, and $\mathcal{R}(\alpha, \beta, u) = 0$ has a unique solution in u, which we denote by u^*.

Proof: See Appendix B. ∎

The point u^* has a pivotal role in specifying the optimal detection decision rules. We define the constants $\{\ell_i : i \in \{0, \ldots, T+2\}\}$ as the dual variables in the Lagrange function associated with the convex problem $\mathcal{R}(\alpha, \beta, u^*)$. Given u^* and $\{\ell_i : i \in \{0, \ldots, T+2\}\}$, the optimal detection rules can be characterized in closed-forms, as specified in the following theorem.

Theorem 4 ((q, α, β)-Secure Detection Rules): The optimal decision rule for isolating the compromised coordinates is given by

$$\delta_i(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{37}$$

where we have defined

$$i^* \triangleq \operatorname*{argmin}_{i \in \{0, \ldots, T\}} A_i. \tag{38}$$

Constants $\{A_0, \ldots, A_T\}$ are specified by the data models, u^*, and its associated Langrangian multipliers $\{\ell_i : i \in \{0, \ldots, T+2\}\}$. Specifically, we have

$$A_0 \triangleq \ell_0 f_0(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},0}(\mathbf{Y}) - u^*] + \ell_{T+1} \sum_{i=1}^T \frac{\epsilon_i}{1 - \epsilon_0} f_i(\mathbf{Y}), \tag{39}$$

and for $i \in \{1, \ldots, T\}$, we have

$$A_i \triangleq \ell_i f_i(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u^*] + \ell_{T+1} \sum_{j=1, j \neq i}^T \frac{\epsilon_j}{1 - \epsilon_0} f_j(\mathbf{Y}) + \ell_{T+2} f_0(\mathbf{Y}). \tag{40}$$

Proof: See Appendix C. ∎

In the next corollary, we provide the closed-forms of these decision rules when the distributions $\{f_i(\cdot \mid X) : i \in \{0,\dots,T\}\}$ are all Gaussian.

Corollary 2 ((q,α,β)-Secure Detection Rules in Gaussian Models): When the data models $\{f_i(\cdot \mid X) : i \in \{0,\dots,T\}\}$ have the following Gaussian distributions

$$f_i(\cdot \mid X) \sim \mathcal{N}(\theta_i, X), \qquad \text{for} \quad \theta_i \in \mathbb{R}, \tag{41}$$

where the mean values are distinct, and

$$X \sim \mathcal{X}^{-1}(\zeta, \phi), \tag{42}$$

the optimal decision rule for isolating the compromised coordinates is given by

$$\delta_i(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{43}$$

where we have defined

$$i^* \triangleq \underset{i\in\{0,\dots,T\}}{\operatorname{argmin}} \; A_i. \tag{44}$$

Constants $\{A_0,\dots,A_T\}$ are specified by the data models, u^*, and its associated Langrangian multipliers $\{\ell_i : i \in \{0,\dots,T+2\}\}$. Specifically, we have

$$A_0 \triangleq \ell_0 f_0(\mathbf{Y})(\mathsf{C}^*_{\mathrm{p},0}(\mathbf{Y}) - u^*) + \ell_{T+1} \sum_{i=1}^{T} \frac{\epsilon_i}{1-\epsilon_0} f_i(\mathbf{Y}), \tag{45}$$

and for $i \in \{1,\dots,T\}$ we have

$$A_i \triangleq \ell_i f_i(\mathbf{Y})(\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u^*) + \ell_{T+1} \sum_{\substack{j=1 \\ j\neq i}}^{T} \frac{\epsilon_j}{1-\epsilon_0} f_j(\mathbf{Y}) + \ell_{T+2} f_0(\mathbf{Y}). \tag{46}$$

When the cost function $\mathsf{C}(X, U(\mathbf{Y}))$ is the mean squared error cost, and $\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y})$ is evaluated using (35), we obtain

$$f_i(\mathbf{Y}) = \frac{(\zeta\phi)^{\frac{\zeta}{2}}}{(\zeta\phi + \sum_{r=1}^{n} \|Y_r - \theta_i\|^2)^{\frac{\zeta+n}{2}}} \cdot \frac{\Gamma(\zeta+n)}{2\pi^{\frac{n}{2}}\Gamma(\zeta/2)}. \tag{47}$$

By setting $T = 1$, $n = 1$, $\theta_0 = 0$, $\theta_1 = 2$, and selecting $\zeta = 4, \phi = 1$, Fig. 3 depicts the performance region and the associated (q,β)-security curve, which shows the tradeoff between the quality of the detection and the degradation in the estimation quality. It is noteworthy that this tradeoff is inherently due to the formulation of the secure estimation problem. Essentially, problem $\mathcal{P}(\alpha,\beta)$ as specified in (18), is designed to trade the quality of detection in favor of improving the estimation cost.

Based on all the decision rules specified in the section and the detailed steps of specifying the parameters involved in characterizing the decision rules, we provide Algorithm 1 to summarize all the steps for solving $\mathcal{P}(\alpha,\beta)$ for any feasible pair of α and β.

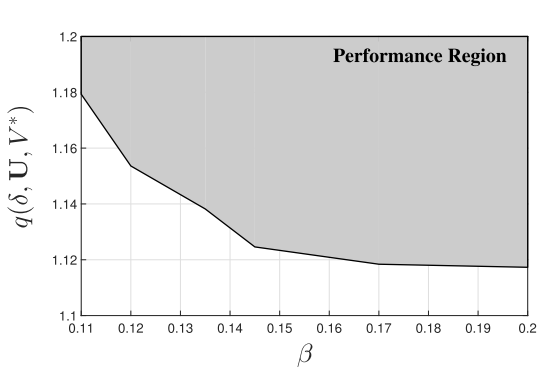

Fig. 3. Performance region for the Gaussian data model.

Algorithm 1 Solving $\mathcal{P}(\alpha,\beta)$

input α and β and evaluate $\beta^*(\alpha)$
if $\beta < \beta^*(\alpha)$ **then**
  $\mathcal{P}(\alpha,\beta)$ not feasible for given choice of α and β
  break
else
  Initialize $u_0 = 0, u_1$
  Evaluate optimal posterior estimation costs in (28)
  repeat
    $\hat{u} \leftarrow (u_0 + u_1)/2$
    for every $\hat{\boldsymbol{\ell}} \succcurlyeq 0$ in the discretized space $\|\hat{\boldsymbol{\ell}}\|_1 = 1$ **do**
      Compute $\boldsymbol{\delta}$ from Theorem 4
      Compute $M(\hat{\boldsymbol{\ell}}) \triangleq \mathcal{R}(\alpha,\beta,\hat{u})$
    end for
    if $\min_{\hat{\boldsymbol{\ell}}} M(\hat{\boldsymbol{\ell}}) \leq 0$ **then**
      $u_1 \leftarrow \hat{u}$
      $\boldsymbol{\ell} \leftarrow \hat{\boldsymbol{\ell}}$
    else
      $u_0 \leftarrow \hat{u}$
    end if
  until $u_1 - u_0 \leq \epsilon$, for ϵ sufficiently small
  $\mathcal{P}(\alpha,\beta) \leftarrow u^* = u_1$
  return Decision rules $\boldsymbol{\delta}$
end if

V. Scalable Secure Parameter Estimation

As the data dimension m grows, the number of possible data models T grows exponentially. This, subsequently, leads to an exponential growth in the complexity of forming the decision rules, e.g., the number of Lagrangian multipliers needed for characterizing the detection rules scales linearly in T. This can render Algorithm 1 computationally prohibitive. In order to circumvent the computational complexity, in this section, we design a scalable approach to secure estimation that exhibits optimality properties too. The core idea is to break down the problem into m coordinate-level problems, treat them individually, and then aggregate the individual decisions. Specifically, the decisions involve a high-level binary decision about whether the data is compromised. If the data is deemed to be compromised, then each coordinate is tested individually, and an estimate of X is formed based on the

data of that coordinate. Individual coordinate-level estimates are then tested for reliability, and combined to form an aggregate estimate for X. Since we need to perform only m single-coordinate binary detection decisions followed by forming a coordinate-level estimation routine, the computational complexity scales only linearly in the number of coordinates m, as opposed to exponentially for forming the optimal decision rules.

A. Binary Attack Detection

In the first stage, we perform a binary test to detect whether the data is compromised at all. This is carried out by solving a binary composite hypothesis testing problem given by

$$\begin{aligned} \hat{\mathsf{H}}_0 \ &: \mathbf{Y} \sim f(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \hat{\mathsf{H}}_1 \ &: \mathbf{Y} \sim \hat{f}(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X), \end{aligned} \tag{48}$$

where $\hat{\mathsf{H}}_0$ is the hypothesis corresponding to the attack-free setting, and $\hat{\mathsf{H}}_1$ signifies to the presence of an attack. Probability distribution $\hat{f}$ is a mixed distribution given by

$$\hat{f}(\mathbf{Y}) \triangleq \frac{1}{1-\epsilon_0} \sum_{i=1}^{T} \epsilon_i f_i(\mathbf{Y}). \tag{49}$$

Similarly to the optimal approach of Section IV, to design the decision rules we define the randomized test $\hat{\boldsymbol{\delta}} \triangleq [\hat{\delta}_0(\mathbf{Y}), \hat{\delta}_1(\mathbf{Y})]$, in which $\hat{\delta}_i(\mathbf{Y}) \in [0,1]$ is the probability of deciding in favor of $\hat{\mathsf{H}}_i$, and we have $\delta_0(\mathbf{Y}) + \delta_1(\mathbf{Y}) = 1$. Furthermore, define $\mathsf{D_d} \in \{\hat{\mathsf{H}}_0, \hat{\mathsf{H}}_1\}$ and $\mathsf{T_d} \in \{\hat{\mathsf{H}}_0, \hat{\mathsf{H}}_1\}$ as the decision and the true hypotheses. Hence, the false alarm rate is given by

$$\mathbb{P}(\mathsf{D_d} = \hat{\mathsf{H}}_1 \mid \mathsf{T_d} = \hat{\mathsf{H}}_0) = \int \hat{\delta}_1(\mathbf{Y}) f(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}, \tag{50}$$

and the miss-detection rate is given by

$$\mathbb{P}(\mathsf{D_d} = \hat{\mathsf{H}}_0 \mid \mathsf{T_d} = \hat{\mathsf{H}}_1) = \int \hat{\delta}_0(\mathbf{Y}) \hat{f}(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{51}$$

Since our objective is to maximize the true detection rate (or minimize (51)), in the first step, we design the decision rule $\hat{\boldsymbol{\delta}}$ using the Neyman-Pearson approach. Following the Neyman-Pearson approach, as noted in Remark 4, we have the following decision rules [31].

Theorem 5 (Attack Detection): The optimal attack detection rule that minimizes the rate of missing the attacks subject to a constraint on the rate of false alarms is given by

$$\hat{\delta}_0(\mathbf{Y}) = \mathbb{1}_{\left\{\frac{\hat{f}(\mathbf{Y})}{f(\mathbf{Y})} < \gamma\right\}} + (1-\varrho) \cdot \mathbb{1}_{\left\{\frac{\hat{f}(\mathbf{Y})}{f(\mathbf{Y})} = \gamma\right\}}, \tag{52}$$

$$\hat{\delta}_1(\mathbf{Y}) = \mathbb{1}_{\left\{\frac{\hat{f}(\mathbf{Y})}{f(\mathbf{Y})} > \gamma\right\}} + \varrho \cdot \mathbb{1}_{\left\{\frac{\hat{f}(\mathbf{Y})}{f(\mathbf{Y})} = \gamma\right\}}, \tag{53}$$

where the threshold γ and the probability term ϱ are chosen such that the false alarm constraint is satisfied with equality.

B. Isolating Compromised Coordinates

The outcome of the binary detection rule in Section V-A is deciding whether there exists an attack in the data. If an attack is deemed to exist, in the second step, we carry out an isolation decision, the role of which is identifying the compromised coordinates. In this subsection, we start by analyzing the optimal isolation rules when an attack is deemed to exist in Section V-B1, and characterize the performance of the optimal decision rules. This analysis will serve as a baseline for comparing the performance of any alternative low-complexity isolation rule. Finally, we provide an isolation rule in Section V-B2 that has low complexity and achieves the optimal performance asymptotically, as the data size n grows.

1) Optimal Isolation Rule: If the attack detection rules specified by (52)-(53) determine that the data in one or more coordinates are compromised, in the next step we aim to identify the compromised coordinates. Isolating the set of compromised coordinates is equivalent to solving the following T-hypothesis testing problem.

$$\mathsf{H}_i : \mathbf{Y} \sim f_i(\mathbf{Y} \mid X), \text{ with } X \sim \pi(X) \tag{54}$$

for $i \in \{1, \ldots, T\}$. Define $\mathsf{D_{is}} \in \{\mathsf{H}_1, \ldots, \mathsf{H}_T\}$ and $\mathsf{T_{is}} \in \{\mathsf{H}_1, \ldots, \mathsf{H}_T\}$ as the decision formed and the true hypothesis, respectively. We define the randomized test $\hat{\boldsymbol{\delta}}_1(\mathbf{Y}) \triangleq [\hat{\delta}_{11}(\mathbf{Y}), \ldots, \hat{\delta}_{1T}(\mathbf{Y})]$ for discerning the correct decision $\mathsf{D_{is}}$, where we have defined $\hat{\delta}_{1i}(\mathbf{Y}) \in [0,1]$ as the probability of deciding in favor of $\hat{\mathsf{H}}_i$. Accordingly, we define $\mathsf{P_{is}}(\hat{\boldsymbol{\delta}}_1)$ as the probability of making an erroneous decision on the problem in (54), given that the decision in the detection step was $\hat{\mathsf{H}}_1$. Therefore, $\mathsf{P_{is}}(\hat{\boldsymbol{\delta}}_1)$ is given by

$$\mathsf{P_{is}}(\hat{\boldsymbol{\delta}}_1) \triangleq \mathbb{P}(\mathsf{D_{is}} \neq \mathsf{T_{is}} \mid \mathsf{D_d} = \hat{\mathsf{H}}_1) \tag{55}$$

$$= \sum_{i=1}^{T} \sum_{j=1, j\neq i}^{T} \mathbb{P}(\mathsf{D_{is}} = \mathsf{H}_j \mid \mathsf{T_{is}} = \mathsf{H}_i, \ \mathsf{D_d} = \hat{\mathsf{H}}_1) \times \mathbb{P}(\mathsf{T_{is}} = \mathsf{H}_i \mid \mathsf{D_d} = \hat{\mathsf{H}}_1). \tag{56}$$

We also denote the error exponent of $\mathsf{P_{is}}(\hat{\boldsymbol{\delta}}_1)$ as the number of the samples n grows according to (when the limit exists)

$$\psi(\hat{\boldsymbol{\delta}}_1) \triangleq -\lim_{n\to\infty} \frac{\log \mathsf{P_{is}}(\hat{\boldsymbol{\delta}}_1)}{n}. \tag{57}$$

In the next theorem, we characterize the optimal decision rule $\hat{\boldsymbol{\delta}}_1$ and the associated error exponent. For this purpose, we denote the Chernoff information between two probability measures with probability density functions g and h by

$$C(g,h) \triangleq -\log \min_{\alpha\in(0,1)} \int g^{\alpha}(x) h^{1-\alpha}(x) \, \mathrm{d}x. \tag{58}$$

Theorem 6: The decision rule $\hat{\boldsymbol{\delta}}_1$ that minimizes $\mathsf{P_{is}}(\hat{\boldsymbol{\delta}}_1)$ is given by

$$\hat{\delta}_{1i}(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{59}$$

where

$$i^* = \arg\max_{j\in\{1,\ldots,T\}} f_j(\mathbf{Y}) \mathbb{P}(\mathsf{T_{is}} = \mathsf{H}_j \mid \mathsf{T_d} = \hat{\mathsf{H}}_1). \tag{60}$$

Proof: See Appendix D. ∎

Note that the optimal decision rules will have the same computational complexity as the optimal rules characterized in Theorem 4.

2) Asymptotically Optimal Isolation Rule: The optimal decision rule in Theorem 6 has similar drawbacks in terms of computational complexity in high dimensions as the optimal decision rules discussed in Section IV. In this subsection, we provide an alternative isolation rule for discerning the compromised coordinates at a lower computational complexity, and show that this rule is optimal in the asymptote of large number of samples, n. We denote the alternative low-complexity decision rule by $\bar{\boldsymbol{\delta}}_{\mathbf{1}}$ and start by providing lower and upper bounds on $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}})$, and show that these bounds have the same error exponents. To specify a lower bound, we first leverage the fact that $\mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}}) \leq \mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1)$, which is true due to the optimality of $\mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}})$, and define $\mathsf{P}_{\mathsf{is}}(X, \bar{\boldsymbol{\delta}}_1)$ as the value of $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1)$ when the unknown parameter X is fully known (e.g., provided by a genie). Clearly, when X is fully known, the error probability decreases, as part of the model uncertainty due to the unknown parameter is removed, and the decision problem reduces to a purely detection problem. To specify the upper bound on $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1)$, we define $\mathsf{P}_{\mathsf{is}}(X_c, \bar{\boldsymbol{\delta}}_1)$ as the value of $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1)$ when X is assumed to take an arbitrary value $X_c \in \mathbb{R}^p$. Note that replacing X by some arbitrarily chosen X_c induces sub-optimality compared to the optimal case in which optimal estimation is performed. The following lemma specifies the bounds on $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1)$.

Lemma 1: There exists some $X_c \in \mathbb{R}^p$, such that

$$\mathsf{P}_{\mathsf{is}}^{l} \triangleq \mathsf{P}_{\mathsf{is}}(X, \hat{\boldsymbol{\delta}}_1) \leq \mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_1) \leq \mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1) \leq \mathsf{P}_{\mathsf{is}}^{u} \triangleq \mathsf{P}_{\mathsf{is}}(X_c, \bar{\boldsymbol{\delta}}_1). \tag{61}$$

This lemma is instrumental to establishing the error exponent of the alternative low-complexity decision rule that we will provide. To proceed, corresponding to each coordinate $l \in \{1, \ldots, m\}$ we define the likelihood ratio term

$$\mathsf{LR}_l(\mathbf{Y}) \triangleq \prod_{r=1}^{n} \frac{g_l^1(Y_r(l))}{g_l^0(Y_r(l))}, \tag{62}$$

where g_l^0 and g_l^1 denote the marginal pdfs of the data at coordinate $l \in \{1, \ldots, m\}$ under the attack-free setting and when the coordinate is compromised, respectively. In the next theorem, we show that a decision rule based on calculating these likelihood ratio terms suffices to reach a decision that achieves the same error exponent as the optimal decision rule specified in (55).

Theorem 7: The isolation rule

$$\bar{\delta}_{1i}(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{63}$$

in which we have defined

$$i^* = \underset{i \in \{1, \ldots T\}}{\arg\max} \prod_{v \in S_i} \mathsf{LR}_v(\mathbf{Y}), \tag{64}$$

has the following error exponent

$$\psi(\bar{\boldsymbol{\delta}}_1) = \min_{i \neq j \in \{1, \ldots, T\}} C(f_i, f_j), \tag{65}$$

which is equal to $\psi(\hat{\boldsymbol{\delta}}_1)$, i.e., the error exponent of the optimal rule.

Proof: See Appendix E. ∎

Therefore, the error probabilities corresponding to the optimal decision rule in Theorem 6 and the low-complexity alternative rule in Theorem 7 decay at the same rate with the increasing number of samples n, rendering the decision rule in Theorem 7 asymptotically optimal. Clearly, the decision rule based on the marginal likelihood ratios significantly reduces the computational complexity for isolating the attacked coordinates. In the next subsection, we discuss how the attack detection and coordinate isolation decisions characterized so far are leveraged for forming an estimate of X.

C. Coordinate-Based Secure Estimation

The decision rules in Section V-A and Section V-B produce decisions about whether each individual coordinate is compromised. In the third stage of the decisions, we form one estimate of X based on all the n samples available at each coordinate. This leads to forming m distinct estimates for X, one corresponding to each coordinate. Clearly, not all the estimates are equally reliable, especially when some of the coordinates are compromised. For this purpose, after forming the m estimates we perform a *reliability* test, the purpose of which is discarding the unreliable estimates and retaining and aggregating the reliable ones. We provide relevant estimation cost functions in Section V-C1, and characterize the reliability test in Section V-C2.

1) Cost Functions: Based on the sequence of the data points at each coordinate, we form an estimate of X corresponding to each coordinate, resulting in m distinct estimates for X. Clearly, different coordinates produce estimates of X with potentially different qualities. Motivated by the fact that the decisions formed in the previous steps are not perfect and a coordinate might yield an unreliable estimate for X, we consider performing a reliability test on the decision produced at each coordinate. For this purpose, at each coordinate l and based on the data available at coordinate l, which we denote by $\mathbf{Y}_l$, we perform a binary test to decide whether the estimate based on the data from coordinate l is reliable or unreliable, denoted by $\mathsf{H}_l^{\mathsf{r}}$ and $\mathsf{H}_l^{\mathsf{u}}$, respectively.

We define $\mathsf{D}_l^{\mathsf{r}} \in \{\mathsf{H}_l^{\mathsf{r}}, \mathsf{H}_l^{\mathsf{u}}\}$ as the decision formed about the reliability of the estimate, and let the randomized test $\bar{\boldsymbol{\delta}}_l(\mathbf{Y}_l) = [\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l), \bar{\delta}_l^{\mathsf{u}}(\mathbf{Y}_l)]$ be the decision rule to decide upon the reliability of the estimate from coordinate l, where $\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)$ is the probability of deciding in favor of $\mathsf{H}_l^{\mathsf{r}}$ and $\bar{\delta}_l^{\mathsf{u}}(\mathbf{Y}_l)$ is the probability of deciding in favor of $\mathsf{H}_l^{\mathsf{u}}$. Furthermore, we define $\boldsymbol{\delta}_l(\mathbf{Y}) \triangleq [\delta_l^0(\mathbf{Y}_l), \delta_l^1(\mathbf{Y}_l)]$, where $\delta_l^1(\mathbf{Y}_l)$ denotes the probability of coordinate l being compromised based on data $\mathbf{Y}$ and the decision rules in Section V-A and Section V-B, and subsequently, $\delta_l^0(\mathbf{Y}_l) = 1 - \delta_l^1(\mathbf{Y}_l)$. Hence, the likelihood of forming a reliable estimate at coordinate l given that we have decided that coordinate l is not compromised is given by

$$\mathbb{P}_0(\mathsf{D}_l^{\mathsf{r}} = \mathsf{H}_l^{\mathsf{r}}) = \int \delta_l^0(\mathbf{Y}_l) \bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l) g_l^0(\mathbf{Y}_l) \mathrm{d}\mathbf{Y}_l. \tag{66}$$

Similarly, the likelihood of forming a reliable estimate at coordinate l given that we have decided that coordinate l is compromised is given by

$$\mathbb{P}_1(\mathsf{D}_l^{\mathsf{r}} = \mathsf{H}_l^{\mathsf{r}}) = \int \delta_l^1(\mathbf{Y}_l) \bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l) g_l^1(\mathbf{Y}_l) \mathrm{d}\mathbf{Y}_l. \tag{67}$$

Furthermore, consider U_l^0 and U_l^1 as the estimates of X at coordinate l when we have decided that the coordinate is attack-free and compromised, respectively. Hence, the cost associated with U_l^j when the decision of the reliability test is $\mathsf{H}_l^{\mathsf{r}}$ is defined as

$$J_l^j(\delta_l^j, \bar{\delta}_l^{\mathsf{r}}, U_l^j) \triangleq \mathbb{E}_j[\mathsf{C}(X, U_l^j) \mid \mathsf{D}_l^{\mathsf{r}} = \mathsf{H}_l^{\mathsf{r}}] \tag{68}$$

$$= \frac{\int\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)\mathsf{C}(U_l^j, X) g_l^j(\mathbf{Y}_l \mid X)\pi(X)\mathrm{d}X\mathrm{d}\mathbf{Y}_l}{\int\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l) g_l^j(\mathbf{Y}_l \mid X)\pi(X)\mathrm{d}X\mathrm{d}\mathbf{Y}_l} \tag{69}$$

$$= \frac{\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)\mathsf{C}_l^j(U_l^j \mid \mathbf{Y}_l) g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l}{\int \delta_j^l(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l) g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l}, \tag{70}$$

where $\mathsf{C}_l^j(U_l^j \mid \mathbf{Y}_l)$ is the posterior estimation cost at coordinate l. We define the optimal average estimation cost as

$$\hat{\mathsf{C}}_l^j(\mathbf{Y}_l) \triangleq \min_{U_l^j} \mathsf{C}_l^j(U_l^j \mid \mathbf{Y}_l), \tag{71}$$

and the optimal coordinate-level estimators as

$$\hat{X}_l^j(\mathbf{Y}_l) \triangleq \arg\min_{U_l^j} \mathsf{C}_l^j(U_l^j \mid \mathbf{Y}_l). \tag{72}$$

2) Reliability Decision Rules: For the probabilities of forming a reliable estimate at coordinate l defined in (66) and (67) we have

$$\mathbb{P}_j(\mathsf{D}_l^{\mathsf{r}} = \mathsf{H}_l^{\mathsf{r}}) = \int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l) g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l, \tag{73}$$

$$\leq \int \delta_l^j(\mathbf{Y}_l) g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l \tag{74}$$

$$= \rho_l^j, \tag{75}$$

where $(1-\rho_l^0)$ and $(1-\rho_l^0)$ are the Type-I and Type-II probabilities of mis-classifying the status of coordinate l. Therefore, the probability of forming a reliable estimate, when coordinate l is deemed to be compromised or attack-free is upper bounded by ρ_l^1 and ρ_l^0, respectively. This implies that only a fraction of the decisions on the true model of data at coordinate l will provide reliable estimates. We wish to have decision rules that control the fraction of the reliable estimates to be beyond a pre-specified level. Obviously, these desired levels should also be within the feasible range. For this purpose, we select $\nu_l^j \in [0, \rho_l^j]$, and impose a lower bound on the fraction of the reliable estimates that the reliability test retains according to

$$\mathbb{P}_j(\mathsf{D}_l^{\mathsf{r}} = \mathsf{H}_l^{\mathsf{r}}) \geq \nu_l^j. \tag{76}$$

Based on the cost functions and decision constraints, the decision rules $\bar{\boldsymbol{\delta}}_l$ and the estimators U_l^j are determined by solving m problems in parallel, where the problem to be solved corresponding to coordinate l is given by

$$\mathcal{S}(\nu_l^j) \triangleq \begin{cases} \min_{\bar{\boldsymbol{\delta}}_l, U_l^j} & J_l^j(\delta_l^j, \bar{\delta}_l^{\mathsf{r}}, U_l^j) \\ \text{s.t.} & \mathbb{P}_j(\mathsf{D}_l^{\mathsf{r}} = \mathsf{H}_l^{\mathsf{r}}) \geq \nu_l^j \end{cases}. \tag{77}$$

Similar to the discussion in Section IV, since the effect of the estimators U_l^j appears only in the cost function $J_l^j(\delta_l^j, \bar{\delta}_l^{\mathsf{r}}, U_l^j)$, the optimization problem in (77) can be decoupled into two subproblems as formalized in the following theorem.

Theorem 8: For given ν_l^0 and ν_l^1, the optimal decision rules for the problems $\mathcal{S}(\nu_l^0)$ and $\mathcal{S}(\nu_l^1)$ are given by

$$\hat{\mathsf{C}}_l^0(\mathbf{Y}_l) \underset{\mathsf{H}_1^{\mathsf{r}}}{\overset{\mathsf{H}_l^{\mathsf{u}}}{\gtrless}} \gamma_l^0 \quad \text{if coordinate } l \text{ is deemed attack-free} \tag{78}$$

$$\hat{\mathsf{C}}_l^1(\mathbf{Y}_l) \underset{\mathsf{H}_1^{\mathsf{r}}}{\overset{\mathsf{H}_l^{\mathsf{u}}}{\gtrless}} \gamma_l^1 \quad \text{if coordinate } l \text{ is deemed compromised,} \tag{79}$$

where γ_l^1 and γ_l^0 are selected such that the constraints in (77) are satisfied with equality. The optimal estimate that solve $\mathcal{S}(\nu_l^0)$ and $\mathcal{S}(\nu_l^1)$ are the estimators $\hat{X}_l^0(\mathbf{Y}_l)$ and $\hat{X}_l^1(\mathbf{Y}_l)$, respectively, defined in (72).

Proof: See Appendix F. ■

For given data $\mathbf{Y}$, performing the reliability test in Theorem 8 retains the estimates deemed reliable. These estimates, subsequently, can be aggregated to form a single estimate for X. The estimation approach, which consists of global binary attack detection, followed by coordinate-level isolation of compromised coordinates and local estimates, and completed by aggregating the reliable coordinate-level estimates, renders a scalable approach to the secure estimation of interest. This approach is significantly less complex as compared to the optimal estimation rules characterized in Theorem 2.

We remark that there exist extensive studies on optimal fusion and aggregation of local decisions, especially for inference objectives, studied under different taxonomies (e.g., distributed estimation). While there is a no general unified theory for optimal aggregation, there exist a broad range of solutions for specific choices of the statistical models and the cost functions with varying degrees of optimality [32]. Representative approaches in the context of sensor networks include [33]–[36], which focus on linearized data models. In these approaches the measurement in each coordinate is a linear combination of the coordinates in X, and they are contaminated by additive Gaussian noise. Specifically, the study in [33] focuses on networks with star topologies and provides an optimal fusion and aggregation strategy. The study in [34] focuses on these settings in which the distribution of the unknown parameter X is bounded, and provides an optimal fusion strategy. The studies in [35] and [36] discuss the optimal fusion strategies for linearized data models under the assumptions of non-Gaussian distributions for the additive noise. Specifically, [35] adopts an entropy-based cost function, and [36] adopts the absolute error as its estimation cost function. The fusion strategies in [33]–[36] can be readily adopted in our framework for a linearized data model under their respective assumptions on the noise distribution and choice of cost function. Furthermore, study in [37] proposes a consensus-based estimation strategy to form an estimate that minimizes the mean squared error for any general data model with no assumption on the noise model. The estimation strategy in [37] can also be adopted in our framework

by selecting a mean squared error cost function, using the locally formed estimates to select the reliable coordinates and then adopting a consensus-based strategy to form the final estimate by repeated rounds of information exchange among the reliable coordinates.

Algorithm 2 Scalable Solution to $\mathcal{P}(\alpha, \beta)$

1: Input feasible ν_l^0 and ν_l^1 for all coordinates l
2: Binary attack detection by (52)-(53)
3: **if** Attack exists **then**
4: Isolate the true model by (63)
5: **end if**
6: Form coordinate-level estimates by (72)
7: Coordinate-level reliability tests by (78)-(79)
8: Aggregate the reliable coordinate-level estimates to form the final estimate

Finally, we also briefly comment on the applicability of the scalable secure estimation framework to the setting with multiple attack models per attack scenario, i.e., when the space $\mathcal{F}_i$ is not a singleton. Earlier we discussed that as the number of attack models per attack scenario increases, the optimal decision rules can be generalized readily, and as long as the models $\{f_1, \ldots, f_T\}$ are fully specified, the pertinent analysis remains intact. For the scalable solution (Algorithm 2), however, the structure of the solution will be impacted. Specifically, when we have multiple attack models per attack scenario, the marginal distribution of the data at each coordinate may have more than one alternative models under the attack, i.e., the decision to select the true model at every coordinate may no longer be binary. In such settings, the decision rules in the binary attack detection step, the optimal isolation rule to select the true model when the data is deemed to be compromised, and the reliability decision rules on the estimates formed at each coordinate remain fundamentally similar with straightforward changes to accommodate the additional number of attack models per coordinate. However, the structure of the low-complexity decision rule must be altered. Specifically, multiple likelihood ratio terms defined in (62) must be evaluated corresponding to all possible attack models at each coordinate and the maximum of the products of the likelihood ratio terms corresponding to all possible combinations of compromised coordinates and their attack models determines the true model for $\mathbf{Y}$. The asymptotic optimality of this low-complexity decision rule can be established using similar technical arguments as in the proof of Theorem 7.

VI. Case Studies: Sensor Networks

We use the example of a sensor network consisting of two sensors and a fusion center (FC) to evaluate the estimation frameworks presented in this paper. Each sensor is collecting a stream of data. Sensor $i \in \{1, 2\}$ collects n measurements, denoted by $\mathbf{Y}_i = [Y_1^i, \ldots, Y_n^i]$, where each sample $Y_j^i \in \mathbb{R}$ in an attack-free environment is related to X according to

$$Y_j^i = h^i X + N_j^i, \tag{80}$$

where h^i models the channel connecting sensor i to the FC and N_j^i accounts for an additive noise. Different noise terms are assumed to be independent and identically distributed (i.i.d.) generated according to a known distribution. We will consider two adversarial scenarios that impact the data model in (80), and evaluate the optimal performance as well as the application of the asymptotically optimal scalable algorithm when the number of the samples n tends to infinity.

A. Case 1: One Sensor Vulnerable to Causative Attacks

We start by considering an adversarial setting in which the data model of the measurements from only one sensor (sensor 1) are vulnerable to a causative attack, while the other sensor (sensor 2) remains attack-free. Under such a setting, we have only one attack scenario, i.e., $T = 1$ and $S_1 = \{1\}$. Accordingly, we have $\epsilon_0 + \epsilon_1 = 1$. Under the attack-free scenario, we assume that the noise terms N_j^i are i.i.d. and distributed according to $\mathcal{N}(0, \sigma_n^2)$, i.e.,

$$Y_j^i \mid X \sim \mathcal{N}(h^i X, \sigma_n^2). \tag{81}$$

When data from sensor 1 is compromised, the actual conditional distribution of $Y_j^1|X$ is distinct from the above distribution assumed by the statistician. The inference objective under such a setting, in principle, becomes similar to the adversarial setting of [2], which focuses on a data injection attack. Hence, in order to be able to compare the performance of the optimal framework with that of [2], we assume that the conditional distribution of $Y_j^1|X$ when sensor 1 is under a causative attack is $\mathcal{N}(h^i X, \sigma_n^2) * \mathsf{Unif}[a, b]$, where $a, b \in \mathbb{R}$ are fixed constants and $*$ denotes convolution. The convolution between the normal distribution and the attack uniform distribution can be implemented as a uniform random shift of the mean of the normal distribution. Therefore, the composite hypothesis test for estimating X and discerning the model in (4) simplifies to the following binary test with the prior probabilities ϵ_0 and ϵ_1, in which we have defined $\mathbf{Y} \triangleq [\mathbf{Y}_1, \mathbf{Y}_2]$.

$$\begin{aligned} \mathsf{H}_0 \ &: \mathbf{Y} \sim f_0(\mathbf{Y} \mid X), \quad \text{with } X \sim \mathcal{N}(0, \sigma^2) \\ \mathsf{H}_1 \ &: \mathbf{Y} \sim f_1(\mathbf{Y} \mid X), \quad \text{with } X \sim \mathcal{N}(0, \sigma^2), \end{aligned} \tag{82}$$

Figure 4 depicts the variations of the estimation quality, captured by q, versus the tolerable miss-detection rate β, where it is observed that the estimation quality improves monotonically as β increases, and it reaches its maximum quality when $\beta = 1$. This observation is in line with what is expected analytically from the formulation of the secure parameter estimation problems in (17) and (18).

A similar setting is studied in [2], where the attack is induced additively into the data of sensor 1 and can be any real number. This setting can be studied in the context of causative attacks where the attacker's mode of compromising the data is adding a disturbance that has a uniform distribution. Therefore, our secure estimation framework can be applied in the context of data injection attacks as well. Figure 4 compares the estimation quality of the methodology developed in this paper, with that obtained by applying the methodology of [2], which characterizes a single point in the (q, β) plane. Specifically, in [2], an estimator is designed to obtain the most

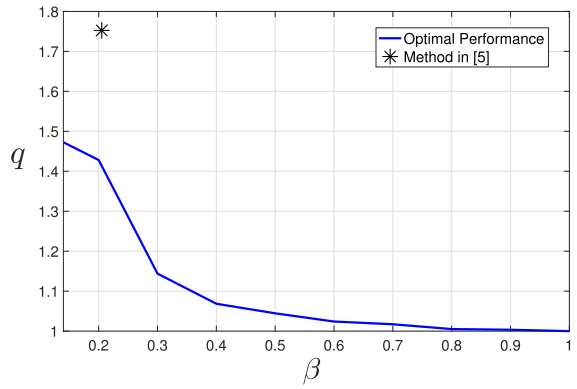

Fig. 4. q versus β for fixed $\alpha^* = 0.1$.

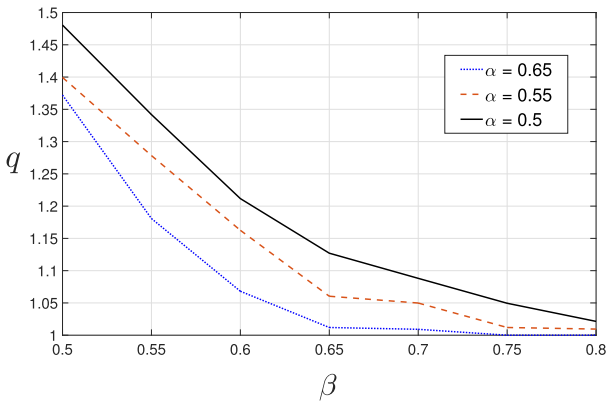

Fig. 5. q versus β for different values of α.

robust estimate by exploring the dependence of the estimation quality on the false alarm probability, using which an optimal false alarm probability α^* is obtained. This in turn, fixes the miss-detection error probability, and does not provide the flexibility to change the miss-detection rate β.

The results presented in Fig. 4 correspond to $\sigma = 3$, $\sigma_n = 1$, $h^1 = 1$, $h^2 = 4$, $a = -40$, and $b = 40$. The upper bound on $\mathsf{P_{fa}}$ is set to $\alpha^* = 0.1$, where α^* is obtained using the methodology in [2].

B. Case 2: Both Sensors Vulnerable to Causative Attacks

We again consider the same model for X, and in this setting, we assume that data from both the sensors are vulnerable to being compromised. We assume that the attacker can compromise the data of at most one sensor at any instant. Under such a setting, we have $T = 2$, $S_1 = \{1\}$, and $S_2 = \{2\}$. Therefore, under the adversarial setting, the sensor measurements follow the following composite hypothesis model

$$\begin{aligned} \mathsf{H}_0 \ &: \mathbf{Y} \sim f_0(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \mathsf{H}_1 \ &: \mathbf{Y} \sim f_1(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \mathsf{H}_2 \ &: \mathbf{Y} \sim f_2(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X), \end{aligned} \tag{83}$$

where H_0 corresponds to the attack-free setting, and hypothesis H_i corresponds to the data of sensor i being compromised. Motivating by the fact that the sensor with the higher gain h^i is expected to generate a better estimate, we explore a scenario in which the sensor with the higher gain is more likely to be attacked. Hence, we select the parameters $h^1 = 1$, and $h^2 = 2$, and set the probabilities $(\epsilon_0, \epsilon_1, \epsilon_2) = (0.2, 0.2, 0.6)$. We set the distribution of X to $\mathsf{Unif}[-2, 2]$. We assume that Y_j^i, for $i \in \{1, 2\}$, given X, is distributed according to $\mathcal{N}(h^i X, 1)$ in the attack-free setting. When sensor i is compromised, we assume that $Y_j^i \mid X \sim \mathcal{N}(h^i X, 5)$ for $i \in \{1, 2\}$.

Figure 5 depicts the performance region defined in Fig. 2 for three different values of α. These regions are the feasible regions of operation for secure estimation. This provides the FC with the flexibility to adjust the emphasis on each of the estimation or detection decisions. As expected, the estimation quality improves monotonically as α and β increase.

The proposed secure estimation framework evaluates the optimal detection and estimation rules that minimize q, and the

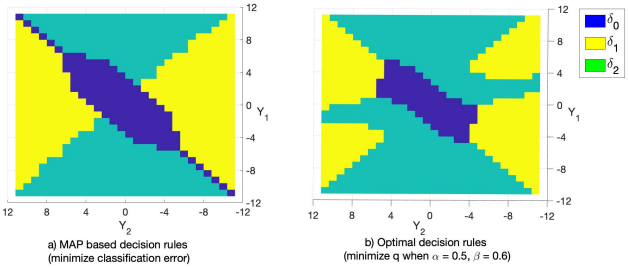

Fig. 6. Decision boundaries for MAP detection and optimal detection rules.

resulting detection rules are different from the ones that focus on minimizing the detection error rate without regards for the estimation quality. These methods include Neyman-Pearson based tests (e.g., GLRT). This is also illustrated by our experiments on the setting described in (83) for one sample collected per sensor. To highlight this difference, in Fig. 6 we depict the decision rules in two different settings. Figure 6 plots the decision regions for (1) a maximum-a-posteriori (MAP) detection rule that minimizes the detection error in detecting the true model from H_0, H_1 and H_2, and (2) the optimal detection rules $\boldsymbol{\delta}$ that minimize the degradation factor $q(\boldsymbol{\delta}, U, V)$ when one sample is collected per sensor. Clearly, the detection rules that minimize q are distinct from the ones that minimize the true model detection error. In both cases, when (Y_1, Y_2) fall in the blue region, we decide in the favor of H_0, when they fall in the yellow region or green region, we decide in the favor of H_1 or H_2, respectively.

C. Case 3: Scalable Secure Parameter Estimation Framework

We compare the estimation performance from the scalable approach developed in Section V and the optimal approach. To start, we illustrate the asymptotic optimality of the decision rule proposed in Theorem 7. Consider a 2-sensor network, where the measurements of at least one sensor are compromised at any instant. The measurements of the sensors follow a similar model as described in (80). Assuming that the parameter $X \sim \mathcal{N}(0, \sigma^2)$, Y_j^i given X is distributed according to $\mathcal{N}(h^i X, \sigma_1^2)$ under the attack-free scenario and according to $\mathcal{N}(h^i X, \sigma_2^2)$ when sensor i is compromised. We illustrate the variations of $-\log(\mathsf{P_{is}}(\bar{\boldsymbol{\delta}}_1))$ versus the number of observations

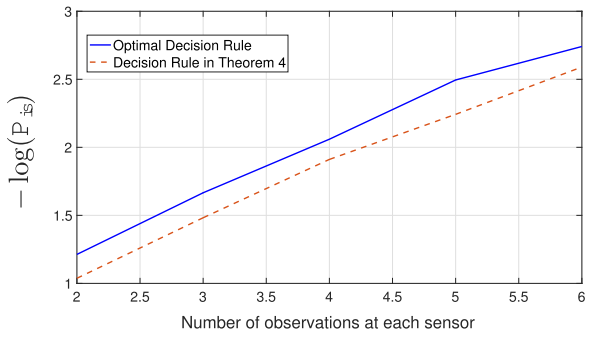

Fig. 7. $-\log(\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_1))$ versus number of observations at each sensor.

TABLE I
COMPARISON OF ESTIMATION PERFORMANCE FROM OPTIMAL DECISION RULES AND HEURISTIC APPROACH

ν_1^0	ν_1^1	ν_2^0	ν_2^1	$\hat{q}$	α	β	q
0.2	0.57	0.2	0.5261	1.48	0.0215	0.5134	1.126
0.25	0.4384	0.2	0.658	1.456	0.0216	0.6484	1.046
0.3	0.4467	0.15	0.6359	1.434	0.0482	0.5946	1.060
0.3	0.4467	0.3	0.4124	1.491	0.0762	0.4198	1.151
0.35	0.4021	0.25	0.4832	1.434	0.0528	0.4812	1.136
0.15	0.689	0.35	0.4124	1.44	0.0476	0.476	1.140

at each sensor in Fig. 7. For the results in Fig. 7, we set $h^1 = 1, h^2 = 4$, $\sigma^2 = 4$, $\sigma_1^2 = 2$, and $\sigma_2^2 = 6$. The error probabilities corresponding to both decision rules decay exponentially at the same rate with the increase in number of observations at each sensor.

In order to compare the performance of the scalable secure estimation framework with that of the optimal framework, we choose ν_l^0 and ν_l^1, for $l \in \{1, 2\}$, according to the steps described in Algorithm 2. We aggregate the reliable local estimates to form an optimal linear estimate at the FC using the fusion strategy for sensor networks with star topology in [33]. The decision on the reliability of the estimate is formed using the decision rules in Theorem 8, following which the local linear estimates from the sensors deemed to provide reliable estimates are aggregated at the FC. To compare with the estimation degradation factor q for an optimal framework, we evaluate the estimation degradation factor for the scalable secure estimation framework and denote it by $\hat{q}$. The estimation performances for the scalable framework and the optimal decision rules are compared in the following table:

For the results presented in Table I, we have set $h^1 = h^2 = 1$. We assume that the parameter X is distributed according to $\mathcal{N}(0, 3)$, N_j^i is distributed according to $\mathcal{N}(0, 1)$ under the attack-free scenario, and according to $\mathcal{N}(0, 1) * \mathsf{Unif}[-10, 10]$ when data from sensor i is compromised. As expected, the optimal decision rules result in superior estimation quality as compared to that obtained from the scalable framework, i.e., $\hat{q} > q$.

VII. CONCLUSION

We have formalized and analyzed the problem of secure parameter estimation under the potential presence of causative attacks on the estimation algorithm. Under causative attacks, the information of the estimation algorithm about the statistical model of the sampled data is compromised. This leads the estimation algorithm to exhibit degraded performance compared to the attack-free setting. We have provided closed-form optimal decision rules that ensure the best estimation quality (minimum estimation cost) while controlling the error in detecting the attacks and isolating the true model of the data. We have shown that the design of optimal estimators is intertwined with the detection rules for deciding upon the true model of the data. Based on this, we have designed the optimal decision rules, which combine both estimation performance and detection power. Based on this vision, the decision-maker can place any desired emphasis on the estimation and detection routines involved. We have also provided case studies by applying the theory developed to sensors networks, where sensors face security vulnerabilities. Finally, to circumvent the computational complexity associated with growing the data dimension or attack complexity, we have provided a low-complexity secure estimation algorithm that is optimal in the asymptote of large data size.

APPENDIX A
PROOF OF THEOREM 2

From (11) we have

$$J_i(\delta_i, U_i) = \mathbb{E}\left[\mathsf{C}(X, U_i(\mathbf{Y})) \,|\, \mathsf{D} = \mathsf{H}_i\right] = \frac{\int_{\mathbf{Y}}\int_{\mathbf{X}} \delta_i(\mathbf{Y})\mathsf{C}(X, U_i(\mathbf{Y}))f_i(\mathbf{Y} \,|\, X)\pi(X)\mathrm{d}X\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}.$$

Using the definition of $\mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y})$ from (25), we find the following lower bound on $J_i(\delta_i, U_i(\mathbf{Y}))$

$$\begin{aligned} J_i(\delta_i, U_i) &= \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})\mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}} \\ &\geq \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \inf\limits_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}, \end{aligned} \tag{84}$$

which implies that

$$J_i(\delta_i, U_i) \geq \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}. \tag{85}$$

Based on the definition of $\hat{X}_i(\mathbf{Y})$ provided in (24), this lower bound is clearly achieved when the estimator is selected as

$$\hat{X}_i(\mathbf{Y}) = \arg \inf_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}), \tag{86}$$

which proves that the estimator characterized in (24) is an optimal estimator that minimizes the cost $J_i(\delta_i, U_i)$. The

corresponding minimum average estimation cost is

$$J_i(\delta_i, \hat{X}_i) = \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}}. \tag{87}$$

Next, we prove that

$$\max_i \min_{\mathbf{U}} \{J_i(\delta_i, U_i)\} = \min_{\mathbf{U}} \max_i \{J_i(\delta_i, U_i)\}. \tag{88}$$

Recall from (12) that the overall estimation cost $J(\boldsymbol{\delta}, \mathbf{U})$ is defined as

$$J(\boldsymbol{\delta}, \mathbf{U}) = \max_i \{J_i(\delta_i, U_i)\}. \tag{89}$$

Define $\mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U})$ as a convex function of $J_i(\delta_i, U_i), i \in \{0, \ldots, T\}$, given by

$$\mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) \triangleq \sum_{i=0}^{T} \Omega_i J_i(\delta_i, U_i), \tag{90}$$

where $\boldsymbol{\Omega} = [\Omega_0, \ldots, \Omega_T]$, and Ω_i satisfy

$$\sum_{i=0}^{T} \Omega_i = 1, \text{ and } \Omega_i \in [0, 1]. \tag{91}$$

We can represent $J(\boldsymbol{\delta}, \mathbf{U})$ as a function of $\mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U})$ in the following form

$$J(\boldsymbol{\delta}, \mathbf{U}) = \max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}).$$

Let $\boldsymbol{\Omega}^* = \{\Omega^*_j : j = 0, \ldots, T\}$ be defined as

$$\boldsymbol{\Omega}^* \triangleq \arg\max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}),$$

where $\Omega^*_j = 1$ if

$$j = \arg\max_i \{J_i(\delta_i, U_i)\}. \tag{92}$$

From (86) and (87), we observe that

$$\begin{aligned} \max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) &= \max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \hat{\mathbf{X}}) \\ &\geq \min_{\mathbf{U}} \max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}). \end{aligned} \tag{93}$$

Also, at the same time, we have

$$\max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) \geq \max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}), \tag{94}$$

which implies that

$$\min_{\mathbf{U}} \max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) \geq \max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}). \tag{95}$$

From (93) and (95), it is concluded that

$$\max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) = \min_{\mathbf{U}} \max_{\boldsymbol{\Omega}} \mathcal{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}), \tag{96}$$

which completes the proof for (88). Using the results in (88) and (87), the cost function $J(\boldsymbol{\delta}, \hat{\mathbf{X}})$ is given by

$$\begin{aligned} J(\boldsymbol{\delta}, \hat{\mathbf{X}}) &= \min_{\mathbf{U}} \max_i \{J_i(\delta_i, U_i)\} \\ &= \max_i \min_{\mathbf{U}} \{J_i(\delta_i, U_i)\} \\ &= \max_i \left\{ J_i(\delta_i, \hat{X}_i) \right\} \qquad (97) \\ &= \max_i \left\{ \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}} \right\}. \qquad (98) \end{aligned}$$

Appendix B
Proof of Theorem 3

Note that the function $J_i(\delta_i, U_i)$ is a quasi-convex function in $\delta_i \in [0, 1]$. To show this, let δ_i^1 and δ_i^2 be two possible values of δ_i such that $\delta_i = \lambda\delta_i^1 + (1-\lambda)\delta_i^2$ for some $\lambda \in [0, 1]$. We have

$$\begin{aligned} &J_i(\delta_i, U_i) \\ &= \frac{\int_{\mathbf{Y}} \int_{\mathbf{X}} (\lambda\delta_i^1(\mathbf{Y}) + (1-\lambda)\delta_i^2(\mathbf{Y})) \mathsf{C}(X, U_i) f_i(\mathbf{Y} \,|\, X) \pi(X) \mathrm{d}X \mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} (\lambda\delta_i^1(\mathbf{Y}) + (1-\lambda)\delta_i^2(\mathbf{Y})) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}} \qquad (99) \\ &= \frac{\lambda \int_{\mathbf{Y}} \int_X \delta_i^1(\mathbf{Y}) \mathsf{C}(\mathbf{X}, U_i) f_i(\mathbf{Y} \,|\, X) \pi(X) \mathrm{d}X \mathrm{d}\mathbf{Y}}{\lambda \int_{\mathbf{Y}} \delta_i^1(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y} + (1-\lambda) \int_{\mathbf{Y}} \delta_i^2(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}} \\ &\quad + \frac{(1-\lambda) \int_{\mathbf{Y}} \int_X \delta_i^2(\mathbf{Y}) \mathsf{C}(X, U_i) f_i(\mathbf{Y} \,|\, X) \pi(X) \mathrm{d}X \mathrm{d}\mathbf{Y}}{\lambda \int_{\mathbf{Y}} \delta_i^1(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y} + (1-\lambda) \int_{\mathbf{Y}} \delta_i^2(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}} \qquad (100) \end{aligned}$$

Note that, for any $a, b, c, d > 0$,

$$\frac{a+b}{c+d} \leq \max\left\{\frac{a}{c}, \frac{b}{d}\right\}. \tag{101}$$

Therefore,

$$J_i(\delta_i, U_i) \leq \max\{J_i(\delta_i^1, U_i), J_i(\delta_i^2, U_i)\}, \tag{102}$$

which implies that $J_i(\delta_i, U_i)$ is quasiconvex in δ_i for any desired non-negative cost function $\mathsf{C}(X, U_i)$.

Since the weighted maximum function preserves the quasi-convexity, it is concluded that $J_i(\delta_i, \hat{X}_i)$ is a quasi-convex function from its definition in (27). Therefore, we can find the solution by solving an equivalent feasibility problem given below [38]. Specifically, for some $u \in \mathbb{R}_+$, it is easily observed that

$$J(\boldsymbol{\delta}, \hat{\mathbf{X}}) \leq u \Leftrightarrow \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) (\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u) \mathrm{d}\mathbf{Y} \leq 0, \tag{103}$$

for all $i \in \{0, \ldots, T\}$. Hence, the feasibility problem equivalent to (23) is given by

$$\begin{aligned} &\tilde{\mathcal{P}}(\alpha, \beta) \\ &= \begin{cases} \min_{\boldsymbol{\delta}} & u \\ \text{s.t.} & \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) (\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u) \mathrm{d}\mathbf{Y} \leq 0, \forall i \\ & \sum_{j=1}^{T} \sum_{i=0, i\neq j}^{T} \frac{\epsilon_j}{1-\epsilon_0} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_j(\mathbf{Y}) \mathrm{d}\mathbf{Y} \leq \beta \\ & \sum_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \mathrm{d}\mathbf{Y} \leq \alpha \end{cases} \end{aligned} \tag{104}$$

If the above problem is feasible for a given u, the problem in (23) must satisfy $\mathcal{P}(\alpha, \beta) \leq u$, and $\mathcal{P}(\alpha, \beta)$ represents the lowest possible value of u for which the problem in (104) is feasible and all its constraints are satisfied. If the problem $\tilde{\mathcal{P}}(\alpha, \beta)$ is infeasible, then $\mathcal{P}(\alpha, \beta) > u$. Given an interval

$[u_0, u_1]$ containing $\tilde{\mathcal{P}}(\alpha,\beta)$, the optimal detection rule $\boldsymbol{\delta}$ and the optimal estimation cost $\mathcal{P}(\alpha,\beta)$ can be determined by a bi-section search between u_0 and u_1 iteratively, solving the feasibility problem in each iteration. To solve the feasibility problem, we define an auxiliary convex optimization problem

$$\mathcal{R}(\alpha,\beta,u) = \begin{cases} \min_{\boldsymbol{\delta}} & \eta \\ \text{s.t.} & \displaystyle\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y})(\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u)\mathrm{d}\mathbf{Y} \le \eta, \quad \forall i \\ & \displaystyle\sum_{j=1}^{T}\sum_{i=0,i\neq j}^{T} \tfrac{\epsilon_j}{1-\epsilon_0}\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_j(\mathbf{Y})\mathrm{d}\mathbf{Y} \le \beta + \eta \\ & \displaystyle\sum_{i=1}^{T}\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y})\mathrm{d}\mathbf{Y} \le \alpha + \eta \end{cases} \tag{105}$$

Algorithm 1 summarises the steps for determining $\mathcal{P}(\alpha,\beta)$.

Algorithm 3 Bi-Section Search

```
Initialize u_0, u_1
repeat
    û ← (u_0 + u_1)/2
    Solve R(α, β, û)
    if R(α, β, û) ≤ 0 then
        u_1 ← û
    else
        u_0 ← û
    end if
until u_1 − u_0 ≤ ε, for ε sufficiently small
P(α, β) ← u_1
```

APPENDIX C
PROOF OF THEOREM 4

To solve the problem in (105), a Lagrangian function is constructed according to

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\delta},\eta,\boldsymbol{\ell}) \triangleq & \sum_{i=0}^{T}\ell_i\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_i(\mathbf{Y})(\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y})-u)\mathrm{d}\mathbf{Y} \\ & + \ell_{T+1}\sum_{j=1}^{T}\sum_{i=0,i\neq j}^{T}\frac{\epsilon_j}{1-\epsilon_0}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_j(\mathbf{Y})\mathrm{d}\mathbf{Y} \\ & - \ell_{T+1}\beta + \ell_{T+2}\sum_{i=1}^{T}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_0(\mathbf{Y})\mathrm{d}\mathbf{Y} \\ & - \ell_{T+2}\alpha, \end{aligned} \tag{106}$$

where $\boldsymbol{\ell} \triangleq [\ell_0,\ldots,\ell_{T+2}]$ are the non-negative Lagrangian multipliers selected to satisfy the constraints in (23), such that

$$\sum_{i=0}^{T+2}\ell_i = 1. \tag{107}$$

The Lagrangian function also involves the term $(1-\sum_{i=0}^{T+2}\ell_i)\eta$ pertinent to the utility function of (105), which is nullified due to (107). Consequently, the Lagrangian dual function is given by

$$\begin{aligned} d(\boldsymbol{\ell}) &\triangleq \min_{\boldsymbol{\delta},\boldsymbol{\eta}} \mathcal{Q}(\boldsymbol{\delta},\eta,\boldsymbol{\ell}) \\ &= \min_{\boldsymbol{\delta}}\left(\sum_{i=0}^{T}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})A_i\mathrm{d}\mathbf{Y}\right) - \ell_{T+1}\beta - \ell_{T+2}\alpha, \end{aligned} \tag{108}$$

where

$$A_0 \triangleq \ell_0 f_0(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},0}(\mathbf{Y}) - u] + \ell_{T+1}\sum_{i=1}^{T}\frac{\epsilon_i}{1-\epsilon_0}f_i(\mathbf{Y}), \tag{109}$$

and for $i \in \{1,\ldots,T\}$,

$$\begin{aligned} A_i \triangleq & \ell_i f_i(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u] \\ & + \ell_{T+1}\sum_{j=1,j\neq i}^{T}\frac{\epsilon_j}{1-\epsilon_0}f_j(\mathbf{Y}) + \ell_{T+2}f_0(\mathbf{Y}). \end{aligned} \tag{110}$$

Therefore, the optimal detection rules that minimize $d(\boldsymbol{\ell})$ are given by:

$$\delta_i(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{111}$$

where

$$i^* = \underset{i\in\{0,\ldots,T\}}{\operatorname{argmin}} A_i. \tag{112}$$

APPENDIX D
PROOF OF THEOREM 6

We can write $\mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}})$ as

$$\begin{aligned} \mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}}) =& \mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}} \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1) \\ =& \mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}} \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_0) \\ & \times \mathbb{P}(\mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_0 \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1) \\ & + \mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}} \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1) \\ & \times \mathbb{P}(\mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1 \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1). \end{aligned} \tag{113}$$

Note that $\mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}} \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_0) = 1$ for any decision rule $\hat{\boldsymbol{\delta}}_{\mathbf{1}}$, and the terms $\mathbb{P}(\mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1 \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1)$ and $\mathbb{P}(\mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_0 \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1)$ are independent of the decision rule $\hat{\boldsymbol{\delta}}_1$. Therefore, minimizing $\mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}})$ is equivalent to minimizing $\mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}} \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)$. Furthermore,

$$\begin{aligned} &\mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}} \mid \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1) \\ &= \frac{\mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{T}_{\mathsf{is}}, \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1 \mid \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)\mathbb{P}(\mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)}{\mathbb{P}(\mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)} \end{aligned} \tag{114}$$

$$\begin{aligned} &= \frac{\mathbb{P}(\mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)}{\mathbb{P}(\mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)} \\ &\times \sum_{i=1}^{T}\left(\sum_{j=1,j\neq i}^{T}\mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_i, \mathsf{D}_{\mathsf{d}} = \hat{\mathsf{H}}_1 \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_j, \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)\right. \\ &\qquad \left.\times \mathbb{P}(\mathsf{T}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{d}} = \hat{\mathsf{H}}_1)\right) \end{aligned} \tag{115}$$

$$= \frac{\mathbb{P}(\mathsf{T_d} = \hat{\mathsf{H}}_1)}{\mathbb{P}(\mathsf{D_d} = \hat{\mathsf{H}}_1, \mathsf{T_d} = \hat{\mathsf{H}}_1)} \times \sum_{i=1}^{T} \Big((1 - \mathbb{P}(\mathsf{D_{is}} = \mathsf{H}_i, \mathsf{D_d} = \hat{\mathsf{H}}_1 \mid \mathsf{T_{is}} = \mathsf{H}_i, \mathsf{T_d} = \hat{\mathsf{H}}_1)) \times \mathbb{P}(\mathsf{T_{is}} = \mathsf{H}_i \mid \mathsf{T_d} = \hat{\mathsf{H}}_1) \Big) \tag{116}$$

$$= \frac{\mathbb{P}(\mathsf{T_d} = \hat{\mathsf{H}}_1)}{\mathbb{P}(\mathsf{D_d} = \hat{\mathsf{H}}_1, \mathsf{T_d} = \hat{\mathsf{H}}_1)} \times \sum_{i=1}^{T} \left((1 - \int_{R_i \cap \hat{R}_1} f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}) \mathbb{P}(\mathsf{T_{is}} = \mathsf{H}_i \,|\, \mathsf{T_d} = \hat{\mathsf{H}}_1) \right), \tag{117}$$

where $R_i \subseteq \mathcal{Y}^n$ is the region corresponding to the decision $\mathsf{D_{is}} = \mathsf{H}_i$ and $\hat{R}_1 \subseteq \mathcal{Y}^n$ is the region corresponding to the decision $\mathsf{D_d} = \hat{\mathsf{H}}_1$. Since we form the decision $\mathsf{D_{is}}$ only when $\mathsf{D_d} = \hat{\mathsf{H}}_1$, we conclude that $R_i \subseteq \hat{R}_1$ and hence, $R_i \cap \hat{R}_1 = R_i$. Therefore,

$$\mathbb{P}(\mathsf{D_{is}} \neq \mathsf{T_{is}} \mid \mathsf{D_d} = \hat{\mathsf{H}}_1, \mathsf{T_d} = \hat{\mathsf{H}}_1) = \frac{\mathbb{P}(\mathsf{T_d} = \hat{\mathsf{H}}_1)}{\mathbb{P}(\mathsf{D_d} = \hat{\mathsf{H}}_1, \mathsf{T_d} = \hat{\mathsf{H}}_1)} \times \sum_{i=1}^{T} \left((1 - \int_{R_i} f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}) \mathbb{P}(\mathsf{T_{is}} = \mathsf{H}_i \,|\, \mathsf{T_d} = \hat{\mathsf{H}}_1) \right), \tag{118}$$

and $\mathbb{P}(\mathsf{D_{is}} \neq \mathsf{T_{is}} \mid \mathsf{D_d} = \hat{\mathsf{H}}_1, \mathsf{T_d})$ is minimized when the regions R_i are chosen using the decision rule

$$\hat{\delta}_{1i}(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{119}$$

where

$$i^* = \underset{j \in \{1,\ldots,T\}}{\arg\max}\, f_j(\mathbf{Y}) \mathbb{P}(\mathsf{T_{is}} = \mathsf{H}_j \mid \mathsf{T_d} = \hat{\mathsf{H}}_1). \tag{120}$$

Appendix E
Proof of Theorem 7

For the decision rule defined in Theorem 7, we denote the decision formed as $\bar{\mathsf{D}}_{\mathsf{is}} \in \{\mathsf{H}_1, \ldots, \mathsf{H}_T\}$ to distinguish it from the decision formed by the optimal decision rule in Lemma 6. Also, from Lemma 1,

$$-\frac{\log \mathsf{P}_{\mathsf{is}}^u}{n} \leq -\frac{\log \mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}})}{n} \leq -\frac{\mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}})}{n} \leq -\frac{\log \mathsf{P}_{\mathsf{is}}^l}{n}. \tag{121}$$

Note that we can write $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}})$ as

$$\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}}) = \int \mathsf{P}_{\mathsf{is}}(X, \bar{\boldsymbol{\delta}}_{\mathbf{1}}) \pi(X) \mathrm{d}X,$$

where $\mathsf{P}_{\mathsf{is}}(X, \bar{\boldsymbol{\delta}}_{\mathbf{1}})$ is the error probability $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}})$ when conditioned on X. Therefore, there exists at least one $X_c \in \mathbb{R}^p$, such that

$$\mathsf{P}_{\mathsf{is}}(X_c, \bar{\boldsymbol{\delta}}_{\mathbf{1}}) \geq \mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}}).$$

For any arbitrary choice of X_c that satisfies the equation above, we analyze the probability $\mathsf{P}_{\mathsf{is}}^u = \mathsf{P}_{\mathsf{is}}(X_c, \bar{\boldsymbol{\delta}}_{\mathbf{1}})$. Next, we find an upper bound on $\mathsf{P}_{\mathsf{is}}^u$. When X is assumed to be X_c, the probability that the decision formed by the rule in Theorem 7 is H_j when the true model is H_i is given by $\mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T_{is}} = \mathsf{H}_i, X_c)$. Assuming that the probabilities ϵ_i, for $i \in \{1, \ldots, T\}$, are independent of the choice of X_c, we have

$$\mathsf{P}_{\mathsf{is}}^u = \sum_{i=1}^{T} \sum_{\substack{j=1 \\ i \neq j}}^{T} \epsilon_j \mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T_{is}} = \mathsf{H}_j, X_c) \tag{122}$$

$$= \sum_{\substack{j=2 \\ i<j}}^{T} (\epsilon_j + \epsilon_i) \mathsf{P}^{ij}, \tag{123}$$

where we have defined

$$\mathsf{P}^{ij} \triangleq \frac{\epsilon_j}{\epsilon_i + \epsilon_j} \cdot \mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T_{is}} = \mathsf{H}_j, X_c) + \frac{\epsilon_i}{\epsilon_i + \epsilon_j} \cdot \mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T_{is}} = \mathsf{H}_i, X_c). \tag{124}$$

Therefore,

$$\mathsf{P}_{\mathsf{is}}^u \leq \frac{T(T-1)}{2} \max_{i \neq j; i,j, \in \{1,\ldots T\}} \mathsf{P}^{ij}, \tag{125}$$

$$\leq \frac{T(T-1)}{2} \times \max_{i \neq j; i,j \in \{1,\ldots T\}} \left\{ \mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T_{is}} = \mathsf{H}_j, X_c), \mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T_{is}} = \mathsf{H}_i, X_c) \right\}. \tag{126}$$

Let $T' \triangleq \frac{T(T-1)}{2}$. Note that

$$\mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T_{is}} = \mathsf{H}_i, X = X_c) \leq \int_{R_{ij}} f_i(\mathbf{Y} \mid X_c) \, \mathrm{d}\mathbf{Y}, \tag{127}$$

where $R_{ij} = \left\{ \mathbf{Y} : \prod_{a \in S_j} \mathsf{LR}_a \geq \prod_{b \in S_i} \mathsf{LR}_b \right\}$. Define B_i as the set of coordinates deemed to be compromised in H_i but not in H_j, i.e., $B_i \triangleq S_i - S_i \cap S_j$, with $|B_i| = r_i$. Similarly, define $B_j \triangleq S_j - S_i \cap S_j$, with $|B_j| = r_j$. Also, since we have g_l^j, for $l \in \{1, \ldots, n\}, j \in \{0, 1\}$, to be conditionally independent distributions given X_c, we have

$$f_i(\mathbf{Y} \,|\, X_c) = \prod_{a \in S_i} \left(\prod_{c=1}^{n} g_a^1(Y_c(a) \,|\, X_c) \right) \times \prod_{b \in \bar{S}_i} \left(\prod_{d=1}^{n} g_b^0(Y_d(b) \,|\, X_c) \right), \tag{128}$$

where $\bar{S}_i \triangleq \{1, \ldots, m\} \backslash S_i$. Then, the region R_{ij} is equivalent to

$$\left\{ \mathbf{Y} : \left(\prod_{a \in B_j} \prod_{c=1}^{n} g_a^1(Y_c(a) \,|\, X_c) \right) \left(\prod_{b \in B_i} \prod_{d=1}^{n} g_b^0(Y_b(d) \,|\, X_c) \right) \geq \left(\prod_{a \in B_i} \prod_{c=1}^{n} g_a^1(Y_c(a) \mid X_c) \right) \left(\prod_{b \in B_j} \prod_{d=1}^{n} g_b^0(Y_d(b) \mid X_c) \right) \right\}. \tag{129}$$

Therefore,

$$\begin{aligned}\mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X_c) &\leq \int_{R_{ij}} \left(\prod_{a\in B_i} \prod_{c=1}^{n} g_a^1(Y_c(a) \mid X_c) \right) \\ &\quad \times \left(\prod_{b\in B_j} \prod_{d=1}^{n} g_b^0(Y_d(b) \mid X_c) \right) \mathrm{d}\mathbf{Y}_{B_i\cup B_j} \quad (130)\\ &= \int_{R_{ij}} \min\{t_1, t_2\} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j}, \quad (131)\end{aligned}$$

where $\mathbf{Y}_{B_i\cup B_j}$ is the data for the coordinates in the set $B_i \cup B_j$ and

$$t_1 \triangleq \left(\prod_{a\in B_i} \prod_{c=1}^{n} g_a^1(Y_c(a)|X_c) \right) \left(\prod_{b\in B_j} \prod_{d=1}^{n} g_b^0(Y_d(b)|X_c) \right), \quad (132)$$

$$t_2 \triangleq \left(\prod_{a\in B_j} \prod_{c=1}^{n} g_a^1(Y_c(a)|X_c) \right) \left(\prod_{b\in B_i} \prod_{d=1}^{n} g_b^0(Y_b(d)|X_c) \right). \quad (133)$$

Using the inequality

$$\min(a,b) \leq a^\lambda b^{1-\lambda}, \ \forall\, a,b>0, \ \text{and} \quad \lambda \in [0,1], \quad (134)$$

and

$$\int_{R_{ij}} f_i(\mathbf{Y} \mid X_c) \, \mathrm{d}\mathbf{Y} \leq \int f_i(\mathbf{Y} \mid X_c) \, \mathrm{d}\mathbf{Y}, \quad (135)$$

from (130) we get

$$\mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X_c) \leq \int t_1^\lambda t_2^{1-\lambda} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j}, \quad (136)$$

for all $\lambda \in [0,1]$. Following a similar line of analysis as in (130)-(136), it can be verified that

$$\mathbb{P}(\bar{\mathsf{D}}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_j, X_c) \leq \int t_1^\lambda t_2^{1-\lambda} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j}. \quad (137)$$

The inequalities in (136) and (137) hold for all $\lambda \in [0,1]$. Hence, from (125), (136), and (137) we get

$$\mathsf{P}_{\mathsf{is}}^u \leq T' \max_{i,j} \min_{\lambda\in[0,1]} \int t_1^\lambda t_2^{1-\lambda} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j}. \quad (138)$$

We use the change of variables $\tilde{Y}_a(b) = Y_a(b) \mid X_c$ to obtain

$$\begin{aligned}&\int t_1^\lambda t_2^{1-\lambda} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j} \\ &= \left(\prod_{a\in B_j} \prod_{c=1}^{n} \int (g_a^1(\tilde{Y}_a(c)))^\lambda (g_a^0(\tilde{Y}_a(c)))^{1-\lambda} \, \mathrm{d}\tilde{Y}_a(c) \right) \\ &\quad \times \left(\prod_{b\in B_i} \prod_{d=1}^{n} \int (g_b^1(\tilde{Y}_b(d)))^\lambda (g_b^0(\tilde{Y}_b(d)))^{1-\lambda} \, \mathrm{d}\tilde{Y}_b(d) \right). \quad (139)\end{aligned}$$

Note that

$$-\frac{\log \mathsf{P}_{\mathsf{is}}^u}{n} = -\frac{1}{n} \log \left(\max_{i,j} \min_{\lambda} \int t_1^\lambda t_2^{1-\lambda} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j} \right). \quad (140)$$

By letting n to grow to infinity, and using the monotonicity of the log function, we have

$$\begin{aligned}&-\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathsf{is}}^u}{n} \\ &= \lim_{n\to\infty} \frac{1}{n} \min_{i,j} \max_{\lambda} -\log \left(\int t_1^\lambda t_2^{1-\lambda} \, \mathrm{d}\mathbf{Y}_{B_i\cup B_j} \right) \quad (141)\\ &= \lim_{n\to\infty} \frac{1}{n} \min_{i,j} \max_{\lambda} \\ &\quad \left(\sum_{a\in B_j} -n \log \left(\int (g_a^1(Y))^\lambda (g_a^0(Y))^{1-\lambda} \, dY \right) \right. \\ &\quad \left. + \sum_{b\in B_i} -n \log \left(\int (g_b^0(Y))^\lambda (g_b^1(Y))^{1-\lambda} \, dY \right) \right). \quad (142)\end{aligned}$$

Therefore,

$$\begin{aligned}&-\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathsf{is}}^u}{n} \\ &= \min_{i,j} \max_{\lambda} \left(\sum_{a\in B_j} -\log \left(\int g_a^1(Y)^\lambda g_a^0(Y)^{1-\lambda} \, dY \right) \right. \\ &\quad \left. + \sum_{b\in B_i} -\log \left(\int g_b^0(Y)^\lambda g_b^1(Y)^{1-\lambda} \, dY \right) \right). \quad (143)\end{aligned}$$

We now find a lower bound on $\mathsf{P}_{\mathsf{is}}^{\mathsf{l}}$. Under the perfect knowledge of X, we have

$$\begin{aligned}\mathsf{P}_{\mathsf{is}}^l &= \sum_{i=1}^{T} \sum_{\substack{j=1\\ j\neq i}}^{T} \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X)\epsilon_i \quad (144)\\ &= \sum_{i=1}^{T} \mathbb{P}(\mathsf{D}_{\mathsf{is}} \neq \mathsf{H}_i \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X)\epsilon_i \quad (145)\\ &\geq \max_{i,j} \{ \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X)\epsilon_i, \\ &\qquad \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_j, X)\epsilon_j \}. \quad (146)\end{aligned}$$

Note that

$$\begin{aligned}&\lim_{n\to\infty} -\frac{\log(\mathsf{P}_{\mathsf{is}}^l)}{n} \\ &\leq \lim_{n\to\infty} -\frac{1}{n} \log \left(\max_{i,j} \{ \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X)\epsilon_i \ , \right. \\ &\qquad \left. \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_j, X)\epsilon_j \} \right) \\ &= \lim_{n\to\infty} -\frac{1}{n} \log \left(\max_{i,j} \{ \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_j \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_i, X) \ , \right. \\ &\qquad \left. \mathbb{P}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_i \mid \mathsf{T}_{\mathsf{is}} = \mathsf{H}_j, X) \} \right). \quad (147)\end{aligned}$$

For further analysis, we adopt the approach used in [39]. Under model H_i, the data points Y_w are distributed according to the pdf f_i, for $w \in \{1, \ldots n\}$ and $i \in \{1, \ldots T\}$. Define $b_{i,j}^\lambda(Y_w)$ as the pdf

$$b_{i,j}^\lambda(Y_w) \triangleq \frac{f_i^\lambda(Y_w \mid X) f_j^{1-\lambda}(Y_w \mid X)}{\int f_i^\lambda(Y_w \mid X) f_j^{1-\lambda}(Y_w \mid X) \mathrm{d}Y_w}, \quad (148)$$

and let

$$\lambda^* = \arg\max_{\lambda\in[0,1]} -\log\left(\int f_i^{\lambda}(\mathbf{Y}\mid X) f_j^{1-\lambda}(\mathbf{Y}\mid X)\ \mathrm{d}\mathbf{Y}\right) \tag{149}$$

$$= \arg\max_{\lambda\in[0,1]} -\sum_{w=1}^{n}\log\left(\int f_i^{\lambda}(Y_w\,|\,X) f_j^{1-\lambda}(Y_w\,|\,X)\ \mathrm{d}Y_w\right) \tag{150}$$

$$= \arg\max_{\lambda\in[0,1]} -n\log\left(\int f_i^{\lambda}(Y_w\mid X) f_j^{1-\lambda}(Y_w\mid X)\ \mathrm{d}Y_w\right). \tag{151}$$

It can be readily verified that

$$\max_{\lambda\in[0,1]} -\log\left(\int f_i^{\lambda}(Y_w\mid X) f_j^{1-\lambda}(Y_w\mid X)\ \mathrm{d}Y_w\right) = D_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_i(.\mid X)) = D_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\mid X)), \tag{152}$$

where $D_{\mathsf{KL}}(f\|g)$ denotes the Kullback-Liebler divergence between distributions f and g. Define the probability measure $\tilde{\mathbb{P}}$ as

$$\tilde{\mathbb{P}}(\mathbf{Y}\mid X) \triangleq \prod_{w=1}^{n} b_{i,j}^{\lambda^*}(Y_w), \tag{153}$$

and define a random process M_r as

$$M_r \triangleq \sum_{w=1}^{r}\left(\log\left(\frac{b_{i,j}^{\lambda^*}(Y_w)}{f_j(Y_w\mid X)}\right) - \mathbb{E}\left[\log\left(\frac{b_{i,j}^{\lambda^*}(Y_w)}{f_j(Y_w\mid X)}\right)\mid F_{w-1}\right]\right), \tag{154}$$

where F_{w-1} is the $\sigma-$field generated by $\{Y_1,\ldots,Y_{w-1}\}$, and the expectation is with respect to the probability measure $\tilde{\mathbb{P}}$. It can be verified that the process M_r is a stable martingale. According to the martingale stability theorem [40], we have

$$\lim_{r\to\infty}\frac{M_r}{r}\xrightarrow{\text{a.s}} 0. \tag{155}$$

This can be equivalently written in the following form:

$$\lim_{r\to\infty}\tilde{\mathbb{P}}\left(\frac{M_r}{r}>\nu\right)=0,\quad \forall\nu>0. \tag{156}$$

Since for a given X the random vectors Y_w are i.i.d., we have

$$\mathbb{E}\left[\log\left(\frac{b_{i,j}^{\lambda^*}(Y_w)}{f_j(Y_w\mid X)}\right)\,\middle|\,F_{w-1}\right] = D_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_i(.\mid X)) = D_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\mid X)). \tag{157}$$

By using (157) and restructuring (156), we obtain

$$\lim_{n\to\infty}\tilde{\mathbb{P}}\left(\prod_{w=1}^{n} f_j(Y_w\,|\,X) > \exp(-nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\,|\,X))-n\nu)\times\prod_{w=1}^{n} b_{i,j}^{\lambda^*}(Y_w)\right)=1. \tag{158}$$

Under the probability measure $\tilde{\mathbb{P}}$, we denote the probability of deciding H_i as the true model when X is given as $\tilde{\mathbb{P}}(\mathsf{D}_{\mathsf{is}} = \mathsf{H}_i \mid X)$ and note that either $\tilde{\mathbb{P}}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_i\mid X)\geq\frac{1}{2}$ or $\tilde{\mathbb{P}}(\mathsf{D}_{\mathsf{is}}=\mathsf{H}_i\mid X)\geq\frac{1}{2}$. Then, assuming that $\tilde{\mathbb{P}}(\mathsf{D}_{\mathsf{is}}=\mathsf{H}_i\mid X)\geq\frac{1}{2}$ holds, from (158) we get

$$\lim_{n\to\infty}\tilde{\mathbb{P}}\left(\prod_{w=1}^{n} f_j(Y_w\,|\,X) > \exp(-nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\,|\,X))-n\nu)\times\prod_{w=1}^{n} b_{i,j}^{\lambda^*}(Y_w), \mathsf{D}_{\mathsf{is}}=\mathsf{H}_i\mid X\right)\geq\frac{1}{2}-\kappa, \tag{159}$$

for any $\kappa>0$. Using (153), we conclude that (159) is equivalent to

$$\lim_{n\to\infty}\int_R\prod_{w=1}^{n} b_{i,j}^{\lambda^*}(Y_w)\ \mathrm{d}Y_w\geq\frac{1}{2}-\kappa, \tag{160}$$

where R is the region

$$\left\{\mathbf{Y}:\{\mathsf{D}_{\mathsf{is}}=\mathsf{H}_i\mid X\}\text{ and }\left\{\prod_{w=1}^{n} b_{i,j}^{\lambda^*}(Y_w\mid X) < \exp(nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\mid X))+n\nu)\times\prod_{w=1}^{n} f_j(Y_w\mid X)\right\}\right\}. \tag{161}$$

In the region R, we have

$$\begin{aligned}\int_R\prod_{w=1}^{n} b_{i,j}^{\lambda^*}(Y_w\mid X)\ \mathrm{d}Y_w &\leq \exp(nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\,|X))+n\nu)\int_R\prod_{w=1}^{n} f_j(Y_w\,|X)\mathrm{d}\mathbf{Y},\\ &\leq \exp(nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\,|X))+n\nu)\int_{R_2}\prod_{w=1}^{n} f_j(Y_w\,|X)\mathrm{d}\mathbf{Y},\\ &= \exp(kD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\,|X))+n\nu)\\ &\quad\times\mathbb{P}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_j\,|\,\mathsf{T}_{\mathsf{is}}=\mathsf{H}_j,X),\end{aligned} \tag{162}$$

where region R_2 is $\{\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_j\mid X\}$ and also, $R\subseteq R_2$. From (160) and (162), it is concluded that

$$\mathbb{P}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_j\mid\mathsf{T}_{\mathsf{is}}=\mathsf{H}_j,X)\geq\left(\frac{1}{2}-\kappa\right)\exp(-nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\mid X))-n\nu), \tag{163}$$

for any $\nu,\kappa>0$. Similarly, if the case $\tilde{\mathbb{P}}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_i\mid X)\geq\frac{1}{2}$ holds,

$$\mathbb{P}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_i\mid\mathsf{T}_{\mathsf{is}}=\mathsf{H}_i,X)\geq\left(\frac{1}{2}-\kappa\right)\exp(-nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_i(.\mid X))-n\nu). \tag{164}$$

Using (147), (163) and (164), we get

$$\begin{aligned}\lim_{n\to\infty}-\frac{\log(\mathsf{P}_{\mathsf{is}}^{l})}{n} &\leq \lim_{n\to\infty}-\frac{1}{n}\max_{i,j}\{\log(\mathbb{P}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_j\mid\mathsf{T}_{\mathsf{is}}=\mathsf{H}_j,X)),\\ &\quad\log(\mathbb{P}(\mathsf{D}_{\mathsf{is}}\neq\mathsf{H}_i\mid\mathsf{T}_{\mathsf{is}}=\mathsf{H}_i,X))\}\end{aligned} \tag{165}$$

$$= \lim_{n\to\infty} \frac{1}{n}\left(-\left(\frac{1}{2}-\kappa\right)+\kappa\nu\right) + \lim_{n\to\infty} \frac{1}{n} \min_{i,j}\{nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_i(.\mid X)), nD_{\mathsf{KL}}(b_{i,j}^{\lambda^*}\|f_j(.\mid X))\}. \quad (166)$$

Using the fact that κ and ν can be made arbitrarily close to 0, and using (152), we get

$$\lim_{n\to\infty} -\frac{\log(\mathsf{P}_{\mathsf{is}}^l)}{n} \le \min_{i,j} \max_{\lambda\in[0,1]} -\log\left(\int f_i^{\lambda}(Y_w\mid X) f_j^{1-\lambda}(Y_w\mid X)\ \mathrm{d}Y_w\right) \quad (167)$$

$$= \min_{i,j}\max_{\lambda}\left\{\sum_{a\in B_j} -\log\left(\int (g_a^1(Y))^{\lambda}(g_a^0(Y))^{1-\lambda}\mathrm{d}Y\right) + \sum_{b\in B_i} -\log\left(\int (g_b^0(Y)^{\lambda}(g_b^1(Y))^{1-\lambda}\mathrm{d}Y\right)\right\} \quad (168)$$

$$= \min_{i\neq j} C(f_i,f_j), \quad (169)$$

where (169) follows after the change of variables in (167). From (143) and (168), it is clear that

$$\lim_{n\to\infty} -\frac{\log \mathsf{P}_{\mathsf{is}}^l}{k} = \lim_{n\to\infty} -\frac{\log \mathsf{P}_{\mathsf{is}}^u}{n}. \quad (170)$$

Then, using (121), we conclude that the error exponents of $\mathsf{P}_{\mathsf{is}}(\hat{\boldsymbol{\delta}}_{\mathbf{1}})$ and $\mathsf{P}_{\mathsf{is}}(\bar{\boldsymbol{\delta}}_{\mathbf{1}})$ are the same and are given by (169).

Appendix F
Proof of Theorem 8

We aim to design the estimator U_l^j and the decision rule $\bar{\boldsymbol{\delta}}_l$ that minimize the utility function $J_l^j(\delta_l^j,\bar{\delta}_l^{\mathsf{r}},U_l^j)$ subject to the constraint

$$\mathbb{P}_j(\mathsf{D}_l^{\mathsf{r}}=\mathsf{H}_l^{\mathsf{r}}) = \int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l \ge 1-\nu_l^j. \quad (171)$$

Since the effect of estimator only appears in $J_l^j(\delta_l^j,\bar{\delta}_l^{\mathsf{r}},U_l^j)$, the optimization problem can be decoupled similarly to the approach followed earlier, and the optimal estimator can be readily verified to be

$$\hat{X}_l^j(\mathbf{Y}_l) = \operatorname*{argmin}_{\mathcal{U}_l^j} J_l^j(\delta_l^j,\bar{\delta}_l^{\mathsf{r}},U_l^j), \quad (172)$$

where $\hat{X}_l^j(\mathbf{Y}_l)$ is defined in (72). In order to design the decision rules, we start by noting that for a decision rule $\bar{\boldsymbol{\delta}}_l$, such that,

$$\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l > 1-\nu_l^j, \quad (173)$$

we can design another decision rule $\boldsymbol{\Delta}_l \triangleq [\Delta_l^{\mathsf{r}}(\mathbf{Y}_l),\Delta_l^{\mathsf{u}}(\mathbf{Y}_l)]$, such that,

$$\int \delta_l^j(\mathbf{Y}_l)\Delta_l^{\mathsf{r}}(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l = 1-\nu_l^j, \quad (174)$$

with the same estimation performance. To show this, we set

$$\Delta_l^{\mathsf{r}}(\mathbf{Y}_l) = \frac{(1-\nu_l^j)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)}{\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l}, \quad (175)$$

which satisfies (171) with equality. Using (173), note that $\Delta_l^{\mathsf{r}}(\mathbf{Y}_l) < \bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)$, which implies that $\boldsymbol{\Delta}_l$ is a valid decision rule. We can easily verify that

$$J_l^j(\delta_l^j,\Delta_l^{\mathsf{r}},\hat{X}_l^j) = J_l^j(\delta_l^j,\bar{\delta}_l^{\mathsf{r}},\hat{X}_l^j), \quad (176)$$

which implies that the estimation performance is the same for both decision rules, and therefore, we can restrict our design for the optimum decision rule to the class of rules that satisfy (171) with equality. Under the equality condition,

$$\int \delta_j^i(\mathbf{Y}_i),\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l = 1-\nu_l^j, \quad (177)$$

in which case minimizing $J_l^j(\delta_l^j,\bar{\delta}_l^{\mathsf{r}},\hat{X}_l^j)$ is equivalent to minimizing $\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l$. Let $\gamma_l^j \ge 0$ be the solution to

$$\mathbb{P}_j(\gamma_l^j \ge \hat{\mathsf{C}}_l^j(\mathbf{Y}_l)) = \int_{\hat{R}} \delta_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l = 1-\nu_l^j, \quad (178)$$

where $\hat{R} \triangleq \left\{\mathbf{Y}_l : \gamma_l^j \ge \hat{\mathsf{C}}_l^j(\mathbf{Y}_l)\right\}$. Hence,

$$\begin{aligned}
&\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l - \gamma_l^j(1-\nu_l^j)\\
&= \int \delta_l^j(Y_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l\\
&\quad - \gamma_l^j\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l\\
&= \int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)(\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)-\gamma_l^j)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l\\
&\ge \int_{\hat{R}} \delta_j^l(\mathbf{Y}_l)(\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)-\gamma_l^j)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l\\
&= \int_{\hat{R}} \delta_j^l(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l - \gamma_l^j\mathbb{P}_j(\gamma_l^j \ge \hat{\mathsf{C}}_l^j(\mathbf{Y}_l))\\
&= \int_{\hat{R}} \delta_l^j(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l - \gamma_l^j(1-\nu_l^j). \quad (179)
\end{aligned}$$

Clearly,

$$\int \delta_l^j(\mathbf{Y}_l)\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l \ge \int_{\hat{R}} \delta_l^j(\mathbf{Y}_l)\hat{\mathsf{C}}_l^j(\mathbf{Y}_l)g_l^j(\mathbf{Y}_l)\mathrm{d}\mathbf{Y}_l, \quad (180)$$

and therefore, the decision rule $\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l)$ given by

$$\bar{\delta}_l^{\mathsf{r}}(\mathbf{Y}_l) = \mathbb{1}_{\{\gamma_l^j \ge \hat{\mathsf{C}}_l^j(\mathbf{Y}_l)\}}, \quad (181)$$

is optimal since it ensures optimal estimation performance and satisfies the constraint in (171).

References

[1] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar, "Can machine learning be secure," in *Proc. ACM Symp. Inf., Comput. Commun. Secur. (ASIACCS)*, Taipei, Taiwan, Mar. 2006, pp. 16–25.

[2] C. Wilson and V. Veeravalli, "MMSE estimation in a sensor network in the presence of an adversary," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2479–2483.

[3] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with Byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 65–75, Sep. 2013.
[4] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. K. Varshney, "Localization in wireless sensor networks: Byzantines and mitigation techniques," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1495–1508, Mar. 2013.
[5] P. Ebinger and S. D. Wolthusen, "Efficient state estimation and Byzantine behavior identification in tactical MANETs," in *Proc. MILCOM IEEE Mil. Commun. Conf.*, Boston, MA, USA, Oct. 2009, pp. 1–7.
[6] J. Zhang, R. S. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1086–1101, Mar. 2015.
[7] J. Zhang and R. S. Blum, "Distributed estimation in the presence of attacks for large scale sensor networks," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2014, pp. 1–6.
[8] A. Vempaty, Y. S. Han, and P. K. Varshney, "Target localization in wireless sensor networks using error correcting codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 697–712, Jan. 2014.
[9] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Countering Byzantine attacks in cognitive radio networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 3098–3101.
[10] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 205–215, Jan. 2013.
[11] A. Vempaty, K. Agrawal, P. Varshney, and H. Chen, "Adaptive learning of Byzantines' behavior in cooperative spectrum sensing," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Cancun, Mexico, Mar. 2011, pp. 1310–1315.
[12] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Monticello, IL, USA, Sep. 2011, pp. 337–344.
[13] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.
[14] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Osaka, Japan, Dec. 2015, pp. 5162–5169.
[15] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, Osaka, Japan, Dec. 2015, pp. 5827–5832.
[16] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 4917–4932, Oct. 2017.
[17] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 2929–2933.
[18] M. Pajic *et al.*, "Robustness of attack-resilient state estimators," in *Proc. ACM/IEEE Int. Conf. Cyber-Phys. Syst. (ICCPS)*, Berlin, Germany, Apr. 2014, pp. 163–174.
[19] C.-Z. Bai and V. Gupta, "On Kalman filtering in the presence of a compromised sensor: Fundamental performance bounds," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 3029–3034.
[20] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 434–444, May 1968.
[21] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?" *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1597–1602, Sep. 1992.
[22] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4215–4229, Jul. 2012.
[23] G. H. Jajamovich, A. Tajer, and X. Wang, "Minimax-optimal hypothesis testing with estimation-dependent costs," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6151–6165, Dec. 2012.
[24] X. Shen, P. K. Varshney, and Y. Zhu, "Robust distributed maximum likelihood estimation with dependent quantized data," *Automatica*, vol. 50, no. 1, pp. 169–174, Jan. 2014.
[25] A. H. Sayed, "A framework for state-space estimation with uncertain models," *IEEE Trans. Autom. Control*, vol. 46, no. 7, pp. 998–1013, Jul. 2001.
[26] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust distributed estimation by networked agents," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3909–3921, Aug. 2017.
[27] K. Chen, V. Gupta, and Y.-F. Huang, "Minimum variance unbiased estimation in the presence of an adversary," in *Proc. IEEE 56th Annu. Conf. Decis. Control (CDC)*, Melbourne, Australia, Dec. 2017, pp. 151–156.
[28] Y. Lin and A. Abur, "Robust state estimation against measurement and network parameter errors," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4751–4759, Sep. 2018.
[29] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3205–3216, Jul. 2017.
[30] A. Fredriksen, D. Middleton, and V. VandeLinde, "Simultaneous signal detection and estimation under multiple hypotheses," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 5, pp. 607–614, Sep. 1972.
[31] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.
[32] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, Jan. 2013.
[33] X. Yuzhe, V. Gupta, and C. Fischione. (2012). *Distributed Estimation*. [Online]. Available: https://www3.nd.edu/~vgupta2/research/publications/xgf13.pdf
[34] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.
[35] Y. Zhou and H. Leung, "Minimum entropy approach for multisensor data fusion," in *Proc. IEEE Signal Process. Workshop Higher-Order Statist.*, Banff, AB, Canada, Jul. 1997, pp. 336–339.
[36] W. Niu, J. Zhu, W. Gu, and J. Chu, "Four statistical approaches for multisensor data fusion under non-Gaussian noise," in *Proc. IITA Int. Conf. Control, Autom. Syst. Eng. (CASE)*, Zhangjiajie, China, Jul. 2009, pp. 27–30.
[37] S. Bandyopadhyay and S.-J. Chung, "Distributed estimation using Bayesian consensus filtering," in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 634–641.
[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[39] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2451–2464, Oct. 2013.
[40] M. Loeve, *Probability Theory II*, 4th ed. New York, NY, USA: Springer, 1978.

Saurabh Sihag (Student Member, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from IIT Kharagpur, India, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests include statistical signal processing, information theory, and high-dimensional statistics.

Ali Tajer (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Columbia University in 2010. He is currently an Associate Professor of electrical, computer, and systems engineering with the Rensselaer Polytechnic Institute. From 2010 to 2012, he was a Post-Doctoral Research Associate with Princeton University. His research interests include mathematical statistics and network information theory, with applications in wireless communications and power grids. He serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and in the past has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, an Editor for the IEEE TRANSACTIONS ON SMART GRID, a Guest Editor-in-Chief for the IEEE TRANSACTIONS ON SMART GRID, and a Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE. He is a senior member of the IEEE and received an NSF CAREER Award in 2016.