This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

Secure Detection: Performance Metric and Sensor Deployment Strategy

Xiaoqiang Ren and Yilin Mo*

***Abstract*—This paper studies how to deploy sensors in the context of detection in adversarial environments. A fusion center is performing a binary hypothesis testing based on measurements from remotely deployed heterogeneous sensors. An attacker may compromise some of the deployed sensors, which send arbitrary measurements to the fusion center. The problems of interest are: 1) to characterize the performance of the system under attack and thus develop a performance metric; 2) to deploy sensors within a cost budget, such that the proposed performance metric is maximized. In this paper, we first present a performance metric by formulating the detection in adversarial environments in a game theoretic way. A Nash equilibrium pair of the detection algorithm and attack strategy, with the deployed sensors given, is provided and the corresponding detection performance is adopted as the performance metric. We then show that the optimal sensor deployment can be determined approximately by solving a group of unbounded knapsack problems. We also show that the performance metric gap between the optimal sensor deployment and the optimal one with sensors being identical is within a fixed constant for any cost budget. The main results are illustrated by numerical examples.**

***Index Terms*—Secure detection, Byzantine attacks, sensor deployment, game**

I. Introduction

Background: Network embedded sensors are widely used to monitor systems. However, due to the limited capacity and sparsely spatial deployment, they are vulnerable to malicious attacks. An attacker may compromise the sensors and send arbitrary messages, break the communication channels and tamper with the transmitted data, or just launch jamming noises to block the communication channels. Such attacks have motivated many researches on how to process information securely under attack in the context of networked estimation, detection and control [1]–[5].

Our Work and its Contributions: We consider binary hypothesis testing with sensors under Byzantine attack. A fusion center is performing a binary hypothesis testing based on measurements from a group of deployed heterogeneous sensors, among which some may be compromised by an attacker. The measurements of a compromised sensor can be arbitrarily manipulated. The fusion center knows the number of the compromised sensors[1], but does not know what sensors are compromised.

The following two problems are to be addressed in this paper. First, given the deployed sensors, what is a reasonable metric to characterize the performance of the system under attack? Second, suppose that the system manager can choose what sensors to deploy from a sensor pool, in which different types of sensors incur different costs and provide information of different quality. Then with a sensor cost budget, what is the optimal sensor deployment strategy to maximize the proposed performance metric?

The main contributions of this work are summarized as follows:

1) We provide a performance metric by formulating the detection in adversarial environments in a game theoretic way. A Nash equilibrium pair of detection algorithm and attack strategy is identified (Theorem 3) and the corresponding detection performance is adopted as the metric.
2) To solve the sensor deployment problem, two heuristics are provided: 1) A suboptimal sensor deployment strategy, which maximizes a upper bound of the performance metric, is given by solving a group of unbounded knapsack problems (UKPs). We also prove that this suboptimal strategy is indeed the optimal one in certain cases, for example, detecting mean shift of Gaussian noises. 2) We consider using identical sensors and show that the performance metric gap between the optimal sensor deployment and the optimal identical sensors deployment is within a fixed constant for any cost budget (Theorem 5). Notice that this result does not rely on the approximation techniques used in the above UKP formulation.

Related Literature: Minimax robust detection has been studied over decades to deal with the uncertainties of input data [6]–[8]. The classical approach is to identify the "least favorables" among the allowable set of uncertainties first, and then perform the classic probability ratio test. These allowable sets usually consist of a nominal element and others that are somewhat "close" to this nominal one [7]. This "nice" structure may not hold in security settings. Also, it is usually difficult to identify the "least favorables" for arbitrary uncertainties as there does not exist a systematic way to do so.

Xiaoqiang Ren was with the School of Electrical and Electronics Engineering, Nanyang Technological University when the manuscript was written. He is now with the ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute of Technology, 114 28 Stockholm, Sweden. Email: xiaoqren@kth.se

Yilin Mo is with the School of Electrical and Electronics Engineering, Nanyang Technological University. Email: ylmo@ntu.edu.sg

*: Corresponding Author.

This work was supported in part by the Academic Research Fund Tier 1 project under Grant No. M4011504.040 from the Ministry of Education, Singapore. The work of Xiaoqiang Ren was supported in part by the Knut and Alice Wallenberg Foundation and the Swedish Research Council.

[1]This number can also be interpreted as how many bad sensors the system can or is willing to tolerate, which is a design parameter.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

Detection under Byzantine attack has also been studied in [9], [10]. In [9], the authors took the respective of an intruder and found the optimal attacks to minimize the Kullback–Leibler divergence of the manipulated measurements. The manipulated measurements are assumed to be independent and identical, which limits the capability of the adversary. On the contrary, this paper assumes that the measurements of the compromised sensors can be arbitrary, which might be correlated. Also, we adopt a different performance index (i.e., Chernoff information), the analysis method is thus fundamentally different. In [10], the Byzantine sensors can collaborate when generating malicious data. The optimal detector when the number of Byzantine sensors is less than $1/2$ is shown to be of a threshold structure. The authors only focus on one-step detector, while we consider an infinite time sequence of detectors, which is more challenging. Besides, the problem in this paper is formulated in a game-theoretic way, while the above two works take the perspective of either an attacker or a system manager.

Detection with Byzantine sensors have also been studied in a game-theoretic way [11], where binary and noisy sensors are used. A zero-sum game was formulated and (approximate) equilibrium was obtained. However, as in [10], they focus on one-step scenario, which is contrasted by the asymptotic performance index in this paper. In the asymptotic regime, each strategy consists of an infinite time sequence of behavior rules. This renders the analysis challenging and requires fundamentally different analysis techniques. Notice also that the binary sensor model in [11] restricts its application, and the explicit equilibrium was only obtained under certain conditions. At last, we should note that none of the aforementioned literature has studied the sensor deployment strategy.

This paper extends our previous work [12] from homogeneous sensors to heterogeneous sensors. However, due to the heterogeneity of sensors in this paper, the equilibrium detection algorithm (i.e., f^* in Section IV) is fundamentally different. Our recent work [13] studied the fundamental tradeoff between security and efficiency for given homogeneous sensors, while this paper focus on interactive behaviors of an attacker and a system. The detection algorithm f^* is inspired by [13]. However, since the sensors in this paper are heterogeneous, the analysis techniques are significantly different.

Organization: In Section II, we provide the detection model, attack model, detection performance index and the problem of interest. The large deviation theory is presented in Section III, which is a key supporting analysis tool. In Section IV, we formulate the detection in adversarial environments with given heterogeneous sensors in a game theoretic way, and identify a Nash equilibrium of the detection algorithm and attack strategy. The performance metric is proposed as the detection performance when the equilibrium strategy is played. In Section V, two heuristics of the sensor deployment problem are provided: 1) the problem is relaxed to UKPs, 2) identical sensors are considered. Numerical examples are given in Section VI and concluding remarks are provided in Section VII.

Notations: $\mathbb{R}$ ($\mathbb{R}_+$) is the set of (nonnegative) real numbers. $\mathbb{N}$ ($\mathbb{N}_+$) is the set of nonnegative (positive) integers. For a set $\mathcal{A} \subset \mathbb{R}^n$, $\text{int}(\mathcal{A})$ denotes its interior. The cardinality of finite set $\mathcal{A}$ is denoted as $|\mathcal{A}|$. Given a function $f : \mathbb{R} \to \mathbb{R}$, we use $f^{(1)}(x)$ to denote its first-order derivative at point x. For two vectors $x, y \in \mathbb{R}^n$, $x \cdot y$ denotes their dot product. For a vector $x \in \mathbb{R}^n$, define $\|x\|_0$ as the "zero norm", i.e., the number of nonzero elements of the vector x. $\lfloor x \rfloor$ is the floor function of $x \in \mathbb{R}$. For a vector $x \in \mathbb{R}^n$, the support of x, denoted by $\text{supp}(x)$, is the set of indices of nonzero elements:

$$\text{supp}(x) \triangleq \{i \in \{1, 2, \ldots, n\} : x_i \neq 0\}.$$

The main notations used in this paper are summarized in the following table:

$y(k)$	all sensors' original measurements at time k
$\tilde{y}(k)$	all sensors' manipulated measurements at time k
$\mathbf{y}(k)$	all sensors' accumulated original measurements from time 1 to k
$\mathbf{y}(k)$	all sensors' accumulated manipulated measurements from time 1 to k
μ_i (ν_i)	probability measures for sensor i when hypothesis $\theta = 1$ ($\theta = 0$)
$\lambda_i(\cdot)$	log-likelihood ratio function between μ_i and ν_i
$I_{0,i}(\cdot)$	rate function of λ_i when $\theta = 0$
$I_{1,i}(\cdot)$	rate function of λ_i when $\theta = 1$
C_i	Chernoff information between μ_i and ν_i
$C_{\mathcal{O}}$	Chernoff information between $\prod_{i\in\mathcal{O}} \mu_i$ and $\prod_{i\in\mathcal{O}} \nu_i$
$\mathcal{H}$	set of sensors that will be healthy under the worst-case attacks
$C(d)$	detection performance of sensor deployment strategy d
c_p	Chernoff information of a sensor of type p
u_p	cost of a sensor of type p
U	cost budget
n	number of compromised sensors

II. Problem Formulation

A. Detection Model

Consider the problem of detecting a binary state $\theta \in \{0, 1\}$ using m sensors' measurements. Define the measurement $y(k)$ at time k to be a row vector:

$$y(k) \triangleq \begin{bmatrix} y_1(k) & y_2(k) & \cdots & y_m(k) \end{bmatrix} \in \mathbb{R}^m, \tag{1}$$

where $y_i(k)$ is the scalar measurement from sensor i at time k. For simplicity, we define $\mathbf{y}(k)$ as a vector of all measurements from time 1 to time k:

$$\mathbf{y}(k) \triangleq \begin{bmatrix} y(1) & y(2) & \cdots & y(k) \end{bmatrix} \in \mathbb{R}^{mk}. \tag{2}$$

Given θ, we assume that all measurements from sensor i, $\{y_i(k)\}_{k=1,2,\ldots}$ are independent and identically distributed (i.i.d.), and measurements from different sensors are independent but not necessarily with the same distributions. The probability measure generated by $y_i(k)$ is denoted as ν_i when $\theta = 0$ and it is denoted as μ_i when $\theta_i = 1$. In other words, for any Borel-measurable set $\mathcal{A} \subseteq \mathbb{R}$, the probability that $y_i(k) \in \mathcal{A}$ equals $\nu_i(\mathcal{A})$ when $\theta = 0$ and equals $\mu_i(\mathcal{A})$ when $\theta = 1$. We denote the probability space generated by all measurements $y(1)$, $y(2)$, $\ldots$ as $(\Omega_y, \mathcal{F}_y, \mathbb{P}_\theta)$, where for any $l \geq 1$

$$\mathbb{P}_\theta(y_{i_1}(k_1) \in \mathcal{A}_1, \ldots, y_{i_l}(k_l) \in \mathcal{A}_l) = \begin{cases} \nu_{i_1}(\mathcal{A}_1)\nu_{i_2}(\mathcal{A}_2)\ldots\nu_{i_l}(\mathcal{A}_l) & \text{if } \theta = 0, \\ \mu_{i_1}(\mathcal{A}_1)\mu_{i_2}(\mathcal{A}_2)\ldots\mu_{i_l}(\mathcal{A}_l) & \text{if } \theta = 1 \end{cases}$$

when $(i_j, k_j) \neq (i_{j'}, k_{j'})$ for all $j \neq j'$. We further assume that ν_i and μ_i are absolutely continuous with respect to each

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

other. Hence, the log-likelihood ratio $\lambda_i : \mathbb{R} \to \mathbb{R}$ of $y_i(k)$ is well defined as

$$\lambda_i(y_i) \triangleq \log\left(\frac{\mathrm{d}\mu_i}{\mathrm{d}\nu_i}(y_i)\right), \tag{3}$$

where $\mathrm{d}\mu_\mathrm{i}/\mathrm{d}\nu_\mathrm{i}$ is the Radon-Nikodym derivative.

We define $f_k : \mathbb{R}^{mk} \to [0,\, 1]$, the detector at time k, as a mapping from the measurement space $\mathbf{y}(k)$ to the interval $[0,\, 1]$. When $f_k(\mathbf{y}(k)) = 0$, the system makes a decision $\hat{\theta} = 0$, and when $f_k(\mathbf{y}(k)) = 1$, $\hat{\theta} = 1$. When $f_k(\mathbf{y}(k)) = \gamma \in (0,\, 1)$, the system then "flips a biased coin" to choose $\hat{\theta} = 1$ with probability γ and $\hat{\theta} = 0$ with probability $1 - \gamma$. The system's strategy $f \triangleq (f_1,\, f_2,\, \cdots)$ is defined as an infinite sequence of detectors from time 1 to ∞.

B. Attack Model

An adversary wants to deteriorate the performance of the system, the model of which is described as follows.

Assumption 1 (Attacker's knowledge)**.**

1) *The attacker knows the probability measures μ_i, ν_i, and the true state θ.*
2) *At each time, the attacker knows the current and all the historical measurements available at the compromised sensors.*

This assumption is in accordance with Kerckhoffs's principle [14], i.e., the security of the system should not rely on its obscurity. By the knowledge about the sensor, the attacker can develop the probability measures μ_i and ν_i. To obtain the true state, the attacker may deploy its own sensor network. Though it might be difficult to satisfy in practice, this assumption is in fact conventional in literature concerning the worst-case attacks, e.g., [9], [15].

One may verify that the main result (i.e., Theorem 3) remains even if the attacker is "strong" enough to access the measurements from all the sensors (rather than the compromised ones only as in Assumption 1).

In the following, how an attacker, with the above knowledge assumed, affects the detection system is introduced. Let the *manipulated measurements* received by the fusion center at time k be

$$\tilde{y}(k) = y(k) + y^a(k). \tag{4}$$

where $y^a(k) \in \mathbb{R}^m$ is the bias vector injected by the attacker at time k.

Assumption 2 (n-sparse attack)**.**

1) *There exists an index set $\mathcal{I} \subset \mathcal{M} \triangleq \{1, 2, \ldots, m\}$ with $|\mathcal{I}| = n$ such that $\bigcup_{k=1}^{\infty} \mathrm{supp}(y^a(k)) = \mathcal{I}$.*
2) *The system knows the number n, but it does not know the set $\mathcal{I}$.*

Assumption 2 says that the set of compromised sensors is somewhat "constant" over time. This is reasonable in the sense that a time varying set of compromised sensors would require the attacker to abandon the sensors it already controlled. We should note that although the set $\mathcal{I}$ is fixed over time, attacker has the freedom to choose $\mathcal{I}$ at the beginning.

It is also assumed that the number of compromised sensors is upper bounded by n. It is practical to assume that the attacker possesses limited resources since otherwise it would be too pessimistic and the problem becomes trivial. The quantity n might be determined by the *a priori* knowledge about the quality of each sensor. On the other hand, the quantity n may be viewed as a design parameter, which indicates the resilience level that the system is willing to introduce. In general, increasing n will increase the resilience of the detector under attack, which, however, might lead to a more conservative design.

Finally, we should note that we do not assume any pattern of the bias $y_i^a(k)$ for $i \in \mathcal{I}$, i.e., the injected malicious bias may take any value, and may be correlated across the compromised sensors and over time. Compared to the independence assumption in [9], the capability of an attacker is increased in this work. Furthermore, this is more realistic in the sense that the attacker is malicious and intelligent, and will fully utilize the sensors it controlled.

In fact, the same sparse attack model as in Assumption 2 has been widely adopted by literature dealing with Byzantine sensors, e.g., binary hypothesis testing [9], [10], state estimation [2], [16]–[18], and quickest change detection [15].

An admissible attack strategy is any causal mapping from the attacker's available information to a bias vector that satisfies Assumption 2. This is formalized as follows. Let $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$. Define the true measurements of the compromised sensors from time 1 to k as

$$\mathbf{y}_\mathcal{I}(k) \triangleq \begin{bmatrix} y_\mathcal{I}(1) & y_\mathcal{I}(2) & \cdots & y_\mathcal{I}(k) \end{bmatrix} \in \mathbb{R}^{|\mathcal{I}|k}$$

with

$$y_\mathcal{I}(k) \triangleq \begin{bmatrix} y_{i_1}(k) & y_{i_2}(k) & \cdots & y_{i_n}(k) \end{bmatrix} \in \mathbb{R}^{|\mathcal{I}|}.$$

Let $\tilde{\mathbf{y}}(k)$ ($\tilde{\mathbf{y}}^a(k)$, respectively) be defined as all the manipulated measurements (bias vectors, respectively) from time 1 to k. The bias at time k, $y^a(k)$, is chosen as a random function of the attacker's available information at time k:

$$y^a(k) \triangleq g(\mathbf{y}_\mathcal{I}(k), \mathbf{y}^a(k-1), \mathcal{I}, \theta, k), \tag{5}$$

where g is a random function of $\mathbf{y}_\mathcal{I}(k), \mathbf{y}^a(k-1), \mathcal{I}, \theta, k$ such that $\mathbf{y}^a(k)$ satisfies Assumption 2. By random function, we mean that given $\mathbf{y}_\mathcal{I}(k), \mathbf{y}^a(k-1), \mathcal{I}, \theta, k$, the bias $y^a(k)$ might be a random vector with a distribution determined by g. We denote g as an admissible attacker's strategy. Notice that due to the time k and "increasing" $\mathbf{y}_\mathcal{I}(k), \mathbf{y}^a(k-1)$, the definition in (5) does not exclude time-varying attack strategies.

Denote the probability space generated by all manipulated measurements $\tilde{y}(1)$, $\tilde{y}(2)$, ... as $(\Omega,\, \mathcal{F},\, \tilde{\mathbb{P}}_\theta)$. The expectation taken with respect to the probability measure $\tilde{\mathbb{P}}_\theta$ is denoted by $\tilde{\mathbb{E}}_\theta$.

C. Asymptotic Detection Performance

Given the strategy of the system and the attacker, the probability of error at time k can be defined as

$$e(\theta, \mathcal{I}, k) \triangleq \begin{cases} \tilde{\mathbb{E}}_0 f_k(\tilde{\mathbf{y}}(k)) & \text{when } \theta = 0, \\ 1 - \tilde{\mathbb{E}}_1 f_k(\tilde{\mathbf{y}}(k)) & \text{when } \theta = 1. \end{cases} \tag{6}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

Notice that f_k could take any value from $[0,1]$. Hence, the expected value of f_k is used to compute the probability of error. In this paper, we are concerned with the worst-case scenario. As a result, let us define

$$\epsilon(k) \triangleq \max_{\theta=0,1,|\mathcal{I}|=n} e(\theta,\mathcal{I},k). \tag{7}$$

In other words, $\epsilon(k)$ indicates the worst-case probability of error considering all possible sets of compromised sensors and the state θ.

It is quite difficult to analyze $\epsilon(k)$ when k takes finite values since computing the probability of error usually involves numerical integration. Thus, in this work, we consider the asymptotic detection performance, i.e., the exponential rate that the worst-case probability of error goes to zero:

$$\rho \triangleq \liminf_{k\to\infty} -\frac{\log \epsilon(k)}{k}. \tag{8}$$

Clearly, ρ is a function of both the system strategy f and the attacker's strategy g. As such, we will write ρ as $\rho(f,g)$ to indicate such relations. Obviously, the system would like to maximize ρ to make the detection error smaller. On the contrary, the attacker wants to minimize ρ to increase the detection error.

D. Problems of Interest

The following two problems are to be addressed:

1) What is a reasonable performance metric that can characterize the detection performance when the attacker is present?
2) Suppose that the types of the deployed sensors can be selected from a sensor pool, i.e., the distribution pairs of (μ_i,ν_i) (and hence m) in Section II-A are to be determined. Then from the perspective of a system manager, how to deploy sensors, with a sensor cost budget, such that the performance metric is maximized?

III. Preliminary: Large Deviation Theory

In this section, we first introduce the large deviation theory, which is a key supporting technique of this paper. We then impose two quite weak assumptions on the observations of a sensor.

To proceed, we first introduce some definitions. Let $M_\psi(w) \triangleq \int_{\mathbb{R}^p} e^{w\cdot X}\mathrm{d}\psi(X), w\in\mathbb{R}^p$ be the moment generating function for the random vector $X\in\mathbb{R}^p$ that has the probability measure ψ, where $w\cdot X$ is the dot product. Let $\mathrm{dom}_\psi \triangleq \{w\in\mathbb{R}^p | M_\psi(w)<\infty\}$ be the support such that $M_\psi(w)$ is finite. Define the Fenchel–Legendre transform of the function $\log M_\psi(w)$ as

$$I_\psi(x) = \sup_{w\in\mathbb{R}^p}\{x\cdot w - \log M_\psi(w)\},\ x\in\mathbb{R}^p. \tag{9}$$

Theorem 1 (Multidimensional Cramér's Theorem [19])**.** *Suppose $X(1),\ldots,X(k),\ldots$ be a sequence of i.i.d. random vectors and $X(k)\in\mathbb{R}^p$ has the probability measure ψ. Let $\overline{X}(k)\triangleq\sum_{t=1}^k X(t)/k, k\in\mathbb{N}_+$ be the empirical mean. Then if $0\in\mathrm{int}(\mathrm{dom}_\psi)$, the probability $\mathbb{P}(\overline{X}(k)\in\mathcal{A})$ with $\mathcal{A}\subseteq\mathbb{R}^p$ satisfies the large deviation principle, i.e.,*

1) *if $\mathcal{A}$ is closed,*

$$\limsup_{k\to\infty}\frac{1}{k}\log\mathbb{P}(\overline{X}(k)\in\mathcal{A}) \le -\inf_{x\in\mathcal{A}} I_\psi(x).$$

2) *if $\mathcal{A}$ is open,*

$$\liminf_{k\to\infty}\frac{1}{k}\log\mathbb{P}(\overline{X}(k)\in\mathcal{A}) \ge -\inf_{x\in\mathcal{A}} I_\psi(x).$$

To apply the multidimensional Cramér's Theorem, we make assumptions concerning the observations of a sensor as follows. The moment generating function of the log-likelihood ratio $\lambda_i(y_i)$ under $\theta=0$ is given by $M_{0,i}(w)\triangleq\int_{\mathbb{R}} e^{w\lambda_i(y_i)}\mathrm{d}\nu_i(y_i), w\in\mathbb{R}$. The support such that $M_{0,i}(w)<\infty$ is denoted by $\mathrm{dom}_{0,i}$. The quantities $M_{1,i}(w)$ and $\mathrm{dom}_{1,i}$ are defined similarly. We assume that both $\mathrm{dom}_{0,i}$ and $\mathrm{dom}_{1,i}$ contain 0 as an interior point. This is formalized as follows.

Assumption 3. *For any sensor i, there hold $0\in\mathrm{int}(\mathrm{dom}_{0,i})$ and $0\in\mathrm{int}(\mathrm{dom}_{1,i})$.*

Denote the the Kullback-Leibler (K–L) divergences by $\mathrm{KL}(1,i)\triangleq\int_{\mathbb{R}}\lambda_i(y_i)\mathrm{d}\mu_i(y_i)$ and $\mathrm{KL}(0,i)\triangleq-\int_{\mathbb{R}}\lambda_i(y_i)\mathrm{d}\nu_i(y_i)$. To avoid degenerate problems, we assume

Assumption 4. *The K–L divergences are well-defined for any sensor i, i.e., $0<\mathrm{KL}(1,i)<\infty$ and $0<\mathrm{KL}(0,i)<\infty$.*

IV. Performance Metric: a Game-theoretic Approach

In this section, we assume that the set of sensors deployed to collect observations are given, the model of which is the same as in Section II-A. With the sensors fixed, we model the system's strategy f and the attacker's strategy g as a zero-sum game. We then identify a Nash equilibrium (NE) of strategy pair (f^*,g^*). Since $\rho(f,g)$ is unique for any NE (f,g) in our game, we adopt $\rho(f^*,g^*)$ as the performance metric.

We detail the game model $\mathbf{G}$ as follows. The players are the system manager and the attacker, which can adopt any admissible strategy f and g, respectively. The payoff for the system manager and the attacker playing (f,g) is respectively $\rho(f,g)$ and $-\rho(f,g)$.

To present the NE (f^*,g^*), we need the following definitions. For the log-likelihood ratio $\lambda_i(y_i)$, we use $I_{0,i}(x)$ and $I_{1,i}(x)$ to denote As suggested by the reviewer, we have also added a table in the end of introduction. its rate function when $y_i(k)$ follows the distribution ν_i and μ_i, respectively. Given any non-empty set $\mathcal{O}\subset\mathcal{M}$, let $\lambda_\mathcal{O}(y_\mathcal{O})=\sum_{i\in\mathcal{O}}\lambda_i(y_i)$, and $I_{0,\mathcal{O}}(x)$ ($M_{0,\mathcal{O}}(w)$, respectively) be the rate function (moment generating function, respectively) for $\lambda_\mathcal{O}$ when for every $i\in\mathcal{O}$, $y_i(k)$ follows the distribution ν_i. The terms $I_{1,\mathcal{O}}(x)$ and $M_{1,\mathcal{O}}(w)$ are defined similarly. Further define $\mathrm{KL}(j,\mathcal{O})\triangleq\sum_{i\in\mathcal{O}}\mathrm{KL}(j,i)$ for $j\in\{0,1\}$. Here we list some results concerning the above quantities, the proof of which can be found in our work [13].

Theorem 2. *To simplify presentations, $\mathfrak{i}$ might be any element or any subset of the set $\mathcal{M}$. With Assumption 3, the followings hold:*

1) $I_{0,\mathfrak{i}}(0)=I_{1,\mathfrak{i}}(0)$.
2) *For any $x\in\mathbb{R}$, $I_{0,\mathfrak{i}}(x)\ge 0$ and $I_{1,\mathfrak{i}}(x)\ge 0$.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

3) $I_{0,\mathrm{i}}(x)$ ($I_{1,\mathrm{i}}(x)$) *is non-decreasing (non-increasing) on* $[-\mathrm{KL}(0,\mathrm{i}), \mathrm{KL}(1,\mathrm{i})]$.

To simplify presentations, we denote

$$C_i \triangleq I_{0,i}(0), \quad C_{\mathcal{O}} \triangleq I_{0,\mathcal{O}}(0). \tag{10}$$

In addition, we let $C_{\mathcal{O}} = 0$ if $\mathcal{O} = \emptyset$. Notice that C_i or $C_{\mathcal{O}}$ is just the Chernoff information [19, Corollary 3.4.6], which can be written as

$$C_i = -\inf_{w\in\mathbb{R}} \log \int_{\mathbb{R}} e^{w\lambda_i(y_i)} \mathrm{d}\nu_i(y_i),$$
$$C_{\mathcal{O}} = -\inf_{w\in\mathbb{R}} \log \int_{\mathbb{R}^{|\mathcal{O}|}} e^{w\sum_{i\in\mathcal{O}}\lambda_i(y_i)} \prod_{i\in\mathcal{O}} \mathrm{d}\nu_i(y_i),$$

respectively.

Let $\mathcal{H}$ be the set of sensors such that

$$\mathcal{H} = \operatorname*{arg\,min}_{\{\mathcal{O}:|\mathcal{O}|\geq(m-2n)^+\}} C_{\mathcal{O}}, \tag{11}$$

where $(m-2n)^+ = \max(m-2n, 0)$ with the understanding that if $(m-2n)^+ = 0$, $\mathcal{H} = \emptyset$. Roughly speaking, $\mathcal{M} \setminus \mathcal{H}$ provides the best information quality among the sets that consist of no more than $2n$ sensors, and, thus, will be attacked by the adversary. Therefore, the sensors in $\mathcal{H}$, which are less informative, will not be attacked by the adversary.

We now present the equilibrium attack strategy g^*. Let $\mathcal{I}', \mathcal{I}'' \subset \mathcal{M} \setminus \mathcal{H}$ be two sets such that $\mathcal{I}' \cup \mathcal{I}'' = \mathcal{M} \setminus \mathcal{H}$. The attack strategy g^* is as follows.

(i). When $\theta = 0$, sensors in $\mathcal{I}'$ are compromised and the distributions are flipped, i,e., the measurements of sensor i in $\mathcal{I}'$ are i.i.d. as μ_i.

(ii). When $\theta = 1$, sensors in $\mathcal{I}'' \setminus \mathcal{I}'$ are compromised and the distributions are flipped, i.e., the measurements of sensor i in $\mathcal{I}'' \setminus \mathcal{I}'$ are i.i.d. as ν_i.

We then present the equilibrium detection strategy f^*: at each time k, it is implemented as follows:

1) Compute the empirical mean of the likelihood ratio from time 1 to time k for each sensor i:

$$\begin{aligned}\bar{\lambda}_i(k) &\triangleq \sum_{t=1}^{k} \lambda_i(\tilde{y}_i(t))/k \\ &= \frac{k-1}{k}\bar{\lambda}_i(k-1) + \frac{1}{k}\lambda_i(\tilde{y}_i(k))\end{aligned} \tag{12}$$

with $\bar{\lambda}_i(0) = 0$.

2) Compute $I_{0,i}(\bar{\lambda}_i(k))$ and $I_{1,i}(\bar{\lambda}_i(k))$ for each i. Then compute the following quantities:

$$\delta(0,k) \triangleq \min_{\mathcal{O}\subset\mathcal{M},|\mathcal{O}|=m-n} \sum_{i\in\mathcal{O}} I_{0,i}(\bar{\lambda}_i(k)), \tag{13}$$

$$\delta(1,k) \triangleq \min_{\mathcal{O}\subset\mathcal{M},|\mathcal{O}|=m-n} \sum_{i\in\mathcal{O}} I_{1,i}(\bar{\lambda}_i(k)). \tag{14}$$

3) If $\delta(0,k) \leq \delta(1,k)$, make a decision $\hat{\theta} = 0$; make a decision $\hat{\theta} = 1$ otherwise.

We are ready to present the main result and the proof is presented in Appendix A.

Theorem 3. *The system's strategy f^* and the attacker's strategy g^* have the following properties:*

1) *The pair (f^*, g^*) is a Nash equilibrium of the game $\mathbf{G}$, i.e., the following holds: for any system strategy f and attack strategy g,*

$$\rho(f, g^*) \leq \rho(f^*, g^*) \leq \rho(f^*, g).$$

2) $\rho(f^*, g^*) = C_{\mathcal{H}}$.

It is well known that $\rho(f, g) = C_{\mathcal{H}}$ for any equilibrium of strategy profile (f, g) since $\mathbf{G}$ is a two-player zero-sum game [20]. Therefore, we adopt $C_{\mathcal{H}}$ as the performance metric. This performance metric is reasonable since: 1) an intelligent and malicious adversary will launch the most dangerous attacks. 2) $C_{\mathcal{H}}$ is the *best* performance the system *can achieve* when faced with the *worst-case* n-sparse attack.

Remark 1. *Notice that in the equilibrium attack strategy g^*, the attacker confines itself to compromising the sensors in the set $\mathcal{M} \setminus \mathcal{H}$ and does not attack the sensors in $\mathcal{H}$ at all. Notice also that under g^*, the compromised sensors generate i.i.d. measurements, though we do not restrict the statistical pattern of a possible attack strategy. This is in accordance with the intuition that a powerful attacker would make the compromised sensors generate measurements with an identical distribution under two hypotheses, and, thus, provide no information to a detection system. Under attack strategy g^*, the measurements of the sensor i in $\mathcal{I}'$ ($\mathcal{I}'' \setminus \mathcal{I}'$, respectively) are i.i.d. as μ_i (ν_i, respectively) whatever θ is.*

Remark 2. *The computational complexity of steps 1) and 2) of f^* is $O(m)$ and $O(m\log(m))$, respectively. To compute $\delta(0,k)$, one can first sort $I_{0,i}(\bar{\lambda}_i(k))$ in ascending order and then sum the first $m-n$ elements. The quantity $\delta(1,k)$ can be computed similarly. Therefore, the total computational complexity for each time step is $O(m\log m)$.*

Remark 3. *Notice that when $n \geq m/2$, $\rho(f^*, g^*) = C_{\mathcal{H}} = 0$ holds. In other words, when not less than half of the total sensors are compromised, the detection system would be blind. Such "breakdown at $1/2$" phenomenon actually exist pervasively in literature that assume a fusion center processing information collected from possible Byzantine sensors, e.g., binary hypothesis testing [9], [10], state estimation [2], [16]–[18], and quickest change detection [15]. This is in accordance with the intuition that an attack capable of compromising n sensors will make $2n$ sensors behave the same under two different states, hence, provide no information.*

V. Sensor Deployment Strategy

In this section, we address the following problem: with a fixed sensor cost budget, how to deploy sensors such that the resulting $C_{\mathcal{H}}$ is maximized. The problem is formulated and is difficult to solve. We thus provide two heuristics: 1) We relax $C_{\mathcal{O}}$ to its upper bound $\sum_{i\in\mathcal{O}} C_i$ and solve the problem using a group of unbounded knapsack problems (UKPs). This relaxation is indeed exact in certain scenarios, for which sufficient conditions and examples are given. 2) We consider using identical sensors and prove that the performance metric gap between the optimal sensor deployment and the optimal identical sensor deployment is within a fixed constant for any

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

cost budget. We also show that how the results obtained in the above two approximate algorithms can be leveraged to reduce the computation burden of obtaining the optimal sensor deployment strategy.

Let $s \in \mathbb{N}_+$ be the number of the types of available sensors. For simplicity, it is assumed that the number of each type of available sensors is unlimited. Let $d \in \mathbb{N}^s$ denote the sensor deployment strategy with d_p being the number of the p-th type of sensors deployed. Given a sensor deployment strategy d, $C(d)$ denotes its associated $C_{\mathcal{H}}$ introduced in Section IV.

For a sensor of p-th type, we use the Chernoff information defined in (10), denoted by ${c_p}^2$, to quantify its contributions. The costs of a p-th type of sensor (e.g., manufacturing cost and deployment cost) is denoted by u_p. In the sequel, we assume

Assumption 5.

1) *If $u_p \geq u_q$, then $c_p \geq c_q$ for any $1 \leq p, q \leq s$.*
2) *Without loss of any generality, assume that $u_1 \leq u_2 \leq \cdots \leq u_s$.*

The above assumption indicates that: "better" sensors incur more costs. Let U be the sensor cost budget. Define $\mathcal{D}_U$ as the set of admissible sensor deployment strategies:

$$\mathcal{D}_U \triangleq \{d \in \mathbb{N}^s : \sum_{p=1}^{s} d_p u_p \leq U\}. \tag{15}$$

The problem of interest then can be formalized as follows:

$$\textbf{(P1)} \qquad \begin{aligned} \max_{d} \quad & C(d) \\ \text{subject to} \quad & d \in \mathcal{D}_U \end{aligned}$$

A. Unbounded Knapsack Problems Formulation

The main difficulty with the problem **(P1)** lies in the complicated expression of $C(d)$. To circumvent this, we adopt the following approximation:

$$C_{\mathcal{O}} \approx \sum_{i \in \mathcal{O}} C_i, \tag{16}$$

where C_i and $C_{\mathcal{O}}$ are defined in (10). By the following theorem, we know that adopting the approximation in (16) will maximize a upper bound of $C(d)$.

Theorem 4. *It always holds that*

$$C_{\mathcal{O}} \leq \sum_{i \in \mathcal{O}} C_i. \tag{17}$$

If for any i and j in a index set $\mathcal{O}$, the following equality holds:

$$I_{0,i}^{(1)}(0) = I_{0,j}^{(1)}(0) \tag{18}$$

then we have

$$C_{\mathcal{O}} = \sum_{i \in \mathcal{O}} C_i. \tag{19}$$

[2]This is different from C_i, which denotes the Chernoff information of the i-th sensor (instead of i-th type of sensor) of a bunch of deployed sensors.

Proof. Since the observations of different sensors are independent, one has that for any $w \in \mathbb{R}$, $M_{0,\mathcal{O}}(w) = \prod_{i \in \mathcal{O}} M_{0,i}(w)$. Therefore

$$\begin{aligned} C_{\mathcal{O}} &= \sup_{w \in \mathbb{R}} \Big\{ -\sum_{i \in \mathcal{O}} \log M_{0,i}(w) \Big\} \\ &\leq \sum_{i \in \mathcal{O}} \sup_{w \in \mathbb{R}} \{-\log M_{0,i}(w)\} = \sum_{i \in \mathcal{O}} C_i, \end{aligned} \tag{20}$$

which proves (17). Furthermore, notice that

$$I_{0,i}^{(1)}(0) = \arg\sup_{w \in \mathbb{R}^n} \{-\log M_{0,i}(w)\}. \tag{21}$$

Then if (18) holds, the inequality in (20) becomes equality, which proves (19). □

Remark 4. *Notice that if there exists "symmetry" between distribution μ_i and ν_i, then $I_{0,i}^{(1)}(0) = 0.5$ holds for any i [13] and, therefore,* (19) *holds. The symmetry is formalized as follows: if for each sensor i, there exists a constant a_i such that for any Borel measurable set $\mathcal{A}$, we have*

$$\mu_i(a_i + \mathcal{A}) = \nu_i(a_i - \mathcal{A}).$$

One example of the symmetric distribution arises in detecting the mean shift in Gaussian noises. Specifically, each $y_i(k)$ satisfies the following equation:

$$y_i(k) = a_i\theta + v_i(k),$$

where $a_i \neq 0$ and $v_i(k) \sim \mathcal{N}(\bar{v}_i, \sigma_i^2)$ is i.i.d. Gaussian distributed.

Under the approximation in (16) and Assumption 5, an explicit expression of $C(d)$ can be obtained. To this end, for any sensor deployment strategy d, let $\tau(d)$ be the type of its $(2n+1)$-th best sensor, i.e., $\tau(d)$ is chosen such that:

$$\sum_{p=\tau(d)+1}^{s} d_p \leq 2n, \ \sum_{p=\tau(d)}^{s} d_p > 2n. \tag{22}$$

Further define $\phi(d)$ as the number of the remaining $\tau(d)$-th type of sensors after "removing" the $2n$ best sensors, i.e.,

$$\phi(d) \triangleq \sum_{p=\tau(d)}^{s} d_p - 2n. \tag{23}$$

Then one verifies that the problem **(P1)** can be transformed to:

$$\textbf{(P2)} \qquad \begin{aligned} \max_{d} \quad & \sum_{p=1}^{\tau(d)-1} d_p c_p + \phi(d) c_{\tau(d)} \\ \text{subject to} \quad & d \in \mathcal{D}_U \end{aligned}$$

The above problem is still quite complicated to solve since $\tau(d)$ is nonconvex with respect to d_p. In the following, we transform it *equivalently* to a group of UKPs. With Assumption 5, one easily obtains the following lemma:

Lemma 1. *If d is a solution to problem* **(P2)**, *then the following d' is also a solution:*

$$d'_p = d_p, \ \textit{for } 1 \leq p \leq \tau(d) - 1 \tag{24}$$

$$d'_{\tau(d)} = \phi(d) + 2n \tag{25}$$

$$d'_p = 0, \ \textit{for } p \geq \tau(d) + 1 \tag{26}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

The intuition of Lemma 1 is as follows. The $2n$ most expensive sensors will not contribute to the detection performance. Therefore, we should choose them to be as cheap as possible in the sense that they are the same as the $(2n+1)$-th most expensive one.

For $1 \leq q \leq s$, suppose that the $2n$ most expensive sensors are of the q-th type, then Lemma 1 gives that **(P2)** can be transformed to

$$\begin{aligned} \mathbf{C}_q \triangleq \max_{d} \quad & \sum_{p=1}^{q-1} d_p c_p + (d_q - 2n) c_q \\ \text{s.t.} \quad & d_p = 0, \text{ for } p > q, \text{ and } d_q \geq 2n \\ & d \in \mathcal{D}_U. \end{aligned}$$

Now to obtain the optimal solution to **(P2)**, we enumerate q over $\{1, \ldots, s\}$ to find the maximum $\mathbf{C}_q$.

Remark 5. *Notice that obtaining $\mathbf{C}_q$ is an UKP, which can be solved using dynamic programming or branch and bound techniques. The computational complexity of a dynamic programming approach is $O(qU)$. The computational complexity of obtaining the maximal one among $\{\mathbf{C}_q\}_{1\leq q\leq s}$ is $O(s)$. Therefore, the total computational complexity of obtaining the optimal solution of* **(P2)** *is $O(s^2U)$.*

In the following, we provide a sufficient condition under which the strategy obtained by the above UKP algorithm is indeed optimal.

Proposition 1. *Let d^{ukp} be the solution to* **(P2)**. *If* (18) *holds for any i, j in the set* $\mathrm{supp}(d^{\mathrm{ukp}})$*, then d^{ukp} is also the solution to* **(P1)**.

Proof. Given a sensor deployment strategy d, we denote by $\overline{C}(d)$ the objective function in **(P2)**, i.e.,

$$\overline{C}(d) \triangleq \sum_{p=1}^{\tau(d)-1} d_p c_p + \phi(d) c_{\tau(d)}. \tag{27}$$

Then from the definition of $C(d)$ and (17), it follows that for any strategy d:

$$C(d) \leq \overline{C}(d). \tag{28}$$

Moreover, since (18) holds for any i, j in the set $\mathrm{supp}(d^{\mathrm{ukp}})$, then by Theorem 4, one obtains that

$$C(d^{\mathrm{ukp}}) = \overline{C}(d^{\mathrm{ukp}}). \tag{29}$$

Notice also that d^{ukp} is the solution to **(P2)**, i.e., for any admissible strategy $d \in \mathcal{D}_U$, there holds

$$\overline{C}(d) \leq \overline{C}(d^{\mathrm{ukp}}). \tag{30}$$

Then combing (28)–(30), one concludes Proposition 1. □

In practice, instead of computing the optimal algorithm for **(P1)** directly, which is computation-heavy, one may first obtain the strategy d^{ukp} and check whether it is optimal based on the above proposition. Notice that if d^{ukp} only contain one type of sensors, it will surely satisfy the condition and be optimal.

However, in principle the performance gap between d^{ukp} and the solution to **(P1)** can not be assured. To this end, we propose another heuristic algorithm in the next subsection. Nevertheless, we shall show later in Sections V-C and VI that d^{ukp} will help reduce the search space in computing the solution to **(P1)** significantly.

B. Identical Sensors

In this section, we propose a different heuristic and prove that the performance metric gap between the optimal sensor deployment and the optimal one with sensors being identical is within a fixed constant for any cost budget. Notice that the results in this subsection do *not* rely on the approximation in (16).

Let d^{op} be the solution to **(P1)** and d^{ide} the optimal identical sensors deployment strategy. Then we have the following theorem and the proof is presented in Appendix B.

Theorem 5. *For any $U \geq 0$, there holds $C(d^{\mathrm{ide}}) > C(d^{\mathrm{op}}) - c_{p*}$, where $p^* \triangleq \arg\max_{1\leq p\leq s} c_p/u_p$.*

Notice that the optimal identical sensors deployment strategy d^{ide} can be obtained by solving

$$\max_{p\in\{1,\ldots,s\}} (\lfloor U/u_p \rfloor - 2n) c_p,$$

the computational complexity of which is $O(s)$.

C. Further Discussions

In this subsection, we show how d^{ukp} and d^{ide} can be utilized to reduce the burden of computing d^{op}. In particular, given a strategy $d \in \mathcal{D}_U$, to compute its performance $C(d)$, one need to compute the Chernoff information for every possible set of sensors that is obtained by removing $2n$ sensors, which is of combinatorial nature and is computation-heavy. Therefore, it would be greatly beneficial to prune the set D_U. In the following, we provide some necessary conditions that an optimal sensor deployment strategy should satisfy, by which the set D_U can be pruned. Our simulations in the next section show that this pruning is indeed significant.

Proposition 2. *The optimal strategy d^{op} should satisfy the following:*

1) $\sum_{p=1}^{s} d_p^{\mathrm{op}} u_p g > U - u_1$.
2) $\overline{C}(d^{\mathrm{op}}) \geq \max\{C(d^{\mathrm{ukp}}), C(d^{\mathrm{ide}})\}$.

Proof.

1) Since if $\sum_{p=1}^{s} d_p u_p g \leq U - u_1$, then one may just deploy one more sensor of type 1. The resulting performance would be better.
2) This directly follows from (28) and the fact the optimal performance is lower bounded by $C(d^{\mathrm{ukp}})$ and $C(d^{\mathrm{ide}})$.

□

VI. Numerical Examples

Example 1. *We show that $\rho(f^*, g^*) = C_{\mathcal{H}}$ holds. To this end, we assume that there are totally $m = 10$ sensors deployed. The first 5 sensors are of the same type: the observations are Bernoulli distributed under each hypothesis, i.e., $\mathbb{P}_0(y_i = 1) = 0.2$ and $\mathbb{P}_1(y_i = 1) = 0.6$. The remaining 5 sensors also*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

have Bernoulli distributed observations: $\mathbb{P}_0(y_i = 1) = 0.2$ *and* $\mathbb{P}_1(y_i = 1) = 0.8$. *We let* n *vary from* 1 *to* 4 *and simulate* $\rho(f^*, g^*)$. *To simulate the performance with high accuracy, we adopt the importance sampling approach [21]. The results are given in Table I, where the simulated* $\rho(f^*, g^*)$ *is quite close to* $C_{\mathcal{H}}$.

TABLE I: $\rho(f^*, g^*)$ and $C_{\mathcal{H}}$ for different n.

n	$C_{\mathcal{H}}$	$\rho(f^*, g^*)$
1	1.1297	1.1294
2	0.6835	0.6831
3	0.3685	0.3683
4	0.1842	0.1841

Example 2. *In this example, we assume there are three different types of sensors, the details of which are summarized in Table II. It is assumed that* $n = 3$ *and 200 different valued* U *are randomly chosen from in* $[10, 350]$. *In Fig. 1, we plot performance of different deployment strategies:* d^{ukp} *the solution to* **(P2)**, d^{ide} *the optimal identical sensor deployment strategy, and* d^{op} *the solution to* **(P1)**. *To make the simulation results clearer, the performance gap is also illustrated. One may see that both* $C(d^{\mathrm{ukp}})$ *and* $C(d^{\mathrm{ide}})$ *are quite close to* $C(d^{\mathrm{op}})$ *for every cost budget* U. *In particular, the performance gap between* d^{ide} *and* d^{op} *satisfies:*

$$C(d^{\mathrm{op}}) - C(d^{\mathrm{ide}}) < 0.25 < c_{p^*} = 0.2762,$$

which verifies Theorem 5.

Given a cost budget U, *to obtain the optimal sensor deployment strategy* d^{op}, *instead of computing the performance* $C(d)$ *for every strategy* d *in* $\mathcal{D}_U$, *we leverage* d^{ukp} *and* d^{ide} *as in Propositions 1 and 2. The numbers of remaining strategies needed to evaluate after the optimality check of* d^{ukp} *and the pruning procedure for different budget* U *are plotted in Fig. 2. The numbers are rather small in all cases, and, thus, the computation complexity is reduced significantly considering that computing* $C(d)$ *is quite computation-heavy.*

TABLE II: Information of sensors.

type	$\mathbb{P}_0(y_i = 1)$	$\mathbb{P}_1(y_i = 1)$	c	$I^{(1)}_{0,i}(0)$	u	c/u
1	0.2	0.8	0.2231	0.5	1.5	0.1487
2	0.8	0.15	0.2762	0.49	1.7	0.1625
3	0.05	0.65	0.2832	0.56	1.8	0.1573

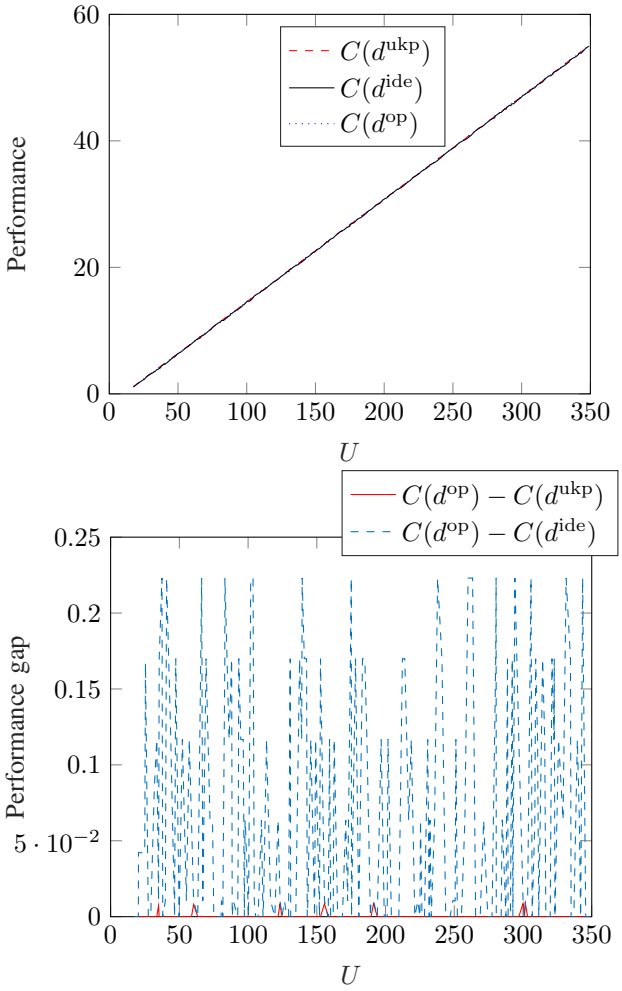


Fig. 1: Upper: Performance of different sensor deployment strategies as functions of the total cost budget U. Lower: Performance gap between d^{ukp} (or d^{ide}) and d^{op}.

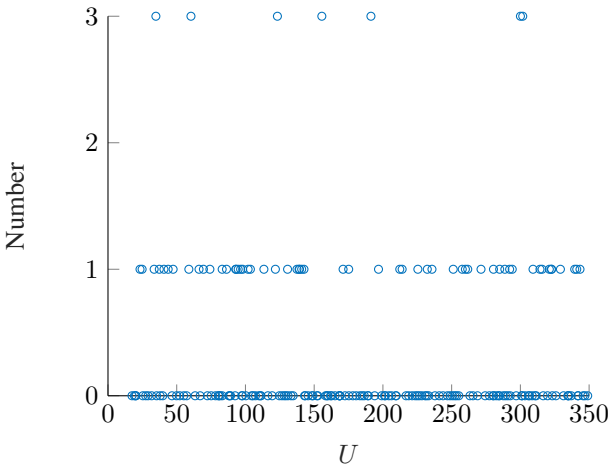


Fig. 2: Number of remaining sensor deployment strategies that are required to evaluate after the optimality check of d^{ukp} as in Proposition 1 and the pruning indicated in Proposition 2.

VII. Conclusion and Future Work

In this paper, we study the sensor deployment strategy in the context of detection in adversarial environments. The deployed heterogeneous sensors send measurements to a fusion center to perform a binary hypothesis testing, among which some may be compromised by an attacker and their measurements can be arbitrarily manipulated. We first formulate the detection with an attacker present as a zero-sum game and present a Nash equilibrium pair of the detector and attack strategy. A performance metric is provided as the detection performance when the equilibrium pair is played. We then assume that the system manager, with a sensor cost budget, needs to decide

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

what sensors to deploy such that the above performance metric is maximized. The problem is difficult to solve. Therefore, we provide two heuristics: 1) A upper bound of the performance metric is used as the objective and the problem is solved by a group of unbounded knapsack problems (UKPs). We further give sufficient conditions and examples where the relaxation is indeed exact. 2) We consider using identical sensors and show that the performance metric gap between the optimal sensor deployment and the optimal one with sensors being of same type is within a fixed constant for any cost budget. Notice that this result does not rely on the approximation techniques used in the above UKP formulation. Future works include investigating sensor deployment strategies with Byzantine attacks in other settings, such as networked estimation and distributed optimization.

APPENDIX A
THE PROOF OF THEOREM 3

We present two lemmas (i.e., Lemmas 2 and 3) on the properties of g^* and f^*, based on which the NE (f^*, g^*) and the equality $\rho(f^*, g^*) = C_{\mathcal{H}}$ are established.

Lemma 2. *For any detection strategy f, $\rho(f, g^*) \leq C_{\mathcal{H}}$ holds.*

Proof. One obtains that under attack strategy g^*, for either $\theta = 1$ or $\theta = 0$, sensor i with $i \in \mathcal{I}'$ will follow the distribution μ_i and sensor i with $i \in \mathcal{I}'' \setminus \hat{\mathcal{I}}$ will follow the distribution ν_i. In other words, only sensors in $\mathcal{H}$ have different distributions when θ is different. Notice that when $m \leq 2n$, $\mathcal{H} = \emptyset$, which means that every sensor will have the exact same distribution regardless of whether $\theta = 0$ or $\theta = 1$. Therefore, $\rho(f, g^*) = 0$ for any g.

When $m > 2n$, consider a detection strategy $f^\dagger = (f_1^\dagger, f_2^\dagger, \cdots)$ as follows[3]:

$$f_k^\dagger(\tilde{\mathbf{y}}(k)) = \begin{cases} 0 & \text{if } \sum_{i\in\mathcal{H}} \bar{\lambda}_i(k) < 0, \\ 1 & \text{if } \sum_{i\in\mathcal{H}} \bar{\lambda}_i(k) \geq 0, \end{cases} \tag{31}$$

where $\bar{\lambda}_i(k)$, the empirical mean of the likelihood ratio from time 1 to time k for each sensor i, is defined in (12). It is well known from the classical Bayesian detection theory [22] that, the detector $f^\dagger$ is optimal against the attacker's strategy g^*, in the sense that for any f:

$$\begin{aligned} &\mathbb{E}_0 f_k^\dagger(\tilde{\mathbf{y}}(k)) + \left(1 - \mathbb{E}_1 f_k^\dagger(\tilde{\mathbf{y}}(k))\right) \\ &\leq \mathbb{E}_0 f_k(\tilde{\mathbf{y}}(k)) + (1 - \mathbb{E}_1 f_k(\tilde{\mathbf{y}}(k))), \ \forall k \geq 1. \end{aligned} \tag{32}$$

Notice that for any f and k, there holds

$$\begin{aligned} &\log\left(\mathbb{E}_0 f_k(\tilde{\mathbf{y}}(k)) + (1 - \mathbb{E}_1 f_k(\tilde{\mathbf{y}}(k)))\right) \\ &\leq \log\max\left(\mathbb{E}_0 f_k(\tilde{\mathbf{y}}(k)), 1 - \mathbb{E}_1 f_k(\tilde{\mathbf{y}}(k))\right) + \log 2. \end{aligned}$$

Therefore, (32) implies that for any f:

$$\begin{aligned} &\lim_{k\to\infty} -\frac{\log\max\left(\mathbb{E}_0 f_k(\tilde{\mathbf{y}}(k)), 1 - \mathbb{E}_1 f_k(\tilde{\mathbf{y}}(k))\right)}{k} \\ &\leq \lim_{k\to\infty} -\frac{\log\max\left(\mathbb{E}_0 f_k^\dagger(\tilde{\mathbf{y}}(k)), 1 - \mathbb{E}_1 f_k^\dagger(\tilde{\mathbf{y}}(k))\right)}{k} \\ &= C_{\mathcal{H}}, \end{aligned} \tag{33}$$

[3]Notice that $f^\dagger$ is only optimal when g^* is used, and may have poor performance when another g is used.

where the last equality follows from Cramér's Theorem.

Therefore, for any f, there holds $\rho(f, g^*) \leq C_{\mathcal{H}}$, which completes the proof. □

Lemma 3. *For any attack strategy g, $\rho(f^*, g) \geq C_{\mathcal{H}}$ holds.*

Combining Lemmas 2 and 3, one easily obtains Theorem 3.

The remaining is devoted to the proof of Lemma 3. Notice that when $m \leq 2n$, Lemma 3 is trivial since $C_{\mathcal{H}} = 0$. Therefore, in the following, we focus on the case where $m > 2n$. We first present the following supporting definitions and lemmas.

Lemma 4.

1) *For any $\mathcal{O} \subset \mathcal{M}$, $x \in \mathbb{R}^m$ and $j \in \{0, 1\}$, it holds that*

$$I_{j,\mathcal{O}}\left(\sum_{i\in\mathcal{O}} x_i\right) \leq \sum_{i\in\mathcal{O}} I_{j,i}(x_i). \tag{34}$$

2) *For any $\mathcal{O} \subset \mathcal{M}$, we have the following set inclusions:*

$$\left\{x \in \mathbb{R}^m : \sum_{i\in\mathcal{O}} I_{0,i}(x_i) < C_{\mathcal{O}}\right\} \subseteq \left\{x \in \mathbb{R}^m : \sum_{i\in\mathcal{O}} x_i < 0\right\}, \tag{35}$$

$$\left\{x \in \mathbb{R}^m : \sum_{i\in\mathcal{O}} I_{1,i}(x_i) < C_{\mathcal{O}}\right\} \subseteq \left\{x \in \mathbb{R}^m : \sum_{i\in\mathcal{O}} x_i > 0\right\}. \tag{36}$$

Proof.

1) Notice that since all the measurements are independent from each other (to define $I_{j,\mathcal{O}}$), we can prove that for any $w \in \mathbb{R}$, $M_{j,\mathcal{O}}(w) = \prod_{i\in\mathcal{O}} M_{j,i}(w)$. Therefore,

$$\begin{aligned} I_{j,\mathcal{O}}\left(\sum_{i\in\mathcal{O}} x_i\right) &= \sup_{w\in\mathbb{R}}\left\{\sum_{i\in\mathcal{O}} x_i w - \sum_{i\in\mathcal{O}} \log M_{j,i}(w)\right\} \\ &\leq \sum_{i\in\mathcal{O}} \sup_{w\in\mathbb{R}}\{x_i w - \log M_{j,i}(w)\} = \sum_{i\in\mathcal{O}} I_{j,i}(x_i). \end{aligned}$$

2) We will focus on proving (35), since (36) can be proved in a similar manner. By (34), it is easy to see that

$$I_{0,\mathcal{O}}(\sum_{i\in\mathcal{O}} x_i) \leq \sum_{i\in\mathcal{O}} I_{0,i}(x_i) < C_{\mathcal{O}}$$

By the definition in (10), $I_{0,\mathcal{O}}(0) = C_{\mathcal{O}}$. By Theorem 2, $I_{0,\mathcal{O}}$ is non-decreasing on $[-\mathrm{KL}(0, \mathcal{O}), \mathrm{KL}(1, \mathcal{O})]$, we thus obtain that

$$I_{0,\mathcal{O}}(\sum_{i\in\mathcal{O}} x_i) < C_{\mathcal{O}} \Rightarrow \sum_{i\in\mathcal{O}} x_i < 0.$$

□

Let us introduce the following definition:

Definition 1. *Let $\mathcal{O} \subset \mathcal{M}$, $j \in \{0, 1\}$, define a ball as*

$$\mathrm{Bal}_j(\mathcal{O}) = \left\{x \in \mathbb{R}^m : \sum_{i\in\mathcal{O}} I_{j,i}(x_i) < C_{\mathcal{H}}\right\}.$$

We further define

$$\mathrm{EBal}_j \triangleq \bigcup_{|\mathcal{O}|=m-n} \mathrm{Bal}_j(\mathcal{O}).$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2853110, IEEE Transactions on Signal Processing

We then have the following lemma on the intersections of balls.

Lemma 5. $\mathrm{EBal}_0 \bigcap \mathrm{EBal}_1 = \emptyset$.

Proof. We will prove that for any $\mathcal{O}_0$ and $\mathcal{O}_1$ with cardinality $m-n$, $\mathrm{Bal}_0(\mathcal{O}_0) \bigcap \mathrm{Bal}_1(\mathcal{O}_1) = \emptyset$. To this end, let us define $\mathcal{O} = \mathcal{O}_0 \bigcap \mathcal{O}_1$. Therefore, $|\mathcal{O}| \geq m - 2n$.

Since each $I_{0,i}$ are non-negative by Theorem 2, if $x \in \mathrm{Bal}_0(\mathcal{O}_0)$, then

$$\sum_{i\in\mathcal{O}} I_{0,i}(x_i) \leq \sum_{i\in\mathcal{O}_0} I_{0,i}(x_i) < C_{\mathcal{H}} \leq C_{\mathcal{O}},$$

where the last inequality is due to the definition of $\mathcal{H}$ in (11). Now by Lemma 4, we know that $\sum_{i\in\mathcal{O}} x_i < 0$. Similarly, one can prove if $x \in \mathrm{Bal}_1(\mathcal{O}_1)$, then $\sum_{i\in\mathcal{O}} x_i > 0$. Therefore

$$\mathrm{Bal}_0(\mathcal{O}_0) \bigcap \mathrm{Bal}_1(\mathcal{O}_1) = \emptyset,$$

which finishes the proof. □

We now come to the main body of the proof of Lemma 3. Let $\bar{\lambda}(k) = (\bar{\lambda}_1(k), \ldots, \bar{\lambda}_m(k))$, then Lemma 5 means that for $j = \{0, 1\}$, if $\bar{\lambda}(k) \in \mathrm{EBal}_j$, i.e., $\delta(j,k) < C_{\mathcal{H}}$, then there holds $\delta(1-j,k) \geq C_{\mathcal{H}}$ and, thus, $\delta(1-j,k) > \delta(j,k)$, where $\delta(j,k)$ is defined in (13) and (14). Therefore,

$$f_k^*(\tilde{\mathbf{y}}(k)) = \begin{cases} 0 & \text{if } \bar{\lambda}(k) \in \mathrm{EBal}_0, \quad (37a) \\ 1 & \text{if } \bar{\lambda}(k) \in \mathrm{EBal}_1. \quad (37b) \end{cases}$$

The values of $f_k^*(\tilde{\mathbf{y}}(k))$ when $\lambda(k)$ is not in EBal_0 or EBal_1 do not affect the following analysis. Notice that if $x \in \mathrm{Bal}_j(\mathcal{M})$ and $\|x - x'\|_0 \leq n$, then

$$x' \in \mathrm{EBal}_j, \tag{38}$$

which implies that if the unmodified $(\bar{\lambda}_1, \ldots, \bar{\lambda}_m) \in \mathrm{Bal}_j(\mathcal{M})$ and no more than n sensors are compromised, then the compromised $(\bar{\lambda}_1, \ldots, \bar{\lambda}_m)$ will be in EBal_j. One thus obtains that under any attacks, there holds

$$\begin{aligned} &\limsup_{k\to\infty} \frac{1}{k} \log \tilde{\mathbb{P}}_0(f_k^* = 1) \\ &\leq \limsup_{k\to\infty} \frac{1}{k} \log \mathbb{P}_0(\bar{\lambda}_k \in \mathbb{R}^m \setminus \mathrm{Bal}_0(\mathcal{M})) \\ &\leq -\inf_{x\in\mathbb{R}^m\setminus\mathrm{Bal}_0(\mathcal{M})} \sum_{i=1}^m I_{0,i}(x_i) \\ &= -C_{\mathcal{H}}, \end{aligned} \tag{39}$$

where the first inequality follows from (37a) and (38); the second inequality holds because of the multi-dimensional Cramér's Theorem, and the fact that $\mathbb{R}^m \setminus \mathrm{Bal}_0(\mathcal{M})$ is closed and the observations at different sensors are independent.

Similarly, one obtains that under any attacks, there holds

$$\limsup_{k\to\infty} \frac{1}{k} \log \tilde{\mathbb{P}}_1(f_k^* = 0) \leq -C_{\mathcal{H}}. \tag{40}$$

Combing (39) and (40), one concludes Lemma 3.

APPENDIX B
THE PROOF OF THEOREM 5

Given the optimal solution d^{op}, define its most "economical" type of sensor to be

$$p^*(d^{\mathrm{op}}) \triangleq \arg\max_{p\leq\tau(d^{\mathrm{op}})} \quad c_p/u_p,$$

where the function $\tau(\cdot)$ is defined in (22). With abuse of notation, we use $p^*(d^{\mathrm{op}})$ to emphasize its dependence on the strategy used d^{op}, which is contrasted with the universally most economical type p^* defined in Theorem 5. We create a strategy d' using identical sensors. Let $d'_{p^*(d^{\mathrm{op}})} = \lfloor U/u_{p^*(d^{\mathrm{op}})} \rfloor$ with other elements of d' being zero. In the following, we derive the performance gap between d' and d^{op}.

Notice that we assume $u_1 \leq u_2 \leq \cdots \leq u_s$. Since none of the best $2n$ sensors in strategy d^{op} is cheaper than $u_{\tau(d^{\mathrm{op}})}$ and hence $u_{p^*(d^{\mathrm{op}})}$, one obtains that

$$(U/u_{p^*(d^{\mathrm{op}})} - 2n)u_{p^*(d^{\mathrm{op}})} \geq \sum_{p=1}^{\tau(d^{\mathrm{op}})-1} d_p^{\mathrm{op}} u_p + \phi(d^{\mathrm{op}}) u_{\tau(d^{\mathrm{op}})}, \tag{41}$$

where the function $\phi(\cdot)$ is defined in (23). Multiplying both sides of (41) by $c_{p^*(d^{\mathrm{op}})}/u_{p^*(d^{\mathrm{op}})}$, we get

$$(U/u_{p^*(d^{\mathrm{op}})} - 2n)c_{p^*(d^{\mathrm{op}})} \geq \sum_{p=1}^{\tau(d^{\mathrm{op}})-1} d_p^{\mathrm{op}} c_p + \phi(d^{\mathrm{op}}) c_{\tau(d^{\mathrm{op}})}, \tag{42}$$

where we use the fact that $p^*(d^{\mathrm{op}})$ is the most "economical" type of sensors. Furthermore, it follows from (17) that

$$\sum_{p=1}^{\tau(d^{\mathrm{op}})-1} d_p^{\mathrm{op}} c_p + \phi(d^{\mathrm{op}}) c_{\tau(d^{\mathrm{op}})} \geq C(d^{\mathrm{op}}). \tag{43}$$

Notice also that homogeneity of d' yields that

$$\begin{aligned} C(d') &= (\lfloor U/u_{p^*(d^{\mathrm{op}})} \rfloor - 2n)c_{p^*(d^{\mathrm{op}})} \\ &\geq (U/u_{p^*(d^{\mathrm{op}})} - 2n)c_{p^*(d^{\mathrm{op}})} + c_{p^*(d^{\mathrm{op}})} \end{aligned}$$

Combining (42) and (43), one obtains that $C(d') > C(d^{\mathrm{op}}) - c_{p^*(d^{\mathrm{op}})}$. Furthermore, it follows from the definitions of $p^*(d^{\mathrm{op}})$ and p^* that $p^*(d^{\mathrm{op}}) \leq p^*$ and, thus, $c_{p^*(d^{\mathrm{op}})} \leq c_{p^*}$. Notice also that $C(d^{\mathrm{ide}}) \geq C(d')$ holds by the optimality of d^{ide} among strategies that only use identical sensors. One thus concludes the proof.

REFERENCES

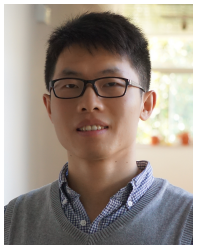
[1] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of Byzantine attacks in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774–786, 2011.

[2] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, 2015.

[3] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems*, vol. 35, no. 1, pp. 24–45, 2015.

[4] X. Ren, J. Wu, S. Dey, and L. Shi, "Attack allocation on remote state estimation in multi-systems: Structural results and asymptotic solution," *Automatica*, vol. 87, pp. 184–194, 2018.

[5] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal DoS attack scheduling in wireless networked control system," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 843–852, 2016.
[6] P. J. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965.
[7] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proceedings of the IEEE*, vol. 73, no. 3, pp. 433–481, 1985.
[8] P. J. Huber, *Robust statistics*. Springer, 2011.
[9] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attacks," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 16–29, 2009.
[10] Y. Mo, J. P. Hespanha, and B. Sinopoli, "Resilient detection in the presence of integrity attacks," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 31–43, 2014.
[11] K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo, "Detection in adversarial environments," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3209–3223, 2014.
[12] J. Yan, X. Ren, and Y. Mo, "Sequential detection in adversarial environment," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec 2017, pp. 170–175.
[13] X. Ren, J. Yan, and Y. Mo, "Binary hypothesis testing with byzantine sensors: Fundamental tradeoff between security and efficiency," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1454–1468, March 2018.
[14] A. Kerckhoffs, *La cryptographie militaire*. University Microfilms, 1978.
[15] G. Fellouris, E. Bayraktar, and L. Lai, "Efficient Byzantine sequential change detection," *IEEE Transactions on Information Theory*, 2017.
[16] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.
[17] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 49–59, 2017.
[18] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 2439–2444.
[19] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.
[20] M. J. Osborne and A. Rubinstein, *A course in game theory*. MIT press, 1994.
[21] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016.
[22] S. Key, "Fundamentals of statistical signal processing, Volume ii: Detection theory," 1993.



Xiaoqiang Ren received the B.E. degree in the Department of Control Science and Engineering from Zhejiang University, Hangzhou, China, in 2012 and the Ph.D. degree in the Department of Electronic and Computer Engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2016. He is currently a postdoctoral researcher at the Department of Automatic, KTH Royal Institute of Technology, Sweden. Prior to this, he was a postdoctoral researcher in the Hong Kong University of Science and Technology from September to November 2016 and Nanyang Technological University from December 2016 to February 2018. His research interests include security of cyber-physical systems, sequential detection, and networked estimation and control.

Yilin Mo is an Assistant Professor in the School of Electrical and Electronic Engineering at Nanyang Technological University. He received his Ph.D. In Electrical and Computer Engineering from Carnegie Mellon University in 2012 and his Bachelor of Engineering degree from Department of Automation, Tsinghua University in 2007. Prior to his current position, he was a postdoctoral scholar at Carnegie Mellon University in 2013 and California Institute of Technology from 2013 to 2015. His research interests include secure control systems and networked control systems, with applications in sensor networks and power grids.