Chapter 9
Secure Estimation Under Model Uncertainty

Saurabh Sihag and Ali Tajer

9.1 Introduction

Cyber-physical systems are deployed in a variety of technical domains such as critical infrastructure, healthcare devices, and transportation. The rapid rise in their applications has exposed them to different vulnerabilities, threats, and attacks (Humayed et al. 2017). An abstract representation consisting of three main components: monitoring, communications, and computation and control, captures the fundamental aspects of cyber-physical systems. The monitoring component observes the environment and communicates with the computation and control component, which in turn processes the observations to form and communicate decisions. Each of these components could potentially be exploited or compromised, causing unexpected behaviors and compromised integrity and performance for the system.

The source of security threats to a cyber-physical system can broadly be categorized into three groups: an attacker with a malicious intent, functional failure of components in the system, and environmental threats such as natural disasters. While the impacts of operational failures of the system due to environmental threats or internal failures can be minimized by robust strategies (Hu et al. 2016), malicious attacks on cyber-physical systems intend to deceive the controller into making highly damaging decisions via well-crafted adversarial strategies. Therefore, specialized security measures are required to mitigate such attacks (Li et al. 2020).

Adversarial attacks that exploit the vulnerabilities of the inference and control algorithms deployed in the cyber-physical systems and potential defense strategies against them have been subjects of active research (Li et al. 2020; Fawzi et al. 2014;

S. Sihag
University of Pennsylvania, Philadelphia, PA 19104, USA
e-mail: saurabh.sihag@pennmedicine.upenn.edu

A. Tajer (✉)
Rensselaer Polytechnic Institute, Troy, NY 12108, USA
e-mail: tajer@ecse.rpi.edu

M. Abbaszadeh and A. Zemouche (eds.), *Security and Resilience in Cyber-Physical Systems*, https://doi.org/10.1007/978-3-030-97166-3_9

Ahmed et al. 2021). The taxonomy of the adversarial attacks on cyber-physical systems can be specified along three axes. The first axis pertains to the influence of the attack, where the attacker is capable of probing the algorithms for vulnerabilities. The attacker can further leverage these vulnerabilities to impose false decisions or outcomes in the system. The second axis pertains to the specificity of the attack, i.e., the attack can be either indiscriminate and affect all decisions made by the system, or targeted to impose false decisions only in specific scenarios. The third axis is related to the violation induced by the attack, where the attack can distort the integrity of the decisions made by the system in specific scenarios or overwhelm the system with malicious inputs, thus rendering it incapable of making any decision (for instance, through denial of service attacks).

In this chapter, we design a statistical inference framework for systems vulnerable to adversarial attacks. Statistical inference leverages the data sampled from a population to deduce its statistical properties. The commonly studied modes of statistical inference are broadly focused on discerning the statistical model of the population or estimating unknown, underlying parameters that characterize the statistical model of the population. Vulnerability to an attack induces uncertainties in the inference decisions, and therefore, must be accounted for in the design of inference algorithms that are resilient to adversarial attacks.

9.1.1 Overview and Contributions

We start by laying the context for the problem studied in this chapter. For this purpose, we consider the canonical parameter estimation problem in which the objective is to estimate a stochastic parameter X, which lies in a known set $\mathscr{X} \subseteq \mathbb{R}^p$, from the data samples $\mathbf{Y} \triangleq [Y_1, \ldots, Y_n]$, where the sample Y_r is distributed according to a statistical model with probability density function (pdf) P_X and lies in a known set $\mathscr{Y} \subseteq \mathbb{R}^m$. In practice, the dimension of the data points m could correspond to the number of data collecting entities in the system. Furthermore, the statistician *assumes* a prior data model for X and Y_r, determined through historical data. We denote the assumed underlying pdfs for X and $\mathbf{Y}$ by π and $f(\cdot \mid X)$, respectively, i.e.,

$$\mathbf{Y} \sim f(\cdot \mid X), \qquad \text{with} \quad X \sim \pi. \tag{9.1}$$

For our analysis, we assume that the pdfs do not have any non-zero probability masses over lower-dimensional manifolds. The objective of the statistician is to formalize a reliable estimator

$$\hat{X}(\mathbf{Y}) : \mathscr{Y}^n \mapsto \mathscr{X}. \tag{9.2}$$

For elaborate discussions on the design of statistical estimators, we refer the readers to Poor (1998). In an adversarial environment, the attacker may launch an attack on

different components of the data model defined in (9.1) to degrade the quality of $\hat{X}(\mathbf{Y})$. Next, we discuss two specific adversarial attack scenarios.

False data injection attacks: The purpose of false data injection attacks is to distort the data samples $\mathbf{Y}$ such that the data model deviates from (9.1) for at least a subset of coordinates in $\mathbf{Y}$.

Causative attacks: The purpose of a causative attack is to compromise the *process that underlies acquiring the statistical models* in (9.1). We emphasize that such an attack is different from false data injection attack because 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 specifying a model for the data.

We remark that the nature of security vulnerabilities that inference algorithms are exposed to in causative attacks is fundamentally distinct from that of the data that faces false data injection attacks. Specifically, in the case of a false data injection attack, the information of the decision algorithm about the data model remains intact, while the data fed to the algorithm is anomalous. Therefore, when the sampled data is compromised, an inference algorithm produces decisions based on the true model for the data in the attack-free scenario, while the data that it receives and processes are compromised. On the other hand, when the historical data leveraged by the statistician to determine the true model are compromised, an inference algorithm functions based on an incorrect model for the data, in which case even un-compromised sampled data produces unreliable decisions. Both attack scenarios mentioned above force the inference algorithm to deviate from its optimal structure and, if not mitigated, may produce decisions that serve the adversary's purposes.

Depending on the specificity and the extent of an adversarial attack, e.g., the fraction of the observed data or training data that is compromised, the true model $f(\cdot \mid X)$ can be assumed to deviate to the space of alternative data models, which we denote by $\mathscr{F}$. The attack can be characterized by alterations in the statistical distributions of any number of the m coordinates of $\mathbf{Y}$. There are two major aspects of selecting $\mathscr{F}$ as a viable model space.

- An attack is effective in degrading the quality of estimation if the compromised model is sufficiently distinct from the model assumed by the statistician for designing the estimator. Hence, even though, in general, $\mathscr{F}$ can be thought of as any representation of possible kernels $f(\cdot \mid X)$ mapping $\mathscr{Y}$ to $\mathbb{R}^m$, only a subset of such mappings pertain to the set of effective attacks.
- There exists a tradeoff between the complexity of the model space and its expressiveness. Specifically, an overly expressive space can represent the possible compromised models with a more refined accuracy, albeit at the expense of more complex statistical inference rules.

We will discuss the specifics of the attack model in Sect. 9.2. Note that the potential adversarial presence induces a new dimension to the estimation problem in (9.2). Specifically, the optimal estimator design hinges on the knowledge of the true statistical model of the measurements $\mathbf{Y}$. However, detecting whether the data model has

been compromised and discerning the true model, itself being an inference task, is never perfect. These observations imply an inherent coupling between the original estimation problem of interest and the introduced auxiliary problem due to potential adversarial behavior (i.e., detecting the presence of an attacker and isolating the true model). Therefore, the quality of the estimator is expected to degrade with respect to an attack-free setting due to uncertainties in the true model in the adversarial setting. Our objective is to characterize the fundamental interplay between the quality of discerning the true model and the degradation in the estimation quality.

9.1.2 Related Studies

The problem of secure inference is studied primarily in the context of sensor networks, where a subset of sensors may be corrupted by an attacker. The study in Wilson and Veeravalli (2016), in particular, considers the problem of secure estimation in a two-sensor network, in which one sensor is assumed to be secured , and the other sensor is vulnerable to attacks. According to the heuristic estimation design in this context, first, a decision is formed on the attacker's activity on the unsecured sensor . If it is deemed to be attacked, then the estimation design relies only on the secured sensor , and otherwise, it uses the data collected at both sensors. In contrast to Wilson and Veeravalli (2016), we consider a model with an arbitrary dimension of data, assume that all data coordinates are vulnerable to the attack, and characterize the optimal secure inference structure, which is distinct from being a detection -driven design studied in Wilson and Veeravalli (2016).

The adversarial setting considered in this chapter has similarities with the widely-investigated Byzantine attack models in sensor networks. In Byzantine attack models, the data corresponding to the compromised sensors is modified arbitrarily by the adversaries with an aim to degrade the inference quality. The impact of Byzantine attacks on the quality of inference and relevant mitigation strategies in sensor networks are discussed in Vempaty et al. (2013). Various detection-driven estimation strategies (i.e., when attack detection precedes and guides the estimation routine) for scenarios where the impacts of the Byzantine attacks on data are characterized by randomly flipped information bits, are discussed in Vempaty et al. (2013), Ebinger and Wolthusen (2009), Zhang et al. (2015), Zhang and Blum (2014). Furthermore, attack-resilient target localization strategies are studied in Vempaty et al. (2013, 2014), where the assumption is that the attacker adopts a fixed strategy that leads to maximum disruption in the inference. In these studies, however, an attacker can deviate from the worst-case attack strategy of incurring the maximum damage, and launch a less impactful but sustained attack, which may remain undetected. Finally, various strategies for isolating the compromised sensors in sensor networks are studied in Rawat et al. (2010), Soltanmohammadi et al. (2013), Vempaty et al. (2011). The emphasis of these studies is primarily detection of attacks or isolating the attacked sensors, whereas this chapter focuses on parameter estimation.

Secure estimation in linear *dynamical* systems that characterize cyber-physical systems has been actively studied in recent years (Fawzi et al. 2011, 2014; Yong et al. 2015; Pajic et al. 2014, 2015; Shoukry et al. 2017; Mishra et al. 2015). The studies with more relevance to the scope of this chapter include Fawzi et al. (2014), Mishra et al. (2015), and Pajic et al. (2014), which investigate robust estimation in dynamic systems. Specifically, a coding-theoretic interplay between the number of sensors compromised by an adversary and the guarantees on perfect system state recovery are characterized in Fawzi et al. (2014), a Kalman filter-based approach for identifying the most reliable set of sensors for inference is investigated in Mishra et al. (2015), and the design of estimators that is robust in the presence of dynamical model uncertainty is studied in Pajic et al. (2014). Furthermore, the degradation impact on estimation performance in a dynamical system consisting of a single sensor network is investigated from the adversary's perspective in Bai and Gupta (2014), where bounds on the degradation in estimation quality with the stealthiness of the attacker are characterized.

Secure estimation is also linked to robust estimation (Shen et al. 2014; Sayed 2001; Al-Sayed et al. 2017; Chen et al. 2017; Lin and Abur 2020; Zhao et al. 2016). These two problems share some aspects (e.g., data model uncertainty), but their inference tasks are distinct. Specifically, besides the estimation objective, both problems also face the problem of resolving uncertainties about the data model. The main distinction between secure estimation and robust estimation lies in their resolution of the model uncertainties, which results in significant differences in the formulation of the problems and the designs of the optimal decision rules. Specifically, in robust estimation , the emphasis is laid on forming the most reliable estimates, and as an intermediate step, the model uncertainty must also be resolved as a second inference task. Resolution of model uncertainties can be executed by a wide range of approaches, which include averaging out the effect of the model or forming an estimate of the model. The ultimate objective of robust estimation is optimizing the estimation quality, and it generally does not account for the quality of the decisions involved in resolving model uncertainty, i.e., model uncertainty resolution will be dictated by the decision rules optimized for producing the best estimates.

The aforementioned studies that study secure estimation, despite their discrepancies, conform to an underlying design principle, which decouples the estimation design from all other decisions involved (e.g., attack detection or attacked sensor isolation), and leads to either detection -driven estimators or estimation -driven detection routines. The sub-optimality of decoupling such intertwined estimation and detection problems into independent estimation and detection routines is well-investigated (Middleton and Esposito 1968; Zeitouni et al. 1992; Moustakides et al. 2012; Jajamovich et al. 2012). In contrast, in secure estimation, our focus is on 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 problem formulation is motivated by our recent work in Sihag and Tajer (2020), which emphasizes the natural coupling between the two inference tasks and requires that the optimal decisions are determined jointly.

9.2 Data Model and Definitions

Our focus is on the estimation problem in (9.2) and in this context, we discuss the data models under the attack-free and adversarial scenarios.

9.2.1 Attack Model

The objective is to form an optimal estimate $\hat{X}(\mathbf{Y})$ (under the general cost functions specified in Sect. 9.2.2) in the potential presence of an adversary. In the attack-free setting, the data is assumed to be generated according to a known model specified in (9.1). In an adversarial setting, an adversary, depending on its strength and desired impact, can launch an attack with the ultimate purpose of degrading the quality of the estimate 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 exist $T = \sum_{i=1}^{K} \binom{m}{i}$ number of attack scenarios, each of which is associated with a distinct data model. To formalize this, we define $\mathscr{S} \triangleq \{S_1, \ldots, S_T\}$ as the set of all possible attack scenarios, where $S_i \subseteq \{1, \ldots, m\}$ describes the set of coordinates of $\mathbf{Y}$ the models of which are compromised under attack scenario $i \in \{1, \ldots, T\}$.

Under the attack scenario $i \in \{1, \ldots, T\}$, if $r \in S_i$, the data model deviates from f to a model in the space $\mathscr{F}_i$. Clearly, the attack can be effective if it encompasses sufficiently distinct models. For our analysis, we assume that $\mathscr{F}_i \triangleq \{f_i(\cdot \mid X)\}$, i.e., $\mathscr{F}_i$ 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 $f_i(\cdot \mid X)$.

In practice, the resources and preferences of the attacker may determine the likelihood of an attack scenario. For instance, attacking one coordinate may be easier or more desirable as compared to others. To account for such likelihoods, we adopt a Bayesian framework in which we define ε_0 as the prior probability of an attack-free scenario and define ε_i as the prior probability of the event that the attacker compromises the data at coordinates specified by S_i. A block diagram of the attack model and the inferential goals is depicted in Fig. 9.1.

9.2.2 Decision Cost Functions

In the adversarial setting, the estimation decision is intertwined with the decision on the true model, and therefore, it constantly faces the uncertainty induced by the action or inaction of the adversary. A decoupled strategy of decisions for isolating the model and estimating the parameter under the isolated model does not generally guarantee optimal performance. In fact, there exist extensive studies on formalizing and analyzing such compound decisions, which generally aim to decouple the

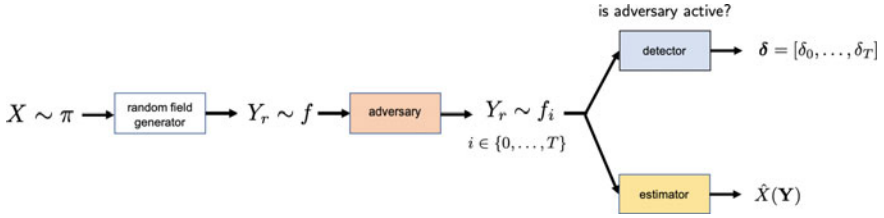


Fig. 9.1 The effect of the adversary on the data model, and the inferential decisions involved. Depending on the adversarial action, the data model may either deviate from f to one among the alternative data models ($\{f_i : i \in \{0, \ldots, T\}\}$) or retain the original data model (given by f_0)

inferential decisions. For instance, in Zeitouni et al. (1992), it is shown that the generalized likelihood ratio test (GLRT), which uses maximum likelihood estimates of unknown parameters in its decision rule, is not always optimal. In Moustakides et al. (2012) and Jajamovich et al. (2012), non-asymptotic frameworks for optimal joint detection and estimation are provided. Specifically, in Moustakides et al. (2012), a binary hypothesis testing problem is studied in a setting where one hypothesis is composite and consists of an unknown parameter to be estimated. In Jajamovich et al. (2012), the principles in Moustakides et al. (2012) are extended to a composite binary hypothesis testing problem in which both hypotheses correspond to composite models. We used similar principles as established in Moustakides et al. (2012) and Jajamovich et al. (2012) in our recent study on secure estimation in Sihag and Tajer (2020). We borrow the principles adopted in Sihag and Tajer (2020) to discuss secure estimation in the context of cyber-physical systems in this chapter. We next discuss the cost functions for true model detection and estimation quality.

9.2.2.1 Attack Detection Costs

Due to the existence of multiple attack scenarios, the true model detection problem can be formulated as the following $(T+1)$-composite hypothesis testing problem.

$$\begin{aligned} \mathsf{H}_0 &: \mathbf{Y} \sim f(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \mathsf{H}_i &: \mathbf{Y} \sim f_i(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X)\,, \quad \text{for } i \in \{1, \ldots, T\}, \end{aligned} \tag{9.3}$$

where H_0 is the hypothesis that represents the attack-free setting, and H_i is the hypothesis corresponding to an attack scenario where the attack is launched at the coordinates in $S_i \in \mathscr{S}$. For the convenience in notation, we denote the attack-free data model by $f_0(\cdot \mid X)$, i.e., $f_0(\cdot \mid X) = f(\cdot \mid X)$. To formalize relevant costs for the detection decisions, we define $\mathsf{D} \in \{\mathsf{H}_0, \ldots, \mathsf{H}_T\}$ as the decision on the hypothesis testing problem in (9.3), and $\mathsf{T} \in \{\mathsf{H}_0, \ldots, \mathsf{H}_T\}$ as the true hypothesis. The true hypothesis is discerned via a general *randomized* test $\boldsymbol{\delta}(\mathbf{Y}) \triangleq [\delta_0(\mathbf{Y}), \ldots, \delta_T(\mathbf{Y})]$, where $\delta_i(\mathbf{Y}) \in [0, 1]$ denotes the probability of deciding in favor of H_i. Clearly

$$\sum_{i=0}^{T} \delta_i(\mathbf{Y}) = 1. \tag{9.4}$$

Hence, the probability of forming a decision in favor of H_j while the true model is H_i is given by

$$\mathbb{P}(\mathsf{D}=\mathsf{H}_j \,|\, \mathsf{T}=\mathsf{H}_i) = \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{9.5}$$

We define P_{md} as the aggregate probability of error in identifying the true model when there exist compromised data coordinates due to attacker's activity, i.e.,

$$\begin{aligned}
\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) &\triangleq \mathbb{P}(\mathsf{D} \neq \mathsf{T} \mid \mathsf{T} \neq \mathsf{H}_0) \\
&= \frac{1}{\mathbb{P}(\mathsf{T} \neq \mathsf{H}_0)} \sum_{i=1}^{T} \mathbb{P}(\mathsf{D} \neq \mathsf{H}_i \mid \mathsf{T} = \mathsf{H}_i)\mathbb{P}(\mathsf{T} = \mathsf{H}_i) \qquad (9.6) \\
&= \sum_{i=1}^{T} \frac{\varepsilon_i}{1-\varepsilon_0} \cdot \mathbb{P}(\mathsf{D} \neq \mathsf{H}_i \mid \mathsf{T} = \mathsf{H}_i). \qquad (9.7)
\end{aligned}$$

Furthermore, we define P_{fa} as the aggregate probability of erroneously deciding that a set of coordinates is compromised while operating in an attack-free scenario. In this context, we have

$$\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta}) \triangleq \mathbb{P}(\mathsf{D} \neq \mathsf{H}_0 \mid \mathsf{T} = \mathsf{H}_0) = \sum_{i=1}^{T} \mathbb{P}(\mathsf{D}=\mathsf{H}_i \,|\, \mathsf{T}=\mathsf{H}_0). \tag{9.8}$$

9.2.2.2 Secure Estimation Costs

In this subsection, we discuss the estimation cost functions that capture the quality of the estimate $\hat{X}(\mathbf{Y})$. For this purpose, we adopt a generic and *non-negative* cost function $\mathsf{C}(X, U(\mathbf{Y}))$ that quantifies the discrepancy between the ground truth X and a generic estimator $U(\mathbf{Y})$. Since the data models under different attack scenarios are distinct, we consider having possibly distinct estimators under each attack scenario. Therefore, we denote the estimate of X under model H_i by $\hat{X}_i(\mathbf{Y})$, and accordingly, we define

$$\hat{\mathbf{X}}(\mathbf{Y}) \triangleq [\hat{X}_0(\mathbf{Y}), \ldots, \hat{X}_T(\mathbf{Y})]. \tag{9.9}$$

Therefore, the estimation cost $\mathsf{C}(X, \hat{X}_i(\mathbf{Y}))$ is relevant only if the decision is H_i. Hence, for a generic estimator $U_i(\mathbf{Y})$ of X under model H_i, we define the *decision-specific average cost function* as

$$J_i(\delta_i, U_i(\mathbf{Y})) \triangleq \mathbb{E}_i[\mathsf{C}(X, U_i(\mathbf{Y})) \mid \mathsf{D} = \mathsf{H}_i], \qquad \forall i \in \{0, \ldots, T\} \tag{9.10}$$

where the conditional expectation is with respect to X and $\mathbf{Y}$. Accordingly, we leverage (9.10) to define an aggregate average estimation cost according to

$$J(\boldsymbol{\delta}, \mathbf{U}) \triangleq \max_{i \in \{0,\ldots,T\}} J_i(\delta_i, U_i(\mathbf{Y})), \tag{9.11}$$

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

$$J_0(V) = \mathbb{E}[\mathsf{C}(X, V(\mathbf{Y}))], \tag{9.12}$$

where the expectation is with respect to X and $\mathbf{Y}$ under model f. Note that J_0 defined in (9.12) corresponds to the scenario in which the attack-free model f is the only possibility for the data model and is, therefore, fundamentally different from $J(\boldsymbol{\delta}, \mathbf{U})$ defined in (9.11). In the analysis, J_0 furnishes a baseline to assess the impact of potential adversarial action on the estimation quality.

9.3 Secure Parameter Estimation

In this section, we formalize the problem of secure estimation. There exists an inherent interplay between the quality of estimating X and the quality of isolation decision to identify the true model governing the data. On the one hand, detecting the adversary's attack model perfectly is not possible. At the same time, the estimation quality critically hinges on the successful isolation of the true data model. Therefore, an imperfection in the decision about the data model is expected to degrade the estimation quality with respect 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 9.1 (*Estimation Degradation Factor*) For a given estimator V in the attack-free scenario, and a secure estimation framework specified by the rules $(\boldsymbol{\delta}, \mathbf{U})$ in the adversarial scenario, we define the estimation degradation factor (EDF) as

$$q(\boldsymbol{\delta}, \mathbf{U}, V) \triangleq \frac{J(\boldsymbol{\delta}, \mathbf{U})}{J_0(V)}. \tag{9.13}$$

Based on Definition 9.1, we define the performance region for secure estimation that encompasses all the pairs of estimation quality $q(\boldsymbol{\delta}, \mathbf{U}, V)$ and detection performance $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$ over the space characterized by all possible decision rules $(\boldsymbol{\delta}, \mathbf{U}, V)$.

Definition 9.2 (*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 $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$.

Next, we leverage the definition of performance region to define the notion of (q, β)-security , which is instrumental for formalizing the secure estimation problem. For this purpose, we first note that the two estimation cost functions involved in the EDF $q(\boldsymbol{\delta}, \mathbf{U}, V)$ can be computed independently, and as a result, their attendant decision rules can be determined independently. For this purpose, we define V^* as the optimal estimator under the attack-free scenario, and J_0^* as the corresponding estimation cost, i.e.,

$$V^* \triangleq \arg\min_{V} J_0(V), \qquad \text{and} \qquad J_0^* \triangleq \min_{V} J_0(V). \tag{9.14}$$

Definition 9.3 ((q, β)-*security*) In the adversarial scenario, an estimation procedure specified by $(\boldsymbol{\delta}, \mathbf{U}, V^*)$ is called (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.,

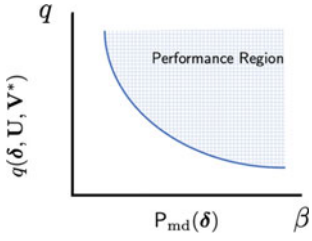
$$q \triangleq \min_{\boldsymbol{\delta}, \mathbf{U}} q(\boldsymbol{\delta}, \mathbf{U}, V^*), \qquad \text{s.t.} \qquad \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) \leq \beta. \tag{9.15}$$

The performance region, and its boundary that specifies the interplay between q and β are illustrated figuratively in Fig. 9.2. Based on the definitions in this subsection, we aim to characterize the region of all simultaneously achievable values of $q(\boldsymbol{\delta}, \mathbf{U}, V^*)$ and $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$ (represented by the dashed region in Fig. 9.2) and the (q, β)-secure decision rules that solve (9.15), and specify the boundary of the performance region (illustrated by a solid line as the boundary of the performance region in Fig. 9.2).

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

$$\mathscr{Q}(\beta) \triangleq \begin{cases} \min_{\boldsymbol{\delta}, \mathbf{U}} & J(\boldsymbol{\delta}, \mathbf{U}) \\ \text{s.t.} & \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) \leq \beta \end{cases}. \tag{9.16}$$

Fig. 9.2 Performance region



We note that although $\mathscr{Q}(\beta)$ ensures that the likelihood of missing an attack is confined below β, it is insensitive to the rate of the false alarms, that is, the rate of erroneously declaring an attack when there is no attack. If it is also desirable to control the rate of false alarms, we can further extend the notion of (q, β)-security as follows.

Definition 9.4 An estimation procedure is (q, α, β)-secure if it is (q, β)-secure and the likelihood of false alarms does not exceed $\alpha \in (0, 1]$.

The (q, α, β)-secure decisions are determined by the optimal decision rules that form the solution to

$$\mathscr{P}(\alpha, \beta) = \begin{cases} \min_{\boldsymbol{\delta}, \mathbf{U}} & J(\boldsymbol{\delta}, \mathbf{U}) \\ \text{s.t.} & \mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) \leq \beta \\ & \mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta}) \leq \alpha \end{cases}. \tag{9.17}$$

Remark 9.1 It is straightforward to verify that $\mathscr{Q}(\beta) = \mathscr{P}(1, \beta)$.

Remark 9.2 (*Feasibility*) The Neyman–Pearson theory (Poor 1998) dictates that the probabilities $\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta})$ and $\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta})$ cannot be made arbitrarily small simultaneously. Specifically, for any given α, there exists a smallest feasible value for β, denoted by $\beta^*(\alpha)$.

We provide the optimal solution to problems $\mathscr{P}(\alpha, \beta)$ and $\mathscr{Q}(\beta)$ in closed-forms in Sect. 9.4.

9.4 Secure Parameter Estimation: Optimal Decision Rules

In this section, we characterize an optimal solution to the general problem $\mathscr{P}(\alpha, \beta)$ to determine the designs for 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 first leverage the expansions of the 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. Based on (9.5) and (9.6), we have

$$\mathsf{P}_{\mathsf{md}}(\boldsymbol{\delta}) = \sum_{i=1}^{T} \frac{\varepsilon_i}{1 - \varepsilon_0} \sum_{\substack{j=0 \\ j \neq i}}^{T} \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{9.18}$$

Similarly, by noting (9.5) and based on (9.8), we have

$$\mathsf{P}_{\mathsf{fa}}(\boldsymbol{\delta}) = \sum_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \, \mathrm{d}\mathbf{Y}. \tag{9.19}$$

By using the expansions in (9.18) and (9.19), the equivalent problem to (9.17) is given by

$$\mathscr{P}(\alpha,\beta)=\begin{cases}\min_{(\boldsymbol{\delta},\boldsymbol{U})} & J(\boldsymbol{\delta},\mathbf{U})\\ \text{s.t.} & \sum\limits_{i=1}^{T}\frac{\varepsilon_i}{1-\varepsilon_0}\sum\limits_{\substack{j=0\\ j\neq i}}^{T}\int_{\mathbf{Y}}\delta_j(\mathbf{Y})f_i(\mathbf{Y})\,\mathrm{d}\mathbf{Y}\leq\beta\\ & \sum\limits_{i=1}^{T}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_0(\mathbf{Y})\,\mathrm{d}\mathbf{Y}\leq\alpha\end{cases}. \tag{9.20}$$

Note that the estimators $\{U_i(\mathbf{Y}) : i \in \{0,\dots,T\}\}$ are restricted to the utility function $J(\boldsymbol{\delta},\mathbf{U})$, which allows us to decouple the problem $\mathscr{P}(\alpha,\beta)$ into two sub-problems, formalized next.

Theorem 9.1 *The optimal secure estimators of X under different models, i.e., $\hat{\mathbf{X}}=[\hat{X}_0,\dots,\hat{X}_T]$ are the solutions to*

$$\hat{\mathbf{X}}=\arg\min_{\mathbf{U}} J(\boldsymbol{\delta},\mathbf{U}). \tag{9.21}$$

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

$$\mathscr{P}(\alpha,\beta)=\begin{cases}\min_{\boldsymbol{\delta}} & J(\boldsymbol{\delta},\hat{\mathbf{X}})\\ s.t. & \sum\limits_{i=1}^{T}\frac{\varepsilon_i}{1-\varepsilon_0}\sum\limits_{\substack{j=0\\ j\neq i}}^{T}\int_{\mathbf{Y}}\delta_j(\mathbf{Y})f_i(\mathbf{Y})\,\mathrm{d}\mathbf{Y}\leq\beta\\ & \sum\limits_{i=1}^{T}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_0(\mathbf{Y})\,\mathrm{d}\mathbf{Y}\leq\alpha\end{cases}. \tag{9.22}$$

By leveraging the design in (9.21) and the decoupled structure of the problem $\mathscr{P}(\alpha,\beta)$ in (9.22), in the following theorem, we discuss optimal designs for the estimators in the secure estimation problem.

Theorem 9.2 ((q,α,β)-secure Estimators) *For the optimal secure estimators $\hat{\mathbf{X}}$, we have:*

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

$$U_i^*(\mathbf{Y})\triangleq\arg\inf_{U_i(\mathbf{Y})}\mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y})\mid\mathbf{Y}), \tag{9.23}$$

where $\mathsf{C}_{\mathrm{p},i}(U(\mathbf{Y})\mid\mathbf{Y})$ is the average posterior cost function denoted by

$$\mathsf{C}_{\mathrm{p},i}(U(\mathbf{Y})\mid\mathbf{Y})\triangleq\mathbb{E}_i\left[\mathsf{C}(X,U(\mathbf{Y}))\mid\mathbf{Y}\right], \tag{9.24}$$

where the conditional expectation in (9.24) is with respect to X under model H_i.

2. *The optimal estimator* $\hat{\mathbf{X}} = [\hat{X}_0, \ldots, \hat{X}_T]$, *specified in* (9.21), *is given by*

$$\hat{X}_i(\mathbf{Y}) = U_i^*(\mathbf{Y}). \tag{9.25}$$

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

$$J(\boldsymbol{\delta}, \hat{\mathbf{X}}) = \max_{i \in \{0,\ldots,T\}} \left\{ \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}} \right\}, \tag{9.26}$$

where we have defined

$$\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) \triangleq \inf_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}). \tag{9.27}$$

Proof See Appendix 1. ■

We next discuss the application of decision rules in Theorem 9.2 in a specific example. Specifically, in the next corollary, we discuss the closed-forms of these decision rules when the distributions $\{f_i(\cdot \mid X) : i \in \{0, \ldots, T\}\}$ are Gaussian.

Corollary 9.1 ((q, α, β)-secure Estimators in Gaussian Models) *When the data models are Gaussian, i.e.,*

$$f_i(\cdot \mid X) \sim \mathscr{N}(\theta_i, X), \qquad \text{for} \quad \theta_i \in \mathbb{R} \tag{9.28}$$

such that the mean values are distinct, and

$$X \sim \mathscr{X}^{-1}(\zeta, \phi), \tag{9.29}$$

where $\mathscr{X}^{-1}(\zeta, \phi)$ *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 the mean squared error, given by*

$$\mathsf{C}(X, U(\mathbf{Y})) = \|X - U(Y)\|^2, \tag{9.30}$$

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*

$$U_i^*(\mathbf{Y}) = \frac{\zeta\phi + \sum_{r=1}^{n} \|Y_r - \theta_i\|_2^2}{\zeta + n - 2}. \tag{9.31}$$

2. *The optimal estimator* $\hat{\mathbf{X}} = [\hat{X}_0, \ldots, \hat{X}_T]$*, specified in* (9.21)*, is given by*

$$\hat{X}_i(\mathbf{Y}) = U_i^*(\mathbf{Y}). \tag{9.32}$$

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

$$J(\boldsymbol{\delta}, \hat{\mathbf{X}}) = \max_{i \in \{0,\ldots,T\}} \left\{ \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) d\mathbf{Y}} \right\}, \tag{9.33}$$

where we have

$$\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) = \frac{2(\zeta\phi + \sum_{r=1}^{n} \|Y_r - \theta_1\|^2)^2}{(\zeta_i + n - 2)^2(\zeta + n - 4)}. \tag{9.34}$$

Next, given the optimal estimators $\hat{\mathbf{X}}$, we provide the optimal detection rules in the next theorem. We note that the decision rules depend on the metrics computed based on the optimal estimation costs, establishing the coupling of estimation and true model detection decisions. We show that by using the solution of the specific auxiliary convex problem in a variational form in the next theorem, we can solve $\mathscr{P}(\alpha, \beta)$ in (9.22).

Theorem 9.3 *For any arbitrary* $u \in \mathbb{R}_+$*, we have* $\mathscr{P}(\alpha, \beta) \leq u$ *if and only if* $\mathbb{R}(\alpha, \beta, u) \leq 0$*, where we have defined*

$$\mathbb{R}(\alpha, \beta, u) = \begin{cases} \min_{\boldsymbol{\delta}} & \eta \\ s.t. & \displaystyle\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) [\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u] \, \mathrm{d}\mathbf{Y} \leq \eta, \quad \forall\, i \in \{0, \ldots, T\} \\ & \displaystyle\sum_{i=1}^{T} \frac{\varepsilon_i}{1 - \varepsilon_0} \sum_{\substack{j=0 \\ j \neq i}}^{T} \int_{\mathbf{Y}} \delta_j(\mathbf{Y}) f_i(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \beta + \eta \\ & \displaystyle\sum_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \, \mathrm{d}\mathbf{Y} \leq \alpha + \eta \end{cases}. \tag{9.35}$$

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

Proof See Appendix 2. ■

The point u^* plays a pivotal role in the structure of optimal detection decision rules. We define the constants $\{\ell_i : i \in \{0, \ldots, T+2\}\}$ as the dual variables in the Lagrange function for the convex problem $\mathbb{R}(\alpha, \beta, u^*)$. Given u^* and $\{\ell_i : i \in$

$\{0, \ldots, T+2\}\}$, we can characterize the optimal detection rules in closed-forms, as specified in the following theorem.

Theorem 9.4 ((q, α, β)-secure Detection Rules) *The optimal decision rules for isolating the compromised coordinates are given by*

$$\delta_i(\mathbf{Y}) = \begin{cases} 1, & \text{if } \ i = i^* \\ 0, & \text{if } \ i \neq i^* \end{cases}, \tag{9.36}$$

where we have defined

$$i^* \triangleq \underset{i \in \{0,\ldots,T\}}{\operatorname{argmin}} \ A_i \ . \tag{9.37}$$

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*

$$A_0 \triangleq \ell_0 f_0(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},0}(\mathbf{Y}) - u^*] + \ell_{T+1} \sum_{i=1}^{T} \frac{\varepsilon_i}{1-\varepsilon_0} f_i(\mathbf{Y}), \tag{9.38}$$

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

$$A_i \triangleq \ell_i f_i(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u^*] + \ell_{T+1} \sum_{j=1, j\neq i}^{T} \frac{\varepsilon_j}{1-\varepsilon_0} f_j(\mathbf{Y}) + \ell_{T+2} f_0(\mathbf{Y}). \tag{9.39}$$

Proof See Appendix 3. ■

In the next corollary, we discuss the application of these decision rules when the distributions $\{f_i(\cdot \mid X) : i \in \{0, \ldots, T\}\}$ are all Gaussian.

Corollary 9.2 ((q, α, β)-secure Detection Rules in Gaussian Models) *When the data models* $\{f_i(\cdot \mid X) : i \in \{0, \ldots, T\}\}$ *have the following Gaussian distributions*

$$f_i(\cdot \mid X) \sim \mathscr{N}(\theta_i, X) \ , \qquad \textit{for} \ \ \theta_i \in \mathbb{R} \tag{9.40}$$

where the mean values are distinct, and

$$X \sim \mathscr{X}^{-1}(\zeta, \phi), \tag{9.41}$$

the optimal decision rules for isolating the compromised coordinates are given by

$$\delta_i(\mathbf{Y}) = \begin{cases} 1, & \text{if } \ i = i^* \\ 0, & \text{if } \ i \neq i^* \end{cases}, \tag{9.42}$$

where we have defined

$$i^* \triangleq \underset{i\in\{0,\ldots,T\}}{\operatorname{argmin}} A_i. \tag{9.43}$$

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*

$$A_0 \triangleq \ell_0 f_0(\mathbf{Y})(\mathsf{C}^*_{\mathrm{p},0}(\mathbf{Y}) - u^*) + \ell_{T+1} \sum_{i=1}^{T} \frac{\varepsilon_i}{1-\varepsilon_0} f_i(\mathbf{Y}), \tag{9.44}$$

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

$$A_i \triangleq \ell_i f_i(\mathbf{Y})(\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u^*) + \ell_{T+1} \sum_{\substack{j=1 \\ j\neq i}}^{T} \frac{\varepsilon_j}{1-\varepsilon_0} f_j(\mathbf{Y}) + \ell_{T+2} f_0(\mathbf{Y}). \tag{9.45}$$

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* (9.34)*, we obtain*

$$f_i(\mathbf{Y}) = \frac{(\zeta\phi)^{\frac{\zeta}{2}}}{\pi^{\frac{n}{2}}\Gamma(\zeta/2)} \cdot \frac{\Gamma(\zeta+n)/2}{(\zeta\phi + \sum_{r=1}^{n} \|Y_r - \theta_i\|^2)^{\frac{\zeta+n}{2}}}. \tag{9.46}$$

Figure 9.3 illustrates the performance region and the corresponding (q, β)-security curve for the case $T = 1$, $n = 1$, $\theta_0 = 0$, $\theta_1 = 2$, $\zeta = 4$, and $\phi = 1$. The (q, β)-security curve in Fig. 9.3 depicts the tradeoff between the quality of the true model detection and the degradation in the estimation quality. Note that this tradeoff is inherently due to secure estimation problem formulation. Essentially, the design of the problem $\mathscr{P}(\alpha, \beta)$ as specified in (9.17) enables the trade of the quality of detection in favor of improving the estimation cost.

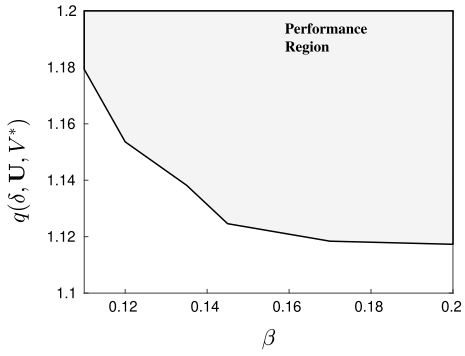


Fig. 9.3 Performance region for the Gaussian data model

We provide Algorithm 9.1, which summarizes all the steps for solving $\mathscr{P}(\alpha, \beta)$ for any feasible pair of α and β, and it encapsulates the decision rules specified by the theorems in this section and the detailed steps of specifying the parameters involved in characterizing the decision rules.

Algorithm 9.1 – Solving $\mathscr{P}(\alpha, \beta)$

Input: α and β and evaluate $\beta^*(\alpha)$
if $\beta < \beta^*(\alpha)$ **then**
  $\mathscr{P}(\alpha, \beta)$ not feasible for given choice of α and β;
  break;
else
  Initialize $u_0 = 0, u_1$;
  Evaluate optimal posterior estimation costs in (9.27);
  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 9.4;
      Compute $M(\hat{\boldsymbol{\ell}}) \triangleq \mathbb{R}(\alpha, \beta, \hat{u})$;
      if $\min_{\hat{\boldsymbol{\ell}}} M(\hat{\boldsymbol{\ell}}) \leq 0$ **then**
        $u_1 \leftarrow \hat{u}, \quad \boldsymbol{\ell} \leftarrow \hat{\boldsymbol{\ell}}$;
      else
        $u_0 \leftarrow \hat{u}$;
  until $u_1 - u_0 \leq \varepsilon$, *for* ε *sufficiently small*;
  $\mathscr{P}(\alpha, \beta) \leftarrow u^* = u_1$;
  Output: Decision rules $\boldsymbol{\delta}$

9.5 Case Studies: Secure Estimation in Sensor Networks

We evaluate the secure estimation framework using the example of a two-sensor network with a fusion center (FC). Each sensor collects a stream of data consisting of n samples. 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 scenario follows the model

$$Y_j^i = h^i X + N_j^i, \tag{9.47}$$

where h^i models the channel connecting sensor i to the FC and N_j^i accounts for the additive channel 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 (9.47) and evaluate the optimal performance.

9.5.1 Case 1: One Sensor Vulnerable to Causative Attacks

We first consider an adversarial setting in which the data model from only one sensor (sensor 1) is vulnerable to an adversarial attack while the other sensor (sensor 2) is secured. Under this setting, we clearly have only one attack scenario, i.e., $T = 1$ and $S_1 = \{1\}$. Accordingly, we have $\varepsilon_0 + \varepsilon_1 = 1$. Under the attack-free scenario, the noise terms N_j^i are distributed according to $\mathcal{N}(0, \sigma_n^2)$, i.e.,

$$Y_j^i \mid X \sim \mathcal{N}(h^i X, \sigma_n^2). \tag{9.48}$$

When sensor 1 is compromised, the actual conditional distribution of $Y_j^1|X$ is distinct from the above distribution. The inference objective under such a setting, in principle, becomes similar to the adversarial setting of Wilson and Veeravalli (2016), which focuses on data injection attack. Hence, for comparison with the performance of the secure estimation framework with that of Wilson and Veeravalli (2016), we assume that the conditional distribution of $Y_j^1|X$ when sensor 1 is under 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. Therefore, the composite hypothesis test for estimating X and discerning the model in (9.3) simplifies to a binary test with the prior probabilities ε_0 and ε_1.

$$\begin{aligned} \mathsf{H}_0 &: \mathbf{Y} \sim f_0(\mathbf{Y} \mid X), \quad \text{with } X \sim \mathcal{N}(0, \sigma^2) \\ \mathsf{H}_1 &: \mathbf{Y} \sim f_1(\mathbf{Y} \mid X), \quad \text{with } X \sim \mathcal{N}(0, \sigma^2). \end{aligned} \tag{9.49}$$

Figure 9.4 shows the variations of the estimation quality, captured by q, versus the miss-detection rate β, where it is observed that the estimation quality improves monotonically with an increase in β, and it reaches its maximum quality as β approaches 1. This observation is in line with the analytic implications of the formulations of the secure parameter estimation problem in (9.16) and (9.17). A similar setting is studied in Wilson and Veeravalli (2016), 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 adversarial attacks where the attacker compromises the data by adding a uniformly distributed disturbance. Figure 9.4 also shows the comparison of the estimation quality of the secure estimation framework in this chapter, with that from the methodology in Wilson and Veeravalli (2016). In Wilson and Veeravalli (2016), the estimator is designed to obtain the most robust estimate corresponding to an optimal false alarm probability α^*, which, in turn, fixes the miss-detection error probability. Therefore, the framework in Wilson and Veeravalli (2016) does not provide the flexibility to change the miss-detection rate β.

The results presented in Fig. 9.4 correspond to $\sigma = 3$, $\sigma_n = 1$, $h^1 = 1$, $h^2 = 4$, $a = -40$, $b = 40$. The upper bound on P_{fa} is set to $\alpha^* = 0.1$, where α^* is obtained using the methodology in Wilson and Veeravalli (2016).

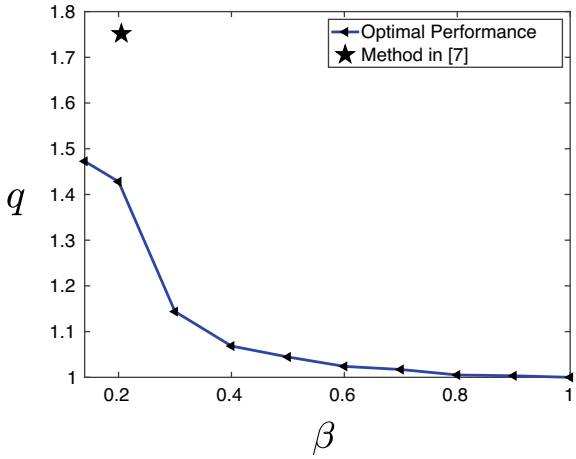


Fig. 9.4 q versus β for fixed $\alpha^* = 0.1$

9.5.2 Case 2: Both Sensors Vulnerable to Adversarial Attacks

We consider the same model for X, and in this setting, we assume that both sensors are vulnerable to attack. The attacker can compromise the data of at most one sensor. Under this setting, we have $T = 2$, $S_1 = \{1\}$, and $S_2 = \{2\}$. Therefore, in the adversarial setting, the following hypothesis model forms the basis of the secure estimation problem

$$\begin{aligned} \mathsf{H}_0 &: \mathbf{Y} \sim f_0(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \mathsf{H}_1 &: \mathbf{Y} \sim f_1(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X) \\ \mathsf{H}_2 &: \mathbf{Y} \sim f_2(\mathbf{Y} \mid X), \quad \text{with } X \sim \pi(X), \end{aligned} \tag{9.50}$$

where H_0 is the attack-free setting and H_i corresponds to sensor i being compromised. Since the sensor with higher gain h^i is expected to provide 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 $(\varepsilon_0, \varepsilon_1, \varepsilon_2) = (0.2, 0.2, 0.6)$. We assume the distribution of X to be $\mathsf{Unif}[-2, 2]$. 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, for $i \in \{1, 2\}$, given X, follows the distribution $\mathcal{N}(h^i X, 5)$.

Figure 9.5 shows the performance region illustrated in Fig. 9.2, which corresponds to the variations of q with β for three different values of α. The region spanned by the plots between q and β for different values of α is the feasible region of operation and allows the FC to adjust the emphasis on either the estimation or detection decisions. As expected, the estimation quality improves monotonically as α and β increase.

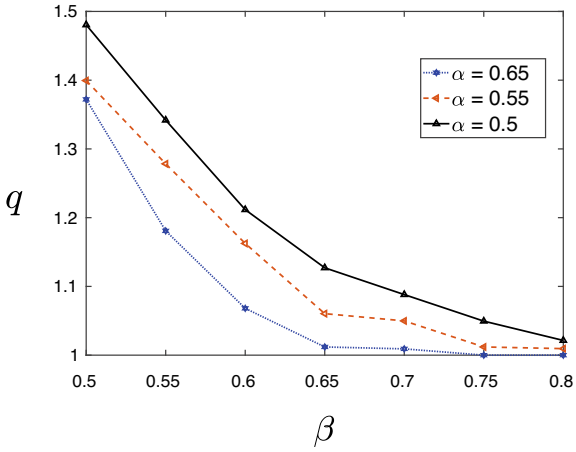


Fig. 9.5 q versus β for different values of α

9.6 Conclusions

We have formalized and analyzed the problem of secure estimation under adversarial attacks on the data model. The possible presence of adversaries results in uncertainty in the statistical model of the data. This further leads the estimation algorithm to exhibit degraded performance compared to the attack-free setting. We have characterized closed-form optimal decision rules that provide the optimal estimation quality (minimum estimation cost) while controlling for the error in detecting the attack and isolating the true model of the data. Our analysis has shown that the design of optimal estimators is intertwined with that of the detection rules to determine the true model of the data. Based on this, we have provided the optimal decision rules that combine the estimation quality with detection power. This allows the decision-maker to place any desired emphasis on the estimation and detection routines involved to study the tradeoff between the two.

Appendix 1

We start the proof of Theorem 9.2 by defining the cost function $J_i(\delta_i, U_i)$ and analyzing a lower bound on it. Our analysis will show that the lower bound on the $J_i(\delta_i, U_i)$ is achieved for the choice of estimator in (9.53). From (9.10), we have

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

By leveraging the definition of $\mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y})$ from (9.24), we have

$$\begin{aligned} J_i(\delta_i, U_i) &= \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})\mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}} \\ &\geq \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \inf\limits_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}\,, \end{aligned} \tag{9.51}$$

which implies that

$$J_i(\delta_i, U_i) \geq \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}\,. \tag{9.52}$$

Using the definition of $\hat{X}_i(\mathbf{Y})$ in (9.23), the above lower bound is achieved when the estimator $U_i(\mathbf{Y})$ is selected to be

$$\hat{X}_i(\mathbf{Y}) = \arg \inf_{U_i(\mathbf{Y})} \mathsf{C}_{\mathrm{p},i}(U_i(\mathbf{Y}) \mid \mathbf{Y}), \tag{9.53}$$

which proves that the estimator in (9.23) is the optimal estimator for minimizing the cost $J_i(\delta_i, U_i)$. The corresponding minimum average estimation cost is

$$J_i(\delta_i, \hat{X}_i) = \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y})\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y})\mathrm{d}\mathbf{Y}}\,. \tag{9.54}$$

Next, we prove that

$$\max_i \min_{\mathbf{U}} \{J_i(\delta_i, U_i)\} \equiv \min_{\mathbf{U}} \max_i \{J_i(\delta_i, U_i)\}\,. \tag{9.55}$$

Recall from (9.11), the estimation cost $J(\boldsymbol{\delta}, \mathbf{U})$ is defined as

$$J(\boldsymbol{\delta}, \mathbf{U}) = \max_{i} \{J_i(\delta_i, U_i)\}. \tag{9.56}$$

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

$$\mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) \triangleq \sum_{i=0}^{T} \Omega_i J_i(\delta_i, U_i), \tag{9.57}$$

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

$$\sum_{i=0}^{T} \Omega_i = 1 \ , \text{ and } \Omega_i \in [0, 1]. \tag{9.58}$$

$J(\boldsymbol{\delta}, \mathbf{U})$ can be represented as a function of $\mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U})$ in the following form

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

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

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

where $\Omega_j^* = 1$ if

$$j = \arg\max_{i} \{J_i(\delta_i, U_i)\}. \tag{9.59}$$

From (9.53) and (9.54), we observe that

$$\begin{aligned} \max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) &= \max_{\boldsymbol{\Omega}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \hat{\mathbf{X}}) \\ &\geq \min_{\mathbf{U}} \max_{\boldsymbol{\Omega}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}). \end{aligned} \tag{9.60}$$

Also, we have

$$\max_{\boldsymbol{\Omega}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) \geq \max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}), \tag{9.61}$$

which implies that

$$\min_{\mathbf{U}} \max_{\boldsymbol{\Omega}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) \geq \max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}). \tag{9.62}$$

From (9.60) and (9.62), it is easily concluded that

$$\max_{\boldsymbol{\Omega}} \min_{\mathbf{U}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}) = \min_{\mathbf{U}} \max_{\boldsymbol{\Omega}} \mathscr{C}(\boldsymbol{\Omega}, \boldsymbol{\delta}, \mathbf{U}), \tag{9.63}$$

which completes the proof for (9.55). Using the results in (9.55) and (9.54), the cost function $J(\boldsymbol{\delta}, \hat{\mathbf{X}})$ is given by

$$\begin{aligned} J(\boldsymbol{\delta}, \hat{\mathbf{X}}) &= \min_{\mathbf{U}} \max_{i} \{J_i(\delta_i, U_i)\} \\ &= \max_{i} \min_{\mathbf{U}} \{J_i(\delta_i, U_i)\} \\ &= \max_{i} \left\{ J_i(\delta_i, \hat{X}_i) \right\} \qquad (9.64) \\ &= \max_{i} \left\{ \frac{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) \mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}}{\int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) \mathrm{d}\mathbf{Y}} \right\} . \qquad (9.65) \end{aligned}$$

Appendix 2

The function $J_i(\delta_i, U_i)$ is a quasi-convex function. The weighted maximum function preserves the quasi-convexity and therefore, $J_i(\delta_i, \hat{X}_i)$ is a quasi-convex function from its definition in (9.26). This allows us to find the solution by solving an equivalent feasibility problem given below (Boyd and Vandenberghe 2004). Specifically, for $u \in \mathbb{R}_+$, it is observed that

$$J(\boldsymbol{\delta}, \hat{\mathbf{X}}) \leq u \equiv \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) (\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u) \mathrm{d}\mathbf{Y} \leq 0, \text{ for } i \in \{0, \ldots, T\} . \quad (9.66)$$

Hence, the feasibility problem equivalent to (9.22) is given by

$$\mathscr{P}(\alpha, \beta) = \begin{cases} \min_{\boldsymbol{\delta}} & u \\ \text{s.t.} & \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_i(\mathbf{Y}) (\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u) \mathrm{d}\mathbf{Y} \leq 0, \quad \forall i \in \{0, \ldots, T\} \\ & \sum_{j=1}^{T} \sum_{i=0, i \neq j}^{T} \frac{\varepsilon_j}{1-\varepsilon_0} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_j(\mathbf{Y}) \mathrm{d}\mathbf{Y} \leq \beta \\ & \sum_{i=1}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) f_0(\mathbf{Y}) \mathrm{d}\mathbf{Y} \leq \alpha \end{cases} . \quad (9.67)$$

The above problem is feasible if $\mathscr{P}(\alpha, \beta) \leq u$, where $\mathscr{P}(\alpha, \beta)$ is the lowest value of u for which the problem is feasible and all constraints are satisfied. Given an interval $[u_0, u_1]$ containing $\mathscr{P}(\alpha, \beta)$, the detection rule $\boldsymbol{\delta}$ and the estimation cost $\mathscr{P}(\alpha, \beta)$ are determined by a bi-section search between u_0 and u_1 iteratively, solving the feasibility problem in each iteration. We define an auxiliary convex optimization problem that allows us to solve the feasibility problem

$$\mathbb{R}(\alpha,\beta,u)=\begin{cases}\min_{\boldsymbol{\delta}} & \eta\\ \text{s.t.} & \displaystyle\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_i(\mathbf{Y})(\mathsf{C}^{*}_{\mathrm{p},i}(\mathbf{Y})-u)\mathrm{d}\mathbf{Y}\le\eta,\quad \forall i\in\{0,\ldots,T\}\\ & \displaystyle\sum_{j=1}^{T}\sum_{i=0,i\neq j}^{T}\frac{\varepsilon_j}{1-\varepsilon_0}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_j(\mathbf{Y})\mathrm{d}\mathbf{Y}\ \le\beta+\eta\\ & \displaystyle\sum_{i=1}^{T}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_0(\mathbf{Y})\mathrm{d}\mathbf{Y}\le\alpha+\eta\end{cases}. \tag{9.68}$$

Algorithm 9.2 summarizes the steps for determining $\mathscr{P}(\alpha,\beta)$.

Algorithm 9.2 Bi-section Search

Input: Initialize u_0, u_1
repeat
  $\hat{u}\leftarrow(u_0+u_1)/2$;
  Solve $\mathbb{R}(\alpha,\beta,\hat{u})$;
  if $\mathscr{J}(\alpha,\beta,\hat{u})\le 0$ **then**
    $u_1\leftarrow\hat{u}$;
  else
    $u_0\leftarrow\hat{u}$;
until $u_1-u_0\le\varepsilon$, *for* ε *sufficiently small*;
Output: $\mathscr{P}(\alpha,\beta)\leftarrow u_1$

Appendix 3

To solve the problem in (9.68), a Lagrangian function is constructed according to

$$\begin{aligned}\mathscr{Q}(\boldsymbol{\delta},\eta,\boldsymbol{\ell})\triangleq&\left(1-\sum_{i=0}^{T+2}\ell_i\right)\eta\\&+\sum_{i=0}^{T}\ell_i\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_i(\mathbf{Y})(\mathsf{C}^{*}_{\mathrm{p},i}(\mathbf{Y})-u)\mathrm{d}\mathbf{Y}\\&+\ell_{T+1}\sum_{j=1}^{T}\sum_{i=0,i\neq j}^{T}\frac{\varepsilon_j}{1-\varepsilon_0}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_j(\mathbf{Y})\mathrm{d}\mathbf{Y}\ -\ell_{T+1}\beta\\&+\ell_{T+2}\sum_{i=1}^{T}\int_{\mathbf{Y}}\delta_i(\mathbf{Y})f_0(\mathbf{Y})\mathrm{d}\mathbf{Y}-\ell_{T+2}\alpha,\end{aligned}$$

where $\boldsymbol{\ell} \triangleq \left[\ell_0, \ldots, \ell_{T+2}\right]$ are the non-negative Lagrangian multipliers selected to satisfy the constraints in (9.22), such that

$$\sum_{i=0}^{T+2} \ell_i = 1. \tag{9.69}$$

The Lagrangian dual function is given by

$$\begin{aligned} d(\boldsymbol{\ell}) &\triangleq \min_{\boldsymbol{\delta}, \eta} \mathscr{Q}(\boldsymbol{\delta}, \eta, \boldsymbol{\ell}) \\ &= \min_{\boldsymbol{\delta}} \left(\sum_{i=0}^{T} \int_{\mathbf{Y}} \delta_i(\mathbf{Y}) A_i \mathrm{d}\mathbf{Y} \right) - \ell_{T+1}\beta - \ell_{T+2}\alpha, \end{aligned} \tag{9.70}$$

where

$$A_0 \triangleq \ell_0 f_0(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},0}(\mathbf{Y}) - u] + \ell_{T+1} \sum_{i=1}^{T} \frac{\varepsilon_i}{1-\varepsilon_0} f_i(\mathbf{Y}), \tag{9.71}$$

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

$$A_i \triangleq \ell_i f_i(\mathbf{Y})[\mathsf{C}^*_{\mathrm{p},i}(\mathbf{Y}) - u] + \ell_{T+1} \sum_{j=1, j\neq i}^{T} \frac{\varepsilon_j}{1-\varepsilon_0} f_j(\mathbf{Y}) + \ell_{T+2} f_0(\mathbf{Y}). \tag{9.72}$$

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

$$\delta_i(\mathbf{Y}) = \begin{cases} 1, & \text{if} \quad i = i^* \\ 0, & \text{if} \quad i \neq i^* \end{cases}, \tag{9.73}$$

where $i^* = \mathrm{argmin}_{i\in\{0,\ldots,T\}} A_i$. Hence, the proof is concluded.

References

C.M. Ahmed, M.A. Umer, B.S.S.B. Liyakkathali, M.T. Jilani, J. Zhou, Machine learning for cps security: applications, challenges and recommendations, in *Machine Intelligence and Big Data Analytics for Cybersecurity Applications* (Springer, 2021), pp. 397–421

S. Al-Sayed, A.M. Zoubir, A.H. Sayed, Robust distributed estimation by networked agents. IEEE Trans. Signal Process. **65**(15), 3909–3921 (2017)

C.Z. Bai, V. Gupta, On Kalman filtering in the presence of a compromised sensor: fundamental performance bounds, in *Proceedings of American Control Conference*, Portland, OR (2014), pp. 3029–3034

S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004)

K. Chen, V. Gupta, Y. Huang, Minimum variance unbiased estimation in the presence of an adversary, in *Proceedings of Conference on Decision and Control (CDC)* (2017), pp. 151–156

P. Ebinger, S.D. Wolthusen, Efficient state estimation and Byzantine behavior identification in tactical MANETs, in *Proceedings of IEEE Military Communications Conference*, Boston, MA (2009)

H. Fawzi, P. Tabuada, S. Diggavi, Secure state-estimation for dynamical systems under active adversaries, in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Monticello, IL (2011), pp. 337–344

H. Fawzi, P. Tabuada, S. Diggavi, Secure estimation and control for cyber-physical systems under adversarial attacks. IEEE Transactions on Automatic Control **59**(6), 1454–1467 (2014)

H. Fawzi, P. Tabuada, S. Diggavi, Secure estimation and control for cyber-physical systems under adversarial attacks. IEEE Trans. Autom. Control **59**(6), 1454–1467 (2014)

F. Hu, Y. Lu, A.V. Vasilakos, Q. Hao, R. Ma, Y. Patil, T. Zhang, J. Lu, X. Li, N.N. Xiong, Robust cyber-physical systems: Concept, models, and implementation. Future generation computer systems **56**, 449–475 (2016)

A. Humayed, J. Lin, F. Li, B. Luo, Cyber-physical systems security'a survey. IEEE Internet of Things Journal **4**(6), 1802–1831 (2017)

G.H. Jajamovich, A. Tajer, X. Wang, Minimax-optimal hypothesis testing with estimation-dependent costs. IEEE Transactions on Signal Processing **60**(12), 6151–6165 (2012)

J. Li, Y. Liu, T. Chen, Z. Xiao, Z. Li, J. Wang, Adversarial attacks and defenses on cyber-physical systems: a survey. IEEE Internet Things J. **7**(6), 5103–5115 (2020)

Y. Lin, A. Abur, Robust state estimation against measurement and network parameter errors, IEEE Trans. Power Syst. 33, (5), pp. 4751–4759 2020

D. Middleton, R. Esposito, Simultaneous optimum detection and estimation of signals in noise. IEEE Trans. Inf. Theory **14**(3), 434–444 (1968)

S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, P. Tabuada, Secure state estimation: optimal guarantees against sensor attacks in the presence of noise, in *Proceedings of IEEE International Symposium on Information Theory*, Hong Kong, China (2015), pp. 2929–2933

G.V. Moustakides, G.H. Jajamovich, A. Tajer, X. Wang, Joint detection and estimation: Optimum tests and applications. IEEE Transactions on Information Theory **58**(7), 4215–4229 (2012)

M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, G.J. Pappas, Robustness of attack-resilient state estimators, in *Proceedings of IEEE International Conference on Cyber-Physical Systems* (2014), pp. 163–174

M. Pajic, P. Tabuada, I. Lee, G.J. Pappas, Attack-resilient state estimation in the presence of noise, in *Proceedings of IEEE Conference on Decision and Control*, Osaka, Japan (2015), pp. 5827–5832

H.V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd edn. (Springer-Verlag, New York, 1998)

A.S. Rawat, P. Anand, H. Chen, P.K. Varshney, Countering Byzantine attacks in cognitive radio networks, in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX (2010), pp. 3098–3101

A.H. Sayed, A framework for state-space estimation with uncertain models. IEEE Trans. Autom. Control **46**(7), 998–1013 (2001)

X. Shen, P.K. Varshney, Y. Zhu, Robust distributed maximum likelihood estimation with dependent quantized data. Automatica **50**(1), 169–174 (Jan. 2014)

Y. Shoukry, P. Nuzzo, A. Puggelli, A.L. Sangiovanni-Vincentelli, S.A. Seshia, P. Tabuada, Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach. IEEE Transactions on Automatic Control **62**(10), 4917–4932 (2017)

S. Sihag, A. Tajer, Secure estimation under causative attacks. IEEE Transactions on Information Theory **66**(8), 5145–5166 (2020)

E. Soltanmohammadi, M. Orooji, M. Naraghi-Pour, Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes. IEEE Transactions on Information Forensics and Security **8**(1), 205–215 (2013)

A. Vempaty, K. Agrawal, P. Varshney, H. Chen, Adaptive learning of Byzantines' behavior in cooperative spectrum sensing, in *Proceedings of IEEE Wireless Communications and Networking Conference*, Cancun, Mexico (2011), pp. 1310–1315

A. Vempaty, L. Tong, P.K. Varshney, Distributed inference with Byzantine data: state-of-the-art review on data falsification attacks. IEEE Signal Process. Mag. **30**(5), 65–75 (2013)

A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, P.K. Varshney, Localization in wireless sensor networks: Byzantines and mitigation techniques. IEEE Transactions on Signal Processing **61**(6), 1495–1508 (2013)

A. Vempaty, Y.S. Han, P.K. Varshney, Target localization in wireless sensor networks using error correcting codes. IEEE Transactions on Information Theory **60**(1), 697–712 (2014)

C. Wilson, V.V. Veeravalli, MMSE estimation in a sensor network in the presence of an adversary, in *Proceedings of IEEE International Symposium on Information Theory*, Barcelona, Spain (2016), pp. 2479–2483

S.Z. Yong, M. Zhu, E. Frazzoli, Resilient state estimation against switching attacks on stochastic cyber-physical systems, in *Proceedings of IEEE Conference on Decision and Control*, Osaka, Japan (2015), pp. 5162–5169

O. Zeitouni, J. Ziv, N. Merhav, When is the generalized likelihood ratio test optimal? IEEE Transactions on Information Theory **38**(5), 1597–1602 (1992)

J. Zhang, R.S. Blum, Distributed estimation in the presence of attacks for large scale sensor networks, in *Proceedings of Conference on Information Sciences and Systems*, Princeton, NJ (2014)

J. Zhang, R.S. Blum, X. Lu, D. Conus, Asymptotically optimum distributed estimation in the presence of attacks. IEEE Transactions on Signal Processing **63**(5), 1086–1101 (2015)

J. Zhao, M. Netto, L. Mili, A robust iterated extended kalman filter for power system dynamic state estimation. IEEE Transactions on Power Systems **32**(4), 3205–3216 (2016)