CHALMERS
UNIVERSITY OF TECHNOLOGY

Design of false data injection attack on distributed process estimation

Citation for the original published paper (version of record):
Choraria, M., Chattopadhyay, A., Mitra, U. et al (2022). Design of false data injection attack on distributed process estimation. IEEE Transactions on Information Forensics and Security, 17: 670-683. http://dx.doi.org/10.1109/TIFS.2022.3146078

(article starts on next page)

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

Design of false data injection attack on distributed process estimation

Moulik Choraria, Arpan Chattopadhyay, Urbashi Mitra, Erik G. Ström

***Abstract*—Herein, design of false data injection attack on a distributed cyber-physical system is considered. A stochastic process with linear dynamics and Gaussian noise is measured by multiple agent nodes, each equipped with multiple sensors. The agent nodes form a multi-hop network among themselves. Each agent node computes an estimate of the process by using its sensor observation and messages obtained from neighboring nodes, via Kalman-consensus filtering. An external attacker, capable of arbitrarily manipulating the sensor observations of some or all agent nodes, injects errors into those sensor observations. The goal of the attacker is to steer the estimates at the agent nodes as close as possible to a pre-specified value, while respecting a constraint on the attack detection probability. To this end, a constrained optimization problem is formulated to find the optimal parameter values of a certain class of linear attacks. The parameters of linear attack are learnt on-line via a combination of stochastic approximation based update of a Lagrange multiplier, and an optimization technique involving either the Karush-Kuhn-Tucker (KKT) conditions or online stochastic gradient descent. The problem turns out to be convex for some special cases. Desired convergence of the proposed algorithms are proved by exploiting the convexity and properties of stochastic approximation algorithms. Finally, numerical results demonstrate the efficacy of the attack.**

***Index Terms*—Attack design, distributed estimation, CPS security, false data injection attack, Kalman-consensus filter, stochastic approximation.**

I. Introduction

In recent times, there have been significant interest in designing cyber-physical systems (CPS) that combine the cyber world and the physical world via seamless integration of sensing, computation, communication, control and learning. CPS has widespread applications such as networked monitoring and control of industrial processes, disaster management, smart grids, intelligent transportation systems, etc. These applications critically depend on estimation of a physical process via multiple sensors over a wireless network. However, increasing use of wireless networks in sharing the sensed data has rendered the sensors vulnerable to various cyber-attacks. In this paper, we focus on *false data injection* (FDI) attacks which is an integrity or deception attack where the attacker modifies the information flowing through the network [2], [3], in contrast to a *denial-of-service* attack where the attacker blocks system resources (*e.g.*, wireless jamming attack [4]). In FDI, the attacker either breaks the cryptography of the data packets or physically manipulates the sensors (*e.g.*, putting a heater near a temperature sensor).

In this paper, we design an attack algorithm that seeks to steer the estimates in all estimators towards a target value in a distributed estimation setting using a Kalman-consensus filter (KCF, see [5]), under a constraint on the attack detection probability. Solving this problem is important in understanding the attack schemes that can be used against a multi-agent system where the attacker seeks to induce the same control action on all agents (e.g., to cause accident in a vehicular system by pushing all vehicles towards one direction), so that suitable countermeasures can be developed. The attack scheme is reminiscent of the popular linear attack scheme [6], but the novelty lies in online learning and optimization of the parameters in the attack algorithm via Karush-Kuhn-Tucker (KKT) conditions, multi-timescale stochastic approximation [7] and simultaneous perturbation stochastic approximation (SPSA [8]). Convergence result is proved for the KKT-based algorithms, and the performance is demonstrated numerically.

A. Related literature

The cyber-physical systems either need to compute the process estimate in a remote estimator (*centralized* case), or often multiple nodes or components of the system need to estimate the same process over time via sensor observations and the information shared over a network (*distributed* case). The area of FDI attack on CPS has received significant attention in recent times [9]. Research on attack design includes developing conditions for undetectable FDI attack [10], design of a linear deception attack scheme to fool the popular χ^2 detector (see [6]), optimal attack design for noiseless systems [11], FDI design to penetrate AC-based bad data detection system [12], etc. The paper [13] designs an optimal attack to steer the state of a control system to a desired target under a constraint on the attack detection probability. Stealth attack design on a quantized networked control system has been solved in [14]. On the other hand, attempts on attack detection

Arpan Chattopadhyay is the co-first author.

Moulik Choraria is with the Department of Electrical and Computer Engineering, UIUC. Arpan Chattopadhyay is with the Department of Electrical Engineering and the Bharti School of Telecom Technology and Management, Indian Institute of Technology (IIT) Delhi. Urbashi Mitra is with the Department of Electrical Engineering, University of Southern California, Los Angeles. Erik G. Ström is with the Department of Signals and Systems, Chalmers University, Sweden. Email: moulik.choraria@epfl.ch, arpanc@ee.iitd.ac.in, ubli@usc.edu, erik.strom@chalmers.se

The work done by Arpan Chattopadhyay was supported by the faculty seed grant and professional development allowance (PDA) of IIT Delhi, and the seed grant from the I-Hub Foundation for Cobotics (IHFC).

The work by Urbashi Mitra was supported by the following grants: ONR N00014-15-1-2550, NSF CCF-1817200, ARO W911NF1910269, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, ONR 503400-78050.

The work by Erik G. Ström was supported in part by Ericsson Research Foundation.

This manuscript is an extended version of our conference paper [1].

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

include centralized (and decentralized as well) schemes for *noiseless* systems [15], coding of sensor output along with χ^2 detector [16], comparing the sensor observations with those coming from from a few *known safe* sensors [17], the attack detection and secure estimation schemes based on innovation vectors in [18], data driven design and detection of FDI [19], neural network based detection of FDI [20], quickest detection of time-varying false data injection attacks in dynamic linear regression models [21], FDI detection in linear parameter varying CPS [22], Gaussian mixture model based detection and secure state estimation [23], and quickest detection of FDI using Markov decision process formulation [24]. Attempts on attack-resilient state estimation include: [25] for *bounded* noise, [26]–[28] for adaptive filter design using stochastic approximation, [29] that uses sparsity models to characterize the switching location attack in a *noiseless* linear system and state recovery constraints for various attack modes. FDI attack and its mitigation in power systems are addressed in [30]–[32]. Attack-resilient state estimation and control in noiseless systems are discussed in [33] and [34]. Performance bound of stealthy attack in a single sensor-remote estimator system using Kalman filter was characterized in [35].

There have also been several recent attempts for attack mitigation in distributed CPS, such as [36] for attack detection and secure estimation, [37] for attack detection in networked control system using a certain *dynamic watermarking* strategy, the paper [38] for distributed Krein space based attack detection in discrete time-varying systems, and [39] for attack detection in power systems. On the other hand, there have also been several works that seek to design attacks against distributed CPS. Authors of [40] have designed an attack scheme to maximize the network-wide estimation error, which is different from our objective of pushing the estimates across nodes towards a target value, while respecting the attack detection constraint. Also, contrary to [40] which adds a simple Gaussian noise to the attacked node's observation, we focus on the class of linear attacks, and provide theoretical convergence results of our proposed online learning based attack schemes. Similarly, unlike our work, the authors of [41] developed conditions for perfect attack in a distributed control system, and also provided design algorithms for perfect and non-perfect attacks.

B. Our Contributions

Our contributions in this paper are the following:

1) Under KCF [5] for distributed estimation, we design a novel attack scheme that steers the estimates in all estimators towards a target value, while respecting a constraint on the attack detection probability under the popular χ^2 detector adapted to the distributed setting.
2) The dynamics of the deviation of the estimates from the target is derived analytically, which is used later to formulate the optimization problem. The updates turn out to be iterative in nature, and this was not available in prior literature.
3) The FDI design problem is cast as an online learning and optimization problem, and solved via KKT conditions (alternatively, SPSA) in the faster timescale to find optimal attack parameters, and by updating a Lagrange multiplier via stochastic approximation at a slower timescale to meet the constraint on the attack detection probability. SPSA is used for online stochastic gradient descent based learning of attack parameters (see [42, Chapter 3]).
4) Theoretical convergence results are proved for KKT-based algorithms. The key challenges in this proof were handling (i) multi-timescale updates, (ii) Markovian evolution of attack parameters, (iii) certain offset terms.
5) Interestingly, the attack algorithms, unlike the linear attack scheme of [6], use a *non-zero mean perturbation* to modify the observation made at a node, and this non-zero mean is an affine function of the process estimate at a node, which was not proposed before in the literature. Next, it is also shown that the optimal solution requires the perturbation to be deterministic, which is counter-intuitive.
6) These works are also extended to the case where the attacker has access to the FDI alarm at each node.

C. Organization

The rest of the paper is organized as follows. System model and the necessary background related to the problem are provided in Section II. Error dynamics expressions under FDI are calculated in Section III. Attack design algorithms are developed in Section IV via KKT conditions, and in Section V via SPSA. Numerical results are presented in Section VI, followed by the conclusions in Section VII. All proofs are provided in the appendices.

II. System Model

In this paper, bold capital letters, bold small letters and capital letters with caligraphic font will denote matrices, vectors and sets respectively. The notation $||\cdot||$, $(\cdot)'$ and $\mathbb{E}(\cdot)$ denote 2-norm, transpose and expectation, respectively.

A. Sensing and estimation model: no attack

We consider a connected, undirected, multi hop wireless network of N agent nodes denoted by $\mathcal{N} \doteq \{1, 2, \cdots, N\}$. The set of neighboring nodes of node k is denoted by $\mathcal{N}_k$, and let $N_k \doteq |\mathcal{N}_k|$. There is a discrete-time stochastic process $\{\boldsymbol{x}(t)\}_{t\geq 0}$ (where $\boldsymbol{x}(t) \in \mathbb{R}^{q\times 1}$ with process dimension q) which is a linear process with Gaussian noise evolving as follows:

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \underbrace{\boldsymbol{w}(t)}_{\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{Q})} \tag{1}$$

where $\boldsymbol{w}(t)$ is zero-mean i.i.d. Gaussian noise with covariance matrix $\boldsymbol{Q}$, and $\boldsymbol{A} \in \mathbb{R}^{q\times q}$ is the process matrix with its spectral radius strictly less than 1.

Each agent node is equipped with one or more sensors which make some observation about the process. The vector

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

observation $\boldsymbol{y}_k(t) \in \mathbb{R}^{n_k \times 1}$ received at node k at time t is given by:

$$\boldsymbol{y}_k(t) = \boldsymbol{H}_k \boldsymbol{x}(t) + \underbrace{\boldsymbol{v}_k(t)}_{\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_k)}, \tag{2}$$

where $\boldsymbol{H}_k \in \mathbb{R}^{n_k \times q}$ is an observation matrix, and $\boldsymbol{v}_k(t)$ is a zero-mean Gaussian observation noise with covariance matrix $\boldsymbol{R}_k$, which is independent across sensors and i.i.d. across t. The pair $(\boldsymbol{A}, \boldsymbol{Q}^{\frac{1}{2}})$ is assumed to be stabilizable, and the pair $(\boldsymbol{A}, \boldsymbol{H}_k)$ is assumed to be observable for each $1 \leq k \leq N$.

At time t, each agent node $k \in \mathcal{N}$ declares an estimate $\hat{\boldsymbol{x}}^{(k)}(t)$ using Kalman consensus filtering (KCF, see [5]) which involves the following sequence of steps:

1) Each node $k \in \mathcal{N}$ computes an intermediate estimate $\bar{\boldsymbol{x}}^{(k)}(t) = \boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)$.
2) Each node $k \in \mathcal{N}$ broadcasts $\bar{\boldsymbol{x}}^{(k)}(t)$ to all $j \in \mathcal{N}_k$.
3) Each node $k \in \mathcal{N}$ computes its final estimate of the process as:

$$\begin{aligned} \hat{\boldsymbol{x}}^{(k)}(t) &= \bar{\boldsymbol{x}}^{(k)}(t) + \boldsymbol{G}_k(\boldsymbol{y}_k(t) - \boldsymbol{H}_k \bar{\boldsymbol{x}}^{(k)}(t)) \\ &\quad + \boldsymbol{C}_k \sum_{j \in \mathcal{N}_k} (\bar{\boldsymbol{x}}^{(j)}(t) - \bar{\boldsymbol{x}}^{(k)}(t)) \end{aligned} \tag{3}$$

Here $\boldsymbol{G}_k$ and $\boldsymbol{C}_k$ are the Kalman and consensus gain matrices used by node k, respectively.

B. The χ^2 detector

Let us define the innovation vector at node k by $\boldsymbol{z}_k(t) := \boldsymbol{y}_k(t) - \boldsymbol{H}_k \boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)$. Let us assume that, under no attack, $\{\boldsymbol{z}_k(t)\}_{t \geq 0}$ reaches its steady-state distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma}_k)$. Under a possible attack, a standard technique (see [6], [17]) to detect any anomaly in $\{\boldsymbol{z}_t\}_{t \geq 0}$ is the χ^2 detector, which tests whether the innovation vector follows the desired Gaussian distribution. The detector *at each agent node* observes the innovation sequence over a pre-specified window of J time-slots, and declares an attack at time τ if $\sum_{t=\tau-J+1}^{\tau} \boldsymbol{z}_k(t)' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{z}_k(t) \geq \eta$, where η is a threshold which can be adjusted to control the false alarm probability. The covariance matrix $\boldsymbol{\Sigma}_k$ can be computed from standard results on KCF as in [5].

C. False data injection (FDI) attack

At time t, sensors associated to any subset of nodes $\mathcal{A}_t \subset \mathcal{N}$ can be under attack. A node $k \in \mathcal{A}_t$ receives an observation:

$$\begin{aligned} \tilde{\boldsymbol{y}}_k(t) &= \boldsymbol{y}_k(t) + \boldsymbol{e}_k(t) \\ &= \boldsymbol{H}_k \boldsymbol{x}(t) + \boldsymbol{e}_k(t) + \boldsymbol{v}_k(t), \end{aligned} \tag{4}$$

where $\boldsymbol{e}_k(t)$ is the error injected by the attacker. The attacker seeks to insert the error sequence $\{\boldsymbol{e}_k(t) : k \in \mathcal{A}_t\}_{t \geq 0}$ in order to introduce error in the estimation. If $\mathcal{A}_t = \mathcal{A}$ for all t, then the attack is called a *static attack*, otherwise the attack is called a *switching location attack*. *We will consider only static attack in this paper, though the theory developed in this paper can be extended to switching location attack.* We assume that the attacker can observe $\hat{\boldsymbol{x}}^{(k)}(t)$ for all $1 \leq k \leq N$ once they are computed by the agent nodes. We also assume that the attacker knows the matrices $\boldsymbol{A}, \boldsymbol{Q}, \{\boldsymbol{H}_k\}_{1 \leq k \leq N}, \{\boldsymbol{R}_k\}_{1 \leq k \leq N}$.

D. The optimization problem

The attacker seeks to steer the estimate at each agent node as close as possible to some pre-defined value $\boldsymbol{x}^*$, while keeping the attack detection probability per unit time across all nodes under some constraint value α. The authors of [6] proposed a linear injection attack to fool the χ^2 detector in a centralized, remote estimation setting. Motivated by [6], we also propose a linear attack, where, at time t, the sensor(s) associated with any node $k \in \mathcal{A}$ modifies the innovation vector as $\tilde{\boldsymbol{z}}_{\boldsymbol{k}}(t) = \boldsymbol{T}_k \boldsymbol{z}_k(t) + \boldsymbol{b}_k(t)$, where $\boldsymbol{T}_k$ is a square matrix and $\boldsymbol{b}_k(t) \sim N(\boldsymbol{\mu}_k(\boldsymbol{\theta}^{(k)}(t-1)), \boldsymbol{S}_k)$ is independent Gaussian with its mean taken as a function of $\boldsymbol{\theta}^{(k)}(t-1) \doteq \hat{\boldsymbol{x}}^{(k)}(t-1) - \boldsymbol{x}^*$. The bias term $\boldsymbol{\mu}_k(\boldsymbol{\theta}^{(k)}(t-1))$ is assumed to take a linear form $\boldsymbol{\mu}_k(\boldsymbol{\theta}^{(k)}(t-1)) = \boldsymbol{M}_k \boldsymbol{\theta}^{(k)}(t-1) + \boldsymbol{d}_k$ for suitable matrix and vector $\boldsymbol{M}_k$ and $\boldsymbol{d}_k$. This is equivalent to modifying the observation vector to $\tilde{\boldsymbol{y}}_{\boldsymbol{k}}(t)$. If $\{\boldsymbol{T}_k, \boldsymbol{S}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1 \leq k \leq N}$ is constant over time t, the attack is called stationary, else non-stationary.

1) Upper bound on the attack detection probability: The probability of attack detection per unit time slot under the χ^2 detector can be upper bounded as:

$$\begin{aligned} P_d &= \limsup_{T \to \infty} \frac{1}{T+1} \sum_{\tau=0}^{T} \mathbb{P}\Big(\cup_{k=1}^{N} \{ \sum_{t=\tau-J+1}^{\tau} \tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t) \geq \eta \} \Big) \\ &\leq \limsup_{T \to \infty} \frac{1}{T+1} \sum_{\tau=0}^{T} \sum_{k=1}^{N} \mathbb{P}(\sum_{t=\tau-J+1}^{\tau} \tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t) \geq \eta) \\ &\leq \limsup_{T \to \infty} \frac{1}{T+1} \sum_{\tau=0}^{T} \sum_{k=1}^{N} \frac{\mathbb{E}(\sum_{t=\tau-J+1}^{\tau} \tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t))}{\eta} \\ &= \frac{J}{\eta} \limsup_{T \to \infty} \frac{1}{T+1} \sum_{\tau=0}^{T} \sum_{k=1}^{N} \mathbb{E}(\tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t)) \end{aligned} \tag{5}$$

Here the two inequalities come from the union bound and the Markov inequality, respectively. Obviously, $\limsup_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=1}^{N} \mathbb{E}(\tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t)) \leq \frac{\alpha\eta}{J}$ will ensure $P_d \leq \alpha$.

2) Using the upper bound to formulate the optimization problem: Motivated by the upper bound (5), the attacker seeks to solve the following constrained optimization problem:

$$\begin{aligned} \min_{\{\boldsymbol{T}_k, \boldsymbol{S}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{k=1}^{N}} \quad & \limsup_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=1}^{N} \mathbb{E} ||\hat{\boldsymbol{x}}^{(k)}(t) - \boldsymbol{x}^*||^2 \\ \text{s.t.} \quad & \limsup_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=1}^{N} \mathbb{E}(\tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t)) \leq \frac{\alpha\eta}{J} \end{aligned} \tag{CP}$$

This problem can be relaxed by a Lagrange multiplier λ to obtain the following unconstrained optimization problem:

$$\begin{aligned} \min_{\{\boldsymbol{T}_k, \boldsymbol{S}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{k=1}^{N}} \quad & \limsup_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=1}^{N} \mathbb{E}(||\hat{\boldsymbol{x}}^{(k)}(t) - \boldsymbol{x}^*||^2 \\ & + \lambda \tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t)) \end{aligned} \tag{UP}$$

The following standard result tells us how to choose λ.

Proposition 1. *Let us consider* (CP) *and its relaxed version* (UP). *If there exists a* $\lambda^* \geq 0$ *and matrices* $\{\boldsymbol{T}_k^*, \boldsymbol{S}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{k=1}^{N}$ *such that (i)* $\{\boldsymbol{T}_k^*, \boldsymbol{S}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{k=1}^{N}$ *is*

the optimal solution of (UP) *under* $\lambda = \lambda^*$*, and (ii) the tuple* $\{\boldsymbol{T}_k^*, \boldsymbol{S}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{k=1}^N$ *satisfies the constraint in* (CP) *with equality, then* $\{\boldsymbol{T}_k^*, \boldsymbol{S}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{k=1}^N$ *is an optimal solution for* (CP) *as well.*

Proposition 1 says that, if we choose an appropriate value for λ^* and solve (UP), we will obtain an optimal solution to (CP). In this section, we provide an on-line learning algorithm to find $(\{\boldsymbol{T}_k^*, \boldsymbol{S}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{k=1}^N, \lambda^*)$. However, we will first analytically characterize the dynamics of the deviation $(\hat{\boldsymbol{x}}^{(k)}(t) - \boldsymbol{x}^*)$ in presence of linear attack, which will be used in developing the attack design algorithm later.

III. Error dynamics under attack

In this section, we will derive iterative updates for the mean squared deviation of the estimates from the target $\boldsymbol{x}^*$, which will be used in formulating the optimal attack design problem in Section IV.

Let us consider an algorithm that maintains iterates $\{\boldsymbol{T}_k(t), \boldsymbol{U}_k(t), \boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\le k\le N}$ and $\lambda(t)$ for $\{\boldsymbol{T}_k, \boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}$ and λ, where $\boldsymbol{U}_k'\boldsymbol{U}_k \doteq \boldsymbol{S}_k$. Since it is difficult to maintain $\boldsymbol{S}_k(t)$ positive definite in an iterative algorithm, we choose to iteratively update $\boldsymbol{U}_k(t)$ and set $\boldsymbol{S}_k(t) = \boldsymbol{U}_k'(t)\boldsymbol{U}_k(t)$.

Let us define the sigma algebra:

$$\begin{aligned}\mathcal{F}_\tau \doteq\ & \sigma(\{\hat{\boldsymbol{x}}^{(k)}(t), \boldsymbol{y}_k(t), \boldsymbol{T}_k(t), \boldsymbol{U}_k(t), \boldsymbol{M}_k(t), \boldsymbol{d}_k(t),\\ & \boldsymbol{b}_k(t), \lambda(t)\}_{1\le k\le N}, \lambda(t) : 1 \le t \le \tau) \end{aligned} \tag{6}$$

This is the information available to the attacker at time $(\tau+1)$ before a new attack. However, let us assume for the sake of analysis that the attacker uses constant $\boldsymbol{T}_k, \boldsymbol{M}_k, \boldsymbol{d}_k, \boldsymbol{U}_k$ respectively, for all $k \in \{1, 2, \cdots, N\}$.

Let $\tilde{\boldsymbol{\phi}}(t) \doteq (\hat{\boldsymbol{x}}(t) - \boldsymbol{x}(t))$, where $\hat{\boldsymbol{x}}(t) \doteq \mathbb{E}(\boldsymbol{x}(t)|\{\boldsymbol{y}_k(\tau)\}_{1\le k\le N, \tau\le t}) = \mathbb{E}(\boldsymbol{x}(t)|\mathcal{F}_t)$ is the MMSE estimate of $\boldsymbol{x}(t)$ under no attack and can be computed by the attacker using a standard Kalman filter. Clearly, $\tilde{\boldsymbol{\phi}}(t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}(t))$ where $\boldsymbol{R}(t)$ can be computed by a standard Kalman filter. Hence, given $\mathcal{F}_t$, $\boldsymbol{x}(t) \sim \mathcal{N}(\hat{\boldsymbol{x}}(t), \boldsymbol{R}(t))$. Also, conditioned on $\mathcal{F}_t$, the distribution of $\boldsymbol{\phi}(t) \doteq (\boldsymbol{x}(t) - \boldsymbol{x}^*)$ is $\mathcal{N}(\hat{\boldsymbol{x}}(t) - \boldsymbol{x}^*, \boldsymbol{R}(t))$. Note that, these quantities can be computed by the attacker via a standard Kalman filter.

Let us also recall that $\boldsymbol{\theta}^{(k)}(t) \doteq \hat{\boldsymbol{x}}^{(k)}(t) - \boldsymbol{x}^*$.

Theorem 1 (Error dynamics)**.** *Under a constant* $\{\boldsymbol{T}_k, \boldsymbol{M}_k, \boldsymbol{d}_k, \boldsymbol{U}_k\}_{1\le k\le N}$*, the quantity* $\mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1})$ *can be expressed as* (7) *and* $\mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1})$ *can be expresed by* (8)*.*

Proof: See Appendix A. ∎

Remark 1. Note that, given $\{\boldsymbol{\theta}^{(k)}(t-1) : 1 \le k \le N\}$, the function $\sum_{k=1}^N \mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1})$ and $\sum_{k=1}^N \mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1})$ are quadratic in $\{\boldsymbol{T}_k, \boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}$. Hence, the function

$$\begin{aligned} & f_t(\{\boldsymbol{T}_k, \boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}, \lambda)\\ \doteq\ & \sum_{k=1}^N \mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2 + \lambda\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1}) \end{aligned} \tag{9}$$

is also quadratic in $\{\boldsymbol{T}_k, \boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}$. In case of non-stationary attack, these results will hold w.r.t. $\{\boldsymbol{T}_k(t), \boldsymbol{U}_k(t), \boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\le k\le N}$. Hence, Theorem 1 will allow us to foumulate quadratically constrained quadratic problems (QCQP) for attack design.

Lemma 1. *The function* $\mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1})$ *is convex in* $\{\boldsymbol{T}_k, \boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}$*. For fixed* $\{\boldsymbol{T}_k\}_{1\le k\le N}$*, the functions* $\mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1})$ *and* $f_t(\{\boldsymbol{T}_k, \boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}, \lambda)$ *are convex in* $\{\boldsymbol{U}_k, \boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\le k\le N}$*.*

Proof: See Appendix B. ∎

Remark 2. Lemma 1 will allow us in the next section to formulate the attack design problem as a convex optimization problem.

1) Stability of $\{\boldsymbol{\theta}^{(k)}(t)\}$*:* Let us consider constant $\{\boldsymbol{T}_k(t), \boldsymbol{U}_k(t), \boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\le k\le N}$ over time. Let us define the matrix $\boldsymbol{M}$ consisting of N^2 blocks (each block is a square matrix) where:

- The (k,k)-th block in $\boldsymbol{M}$ is $(\boldsymbol{A} - \boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A} - N_k\boldsymbol{C}_k\boldsymbol{A})$.
- For $k \ne j$ and $j \in \mathcal{N}_k$, the (k,j)-th block of $\boldsymbol{M}$ is $\boldsymbol{C}_k\boldsymbol{A}$.
- For $k \ne j$ and $j \notin \mathcal{N}_k$, the (k,j)-th block of $\boldsymbol{M}$ is $\boldsymbol{0}$.

Lemma 2. *The error dynamics* $\{\boldsymbol{\theta}^{(k)}(t)\}_{1\le k\le N}$ *is stable if the spectral radius of* $\boldsymbol{M}$ *is less than* 1*.*

Proof: See Appendix C. ∎

Remark 3. Clearly, if we choose $\boldsymbol{T}_k = \boldsymbol{I}$ for $1 \le k \le N$, then the $\{\boldsymbol{\theta}^{(k)}(t) : 1 \le k \le N\}_{t\ge 0}$ process remains stable if the estimates at various nodes are stable under no attack.

Lemma 3. *If the spectral radius of* $\boldsymbol{M}$ *is less than* 1*, then the* $\{\tilde{z}_k(t)\}_{t\ge 0}$ *process is also stable for all* $1 \le k \le N$*.*

Proof: We know that $\tilde{\boldsymbol{z}}(t) = \boldsymbol{T}_k(\boldsymbol{y}_k(t) - \boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)) + \boldsymbol{b}_k(t)$. Since the true observation sequence $\{\boldsymbol{y}_k(t)\}_{t\ge 0}$ is stable, $\{\boldsymbol{b}_k(t)\}_{t\ge 0}$ is i.i.d., and $\{\hat{\boldsymbol{x}}^{(k)}(t)\}_{t\ge 0}$ is stable under FDI (by Lemma 2), the proof follows. ∎

IV. Attack design via direct optimization

In Section III, we derived closed form update equations for the error dynamics. In this section, we will use those update equations to formulate the attack design problem as an optimization problem, and prove its convexity under some special cases. Next, we will apply the well-known Karush-Kuhn-Tucker (KKT) conditions to find $\{\boldsymbol{T}_k^*, \boldsymbol{U}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{1\le k\le N}$ for designing the attack at time t, update the Lagrange multiplier λ iteratively at a slower timescale using stochastic approximation to meet the constraint with equality, and then prove convergence of the proposed algorithms to the set of optimal solutions under convexity.

A. KKT based solution: the LAADE-KKT algorithm

Let us consider the modified constrained problem:

$$\begin{aligned} \min_{\{\boldsymbol{T}_k^*, \boldsymbol{U}_k^*, \boldsymbol{M}_k^*, \boldsymbol{d}_k^*\}_{1\le k\le N}} & \sum_{k=1}^N \mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1})\\ s.t. & \sum_{k=1}^N \mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1}) \le \frac{\alpha\eta}{J} \end{aligned} \tag{MCP1}$$

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

$$\begin{aligned}\mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1}) &= ||(\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}-N_k\boldsymbol{C}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}\boldsymbol{\theta}^{(j)}(t-1)-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*+\boldsymbol{G}_k(\boldsymbol{M}_k\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k)||^2\\ &+\mathrm{Tr}(\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{Q}\boldsymbol{H}_k'\boldsymbol{T}_k'\boldsymbol{G}_k'+\boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{G}_k'+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{R}_k\boldsymbol{T}_k'\boldsymbol{G}_k')\\ &+2\Big((\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}-N_k\boldsymbol{C}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}\boldsymbol{\theta}^{(j)}(t-1)-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*+\boldsymbol{G}_k(\boldsymbol{M}_k\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k))\Big)'\\ &\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\underbrace{\mathbb{E}(\boldsymbol{\phi}(t-1)|\mathcal{F}_{t-1})}_{=\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*}+\underbrace{\mathbb{E}(||\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{\phi}(t-1)||^2|\mathcal{F}_{t-1})}_{=\mathrm{Tr}\left(\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\left(\boldsymbol{R}(t-1)+(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*)(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*)'\right)\boldsymbol{A}'\boldsymbol{H}_k'\boldsymbol{T}_k'\boldsymbol{G}_k'\right)}\end{aligned}\tag{7}$$

$$\begin{aligned}\mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1}) &= \mathrm{Tr}\Big(\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\Big(\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{Q}\boldsymbol{H}_k'\boldsymbol{T}_k'+\boldsymbol{T}_k\boldsymbol{R}_k\boldsymbol{T}_k'+\boldsymbol{S}_k+\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{R}(t-1)\boldsymbol{A}'\boldsymbol{H}_k'\boldsymbol{T}_k'\\ &+[\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}(t-1)-\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{M}_k\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k]\\ &[\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}(t-1)-\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{M}_k\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k]'\Big)\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\Big)\end{aligned}\tag{8}$$

Clearly, applying KKT conditions on the relaxed version of this problem, using a Lagrange multiplier λ, will involve setting the gradient of $f_t(\{\boldsymbol{T}_k,\boldsymbol{U}_k,\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N},\lambda)$ w.r.t. the primal variables $\{\boldsymbol{T}_k,\boldsymbol{U}_k,\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}$ equal to $\mathbf{0}$. However, it turns out that, the function $f_t(\{\boldsymbol{T}_k,\boldsymbol{U}_k,\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N},\lambda)$ is convex (by Lemma 1) but not strictly convex w.r.t. $\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}$, and that the derivative of this function w.r.t. $\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}$ is a function of $\{\boldsymbol{M}_k\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k\}_{1\le k\le N}$, which can lead to many possible solutions. Hence, we introduce a regularization term involving the Frobenius norm of $\{\boldsymbol{M}_k\}_{1\le k\le N}$:

$$\begin{aligned}\min_{\{\boldsymbol{T}_k^*,\boldsymbol{U}_k^*,\boldsymbol{M}_k^*,\boldsymbol{d}_k^*\}_{1\le k\le N}}\ &\sum_{k=1}^N\mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1})+\xi\sum_{k=1}^N||\boldsymbol{M}_k||_F^2\\ s.t.\ &\sum_{k=1}^N\mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1})\le\frac{\alpha\eta}{J}\end{aligned}\tag{MCP}$$

Here $\xi>0$ is a pre-determined constant. Applying KKT conditions on the relaxed version of (MCP), using a Lagrange multiplier λ, will involve setting the gradient of $f_t(\{\boldsymbol{T}_k,\boldsymbol{U}_k,\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N},\lambda)+\xi\sum_{k=1}^N||\boldsymbol{M}_k||_F^2$ w.r.t. the primal variables $\{\boldsymbol{T}_k,\boldsymbol{U}_k,\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}$ equal to $\mathbf{0}$. This yields a set of linear equations (10), (11), (12), (13) of these primal variables.

Lemma 4. *The optimal solution of* (MCP) *yields* $\boldsymbol{U}_k^*=\mathbf{0}$ *and hence* $\boldsymbol{S}_k^*=\mathbf{0}$ *for all* $1\le k\le N$.

Proof: (11) directly shows that $\boldsymbol{U}_k^*=\mathbf{0}$, since $\boldsymbol{G}_k'\boldsymbol{G}_k+\lambda\boldsymbol{\Sigma}_k^{-1}$ is a positive definite matrix. ∎

Remark 4. Lemma 4 tells that $\boldsymbol{b}_k(t)$ can be chosen to be deterministic under the optimal attack.

By solving (10), (12) and (13), we can find $\{\boldsymbol{T}_k^*(\lambda),\boldsymbol{M}_k^*(\lambda),\boldsymbol{d}_k^*(\lambda)\}_{1\le k\le N}$ as a function of λ. Putting these values in the constraint of (MCP) and equating both sides yields λ; then $\{\boldsymbol{T}_k^*(\lambda),\boldsymbol{M}_k^*(\lambda),\boldsymbol{d}_k^*(\lambda)\}_{1\le k\le N}$ can be used for the attack at time t. It is important to note that, $\{\boldsymbol{T}_k^*(\lambda),\boldsymbol{M}_k^*(\lambda),\boldsymbol{d}_k^*(\lambda)\}_{1\le k\le N}$ depend on the estimates, and thus on the history of observations as well.

Note that, (MCP) is a quadratically constrained quadratic problem (QCQP), which is not necessarily convex. Hence, KKT conditions may not yield the globally optimal solution. However, for the special case where $\{\boldsymbol{T}_k\}_{1\le k\le N}$ is fixed, (MCP) becomes a convex optimization problem by Lemma 1, and hence the above KKT-based procedure yields globally optimally solution. This algorithm is called *linear attack algorithm for distributed estimation based on KKT (LAADE-KKT)*.

B. Updating $\lambda(t)$ iteratively: OLAADE-KKT

Note that, solving (CP) will require us to solve a constrained average-cost Markov decision process (MDP; see [43]) to find an optimal policy, since the decision obtained by solving (MCP) at any time will affect the future estimates made at the nodes, and thus the future cost incurred by the attacker as well. Obviously, solving (MCP) will always return a *myopic policy*. However, due to the complicated structure of the problem, especially due to the complex process of evolution of the single stage objective function and constraint function in (CP) over time, we resorted to solve (MCP) as an alternative to solving MDP. However, (MCP) is a one-shot optimization problem where the objective and constraint both are some conditional expectations given the history $\mathcal{F}_{t-1}$, while (CP) is a sequential optimization problem where the objective and constraint are averaged over independent sample paths.

1) OLAADE-KKT-1: Here we will provide an online version of LAADE-KKT, i.e., OLAADE-KKT-1, which will seek to meet the constraint of (CP). This algorithm maintains a running iterate $\lambda(t-1)$, and computes $\boldsymbol{T}_k(t-1)=\boldsymbol{T}_k^*(\lambda(t-1)),\boldsymbol{M}_k(t-1)=\boldsymbol{M}_k^*(\lambda(t-1)),\boldsymbol{d}_k(t-1)=\boldsymbol{d}_k^*(\lambda(t-1))$ to solve (UP) at time t by using the set of linear equations (10), (11), (12), (13). Then it makes the following update:

$$\lambda(t)=[\lambda(t-1)+b(t)(\sum_{k=1}^N\tilde{\tilde{\boldsymbol{z}}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(t)-\frac{\alpha\eta}{J})]_0^{A_0}\tag{14}$$

where $\tilde{\tilde{\boldsymbol{z}}}_k(t)$ is the innovation at node k at time t, which is obtained by applying $\{\boldsymbol{T}_k(t-1)=\boldsymbol{T}_k^*(\lambda(t-1)),\boldsymbol{M}_k(t-1)=\boldsymbol{M}_k^*(\lambda(t-1)),\boldsymbol{d}_k(t-1)=\boldsymbol{d}_k^*(\lambda(t-1))\}_{1\le k\le N}$ on an independently generated/simulated state-observation sequence $\{\tilde{\tilde{\boldsymbol{x}}}(\tau),\tilde{\tilde{\boldsymbol{y}}}(\tau)\}_{0\le\tau\le t}$. Step size sequence $\{b(t)\}_{t\ge0}$ is a sequence of non-negative numbers such that $\sum_{t=0}^\infty b(t)=$

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

Differentiation w.r.t. $\boldsymbol{T}_k$:

$$\begin{aligned}
&\boldsymbol{G}_k'\boldsymbol{G}_k\boldsymbol{T}_k^*\Big[\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{\theta}^{(k)}(t-1)\Big(\boldsymbol{\theta}^{(k)}(t-1)\Big)'\boldsymbol{A}'\boldsymbol{H}_k'+\boldsymbol{H}_k\boldsymbol{Q}\boldsymbol{H}_k'+\boldsymbol{R}_k+\boldsymbol{H}_k\boldsymbol{A}\Big(\boldsymbol{R}(t-1)+(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*)(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*)'\Big)\boldsymbol{A}'\boldsymbol{H}_k'\\
&-\Big(\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{\theta}^{(k)}(t-1)(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*)'\boldsymbol{A}'\boldsymbol{H}_k'+\boldsymbol{H}_k\boldsymbol{A}(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*)(\boldsymbol{\theta}^{(k)}(t-1))'\boldsymbol{A}'\boldsymbol{H}_k'\Big)\Big]\\
&-\boldsymbol{G}_k'\Big[(\boldsymbol{A}-N_k\boldsymbol{C}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}\boldsymbol{\theta}^{(j)}(t-1)-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*+\boldsymbol{G}_k(\boldsymbol{M}_k^*\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k^*)\Big](\boldsymbol{\theta}^{(k)}(t-1))'\boldsymbol{A}'\boldsymbol{H}_k'\\
&+\lambda\boldsymbol{\Sigma}_k^{-1}\boldsymbol{T}_k^*\Big[\boldsymbol{H}_k\boldsymbol{Q}\boldsymbol{H}_k'+\boldsymbol{R}_k+\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{R}(t-1)\boldsymbol{H}_k'\boldsymbol{A}'+\boldsymbol{H}_k\boldsymbol{A}\Big(\hat{\boldsymbol{x}}(t-1)-\hat{\boldsymbol{x}}^{(k)}(t-1)\Big)\Big(\hat{\boldsymbol{x}}(t-1)-\hat{\boldsymbol{x}}^{(k)}(t-1)\Big)'\boldsymbol{H}_k'\boldsymbol{A}'\Big]\\
&+\lambda\boldsymbol{\Sigma}_k^{-1}\Big(\boldsymbol{M}_k^*\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k^*\Big)\Big(\hat{\boldsymbol{x}}(t-1)-\hat{\boldsymbol{x}}^{(k)}(t-1)\Big)\boldsymbol{A}'\boldsymbol{H}_k'=0
\end{aligned}\tag{10}$$

Differentiation w.r.t. $\boldsymbol{U}_k$:

$$\Big(\boldsymbol{G}_k'\boldsymbol{G}_k+\lambda\boldsymbol{\Sigma}_k^{-1}\Big)\boldsymbol{U}_k=0\tag{11}$$

Differentiation w.r.t. $\boldsymbol{M}_k$:

$$\begin{aligned}
&\boldsymbol{G}_k'\boldsymbol{G}_k\boldsymbol{M}_k^*\boldsymbol{\theta}^{(k)}(t-1)(\boldsymbol{\theta}^{(k)}(t-1))'\\
&+2\boldsymbol{G}_k'\Big((\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}-N_k\boldsymbol{C}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}\boldsymbol{\theta}^{(j)}(t-1)-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*+\boldsymbol{G}_k\boldsymbol{d}_k^*\Big)\Big(\boldsymbol{\theta}^{(k)}(t-1)\Big)'\\
&+\boldsymbol{G}_k'\boldsymbol{G}_k\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}\Big(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*\Big)\Big(\boldsymbol{\theta}^{(k)}(t-1)\Big)'+2\lambda\boldsymbol{\Sigma}_k^{-1}\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}\Big(\hat{\boldsymbol{x}}(t-1)-\hat{\boldsymbol{x}}^{(k)}(t-1)\Big)\Big(\boldsymbol{\theta}^{(k)}(t-1)\Big)'\\
&+2\lambda\boldsymbol{\Sigma}_k^{-1}\boldsymbol{M}_k^*\boldsymbol{\theta}^{(k)}(t-1)\Big(\boldsymbol{\theta}^{(k)}(t-1)\Big)'+2\xi\boldsymbol{M}_k=0
\end{aligned}\tag{12}$$

Differentiation w.r.t. $\boldsymbol{d}_k$:

$$\begin{aligned}
&\boldsymbol{G}_k'\Big((\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}-N_k\boldsymbol{C}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}\boldsymbol{\theta}^{(j)}(t-1)-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*+\boldsymbol{G}_k(\boldsymbol{M}_k^*\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k^*)\Big)\\
&+\boldsymbol{G}_k'\boldsymbol{G}_k\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}\Big(\hat{\boldsymbol{x}}(t-1)-\boldsymbol{x}^*\Big)+\lambda\boldsymbol{\Sigma}_k^{-1}\Big(\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}(t-1)-\boldsymbol{T}_k^*\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{M}_k^*\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k^*\Big)=0
\end{aligned}\tag{13}$$

$\infty, \sum_{t=0}^{\infty} b^2(t) < \infty$. The iterations are projected onto a compact interval $[0, A_0]$ to ensure boundedness. The number A_0 is chosen to be sufficiently large so that, if, for any $\lambda^* \geq 0$, the constraint in (MCP) is met with equality under $\{\boldsymbol{T}_k^*(\lambda^*), \boldsymbol{M}_k^*(\lambda^*), \boldsymbol{d}_k^*(\lambda^*)\}_{1\leq k\leq N}$, then $\lambda^* \in [0, A_0)$. This iteration is motivated by the theory of stochastic approximation [7], where the goal is to meet the constraint in (CP) with equality. This algorithm is referred to as OLAADE-KKT-1.

2) OLAADE-KKT-2: However, the constraint in (CP) actually involves an upper bound to the attack detection probability averaged over time. If the attacker has access to the alarms raised by the detectors deployed in various nodes, then that additional information can be used to update $\lambda(t)$. Let the indicator that at least one alarm is raised at time t be denoted by I_t', *which is obtained by applying* $\{\boldsymbol{T}_k(t-1) = \boldsymbol{T}_k^*(\lambda(t-1)), \boldsymbol{M}_k(t-1) = \boldsymbol{M}_k^*(\lambda(t-1)), \boldsymbol{d}_k(t-1) = \boldsymbol{d}_k^*(\lambda(t-1))\}_{1\leq k\leq N}$ *on an independently generated/simulated state-observation sequence* $\{\tilde{\tilde{\boldsymbol{x}}}(\tau), \tilde{\tilde{\boldsymbol{y}}}(\tau)\}_{0\leq\tau\leq t}$. Then, $\lambda(t)$ can be updated as:

$$\lambda(t) = [\lambda(t-1) + b(t)(I_t' - \alpha)]_0^{A_0}\tag{15}$$

Again here A_0 is chosen so large that, for any $\lambda^* \geq 0$ such that the detection probability $P_d(\lambda^*) = \alpha$, we have $\lambda^* < A_0$.

This modified algorithm is called OLAADE-KKT-2. It is interesting to note that OLAADE-KKT-2 is agnostic to the value of η used by the detectors.

3) Complexity reduction: Note that, in OLAADE-KKT-1, $\tilde{\tilde{\boldsymbol{z}}}_k(t)$ is the innovation at node k at time t, when $\{\boldsymbol{T}_k^*(\lambda(t-1)), \boldsymbol{M}_k^*(\lambda(t-1)), \boldsymbol{d}_k^*(\lambda(t-1))\}_{1\leq k\leq N}$ is applied on an independently genereted/simulated state-observation sequence $\{\tilde{\tilde{\boldsymbol{x}}}(\tau), \tilde{\tilde{\boldsymbol{y}}}(\tau)\}_{0\leq\tau\leq t}$. Using an independently generated/simulated state-observation sequence up to time t is necessary for the convergence proof of OLAADE-KKT-1, because a particular noise sequence in the convergence proof need to be Martingale difference noise sequence. Also, at each time t, we need to run this operation over the simulated history over time $\{0, 1, \cdots, t\}$ in order to ensure that an offset term in the proof remains $o(1)$ instead of $O(1)$. Hence, computing $\{\tilde{\tilde{\boldsymbol{z}}}_k(t)\}_{1\leq k\leq N}$ will require $O(t)$ computations at time t, which is not practically feasible. However, we can avoid this $O(t)$ computation by replacing $\tilde{\tilde{\boldsymbol{z}}}_k(t)$ in (14) simply by $\tilde{\boldsymbol{z}}_k(t)$ which is the innovation at node k at time t under the scheme that applies $\{\boldsymbol{T}_k^*(\lambda(\tau-1)), \boldsymbol{M}_k^*(\lambda(\tau-1)), \boldsymbol{d}_k^*(\lambda(\tau-1))\}_{1\leq k\leq N}$ on $\boldsymbol{y}(\tau)$ for all τ. This low complexity version of OLAADE-KKT-1 is denoted by OLAADE-KKT-1-LC.

Similarly, the $O(t)$ computation at time t for OLAADE-KKT-2 can be avoided by replacing I_t' in (15) by I_t which is obtained by applying $\{\boldsymbol{T}_k^*(\lambda(\tau-1)), \boldsymbol{M}_k^*(\lambda(\tau-1)), \boldsymbol{d}_k^*(\lambda(\tau-1))\}_{1\leq k\leq N}$ on $\boldsymbol{y}(\tau)$ for all τ; this low complexity version is henceforth called OLAADE-KKT-2-LC.

While the low-complexity versions are practically feasible, their convergence proof is technically very challenging. Hence, we will only prove convergence of OLAADE-KKT-1 and OLAADE-KKT-2 later in this paper.

C. Convergence analysis of OLAADE-KKT

Here, we discuss convergence properties of OLAADE-KKT-1 and OLAADE-KKT-2, where $\{\boldsymbol{T}_k\}_{1\leq k\leq N}$ are fixed and known, so that (MCP) becomes a convex optimization problem by Lemma 1.

Assumption 1. The matrices $\{\boldsymbol{T}_k\}_{1\leq k\leq N}$ are such that the $\boldsymbol{M}$ matrix of Section III has a spectral radius less than 1.

1) Convergence of OLAADE-KKT-1: Note that, if OLAADE-KKT-1 uses a fixed $\lambda \geq 0$ all the time, then at time t, the attacker takes up the history available up to time $(t-1)$, and computes $\{\boldsymbol{M}_k^*(\lambda, \{\hat{\boldsymbol{x}}^{(j)}(t-1)\}_{1\leq j\leq N}, \hat{\boldsymbol{x}}(t-1)), \boldsymbol{d}_k^*(\lambda, \{\hat{\boldsymbol{x}}^{(j)}(t-1)\}_{1\leq j\leq N}, \hat{\boldsymbol{x}}(t-1))\}_{1\leq k\leq N}$ (which are sample-path-dependent, i.e., dependent on $\{\boldsymbol{y}(\tau)\}_{0\leq\tau\leq t-1}$) which are further used to compute the estimates at time t.

Lemma 5. *For a fixed $\lambda \geq 0$ and under OLAADE-KKT-1 and Assumption 1, the distribution of the sequence of iterates $\{\boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\leq k\leq N, t\geq 0}$ reach a steady state distribution $g_\lambda^*(\cdot)$.*

Proof: By Assumption 1 and Lemma 2, $\{\hat{\boldsymbol{x}}_k(t)\}_{t\geq 0}$ and $\{\hat{\boldsymbol{x}}(t)\}_{t\geq 0}$ are stable. Hence, from (12) and (13), the lemma is proved. ∎

Let us define the distribution of $\{\boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\leq k\leq N, t\geq 0}$ under OLAADE-KKT-1 with a fixed λ as $g_{t,\lambda}(\cdot)$, and the distribution of $\{\boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\leq k\leq N, t\geq 0}$ under OLAADE-KKT-1 with $\lambda(t)$ update as $g_t(\cdot)$. Also, let $\mu_{\lambda,\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}}$ denote a generic decision rule or policy under OLAADE-KKT-1 with a fixed parameter set $\lambda, \{\boldsymbol{M}_k, \boldsymbol{d}_k\}_{1\leq k\leq N}$.

Let us define:

$$\begin{aligned}\Lambda \doteq \{\lambda \in [0, A_0) : \lim_{t\to\infty} \mathbb{E}_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}\sim g_\lambda^*(\cdot)} \\ \mathbb{E}_{\mu_{\lambda,\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}}}[\sum_{k=1}^{N} \tilde{\tilde{\boldsymbol{z}}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\tilde{\boldsymbol{z}}}_k(t)] = \frac{\alpha\eta}{J}\}\end{aligned}$$

Theorem 2 (First main theorem on convergence)**.** *Under Assumption 1 and OLAADE-KKT-1, the iterates $\lambda(t) \to \Lambda$ almost surely, and the limiting distributions satisfy $\lim_{t\to\infty} ||g_t(\cdot) - g_{t,\lambda(t)}(\cdot)||_{TV} = 0$ almost surely.*

Proof: See Appendix D. The proof is based on the theory of stochastic approximation in [7]. ∎

However, it is important to note that the convergence can be sample-path dependent.

2) Convergence of OLAADE-KKT-2: Let us define:

$$\begin{aligned}\Lambda' \doteq \{\lambda \in [0, A_0) : \lim_{t\to\infty} \mathbb{E}_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}\sim g_\lambda^*(\cdot)} \mathbb{E}_{\mu_{\lambda,\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}}} \\ (I_t') = \alpha\}\end{aligned}$$

Theorem 3 (Second main theorem on convergence)**.** *Under Assumption 1 and OLAADE-KKT-2, the iterates $\lambda(t) \to \Lambda'$ almost surely, and the limiting distributions satisfy $\lim_{t\to\infty} ||g_t(\cdot) - g_{t,\lambda(t)}(\cdot)||_{TV} = 0$ almost surely.*

Proof: The proof is very similar to that of Theorem 3, except that we use I_t' instead of $\sum_{k=1}^{N} \tilde{\tilde{\boldsymbol{z}}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\tilde{\boldsymbol{z}}}_k(t)$ in this proof. Hence, we omit details of the proof. ∎

V. Attack design via SPSA

In this section, we propose an *online linear attack algorithm for distributed estimation using SPSA* (OLAADE-SPSA) that allows us to avoid solving the KKT equations at each time t. The OLAADE-SPSA algorithm involves two-timescale stochastic approximation [7], which is basically a stochastic gradient descent algorithm with a noisy gradient estimate; (UP) is solved via SPSA in the faster timescale, and λ is updated in the slower timescale.

A. Description of OLAADE-SPSA

The algorithm (described in the next page) requires three positive step size sequences $\{a(t)\}_{t\geq 0}$, $\{b(t)\}_{t\geq 0}$ and $\{c(t)\}_{t\geq 0}$ satisfying the following criteria: (i) $\sum_{t=0}^{\infty} a(t) = \sum_{t=0}^{\infty} b(t) = \infty$, (ii) $\sum_{t=0}^{\infty} a^2(t) < \infty, \sum_{t=0}^{\infty} b^2(t) < \infty$, (iii) $\lim_{t\to\infty} \frac{b(t)}{a(t)} = 0$, (iv) $\lim_{t\to\infty} c(t) = 0$, and (v) $\sum_{t=0}^{\infty} \frac{a^2(t)}{c^2(t)} < \infty$. The first three conditions are standard requirements for two-timescale stochastic approximation. The fourth condition ensures that the gradient estimate is asymptotically unbiased, and the fifth condition is required for the convergence of SPSA.

B. Discussion of OLAADE-SPSA

1) If $\{\boldsymbol{T}_k\}_{1\leq k\leq N}$ is kept fixed, then the first update in step 4 of OLAADE-SPSA is not required.
2) The OLAADE-SPSA algorithm combines the online stochastic gradient descent (OSGD) algorithm [42, Chapter 3] with two-timescale stochastic approximation of [7]. The $\lambda(t)$ iterate is updated in the slower timescale to meet either the constraint in (CP) or the exact attack detection probability constraint with equality. In the faster timescale, OSGD is used for solving (UP). Since $\lim_{t\to\infty} \frac{b(t)}{a(t)} = 0$, the faster timescale iterates $\{\boldsymbol{T}_k(t), \boldsymbol{M}_k(t), \boldsymbol{d}_k(t)\}_{1\leq k\leq N}$ view the slower timescale iterate $\lambda(t)$ as quasi-static, while the $\lambda(t)$ iteration finds the faster timescale iterates as almost equilibriated; as if, the faster timescale iterates are varied in an inner loop and the slower timescale iterate is varied in an outer loop.
3) Steps $1-4$ of OLAADE-SPSA are basically using SGD, but via simultaneous perturbation stochastic approximation (SPSA; see [8]). SPSA allows us to avoid coordinate wise perturbation for gradient estimation of the function under consideration, by providing a zero-mean random perturbation to all coordinates (entries) of a vector or matrix variable simultaneously and independently. Steps $1-4$ of OLAADE-SPSA is equivalent to one iteration of SGD by using SPSA, where the time-varying function to optimize is $\sum_{k=1}^{N} \mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2 + \lambda(t-1)\tilde{\boldsymbol{z}}_k(t)' \boldsymbol{\Sigma}_k^{-1} \tilde{\boldsymbol{z}}_k(t) + \xi||\boldsymbol{M}_k||_F^2 | \mathcal{F}_{t-1})$.
4) All iterates are projected onto various large but compact intervals to ensure boundedness.

VI. Numerical results

We consider a distributed system with $N = 6$ agent nodes and consider two different network topologies: the 3-

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

The OLAADE-SPSA algorithm

Input: $\{a(t)\}_{t\geq 0}$, $\{b(t)\}_{t\geq 0}$, $\{c(t)\}_{t\geq 0}$, α, η, J, A_0.
Initialization: $\boldsymbol{T}_k(0)$, $\boldsymbol{M}_k(0)$, $\boldsymbol{d}_k(0)$ for all $k \in \mathcal{N}$, $\lambda(0), \{\hat{\boldsymbol{x}}^{(k)}(0)\}_{1\leq k\leq N}$, $\hat{\boldsymbol{x}}(0)$
For $t = 1, 2, 3, \cdots$**:**

1) For each $1 \leq k \leq N$, the attacker generates random matrices $\boldsymbol{\Delta}^{(k)}(t)$, $\boldsymbol{\Pi}^{(k)}(t)$ and $\boldsymbol{\beta}^{(k)}(t)$ having same dimensions as $\boldsymbol{T}_k(t-1)$, $\boldsymbol{M}_k(t-1)$ and $\boldsymbol{d}_k(t-1)$ respectively, whose entries are uniformly and independently chosen from the set $\{-1, 1\}$.
2) The attacker computes $\boldsymbol{T}_k^+ \doteq \boldsymbol{T}_k(t-1) + c(t)\boldsymbol{\Delta}^{(k)}(t)$, $\boldsymbol{T}_k^- \doteq \boldsymbol{T}_k(t-1) - c(t)\boldsymbol{\Delta}^{(k)}(t)$, $\boldsymbol{M}_k^+ \doteq \boldsymbol{M}_k(t-1) + c(t)\boldsymbol{\Pi}^{(k)}(t)$, $\boldsymbol{M}_k^- \doteq \boldsymbol{M}_k(t-1) - c(t)\boldsymbol{\Pi}^{(k)}(t)$, $\boldsymbol{d}_k^+ \doteq \boldsymbol{d}_k(t-1) + c(t)\boldsymbol{\beta}^{(k)}(t)$, $\boldsymbol{d}_k^- \doteq \boldsymbol{d}_k(t-1) - c(t)\boldsymbol{\beta}^{(k)}(t)$, for all $1 \leq k \leq N$.
3) The attacker computes:

$$\kappa_t^+ \doteq \sum_{j=1}^{N} \mathbb{E}\Bigg((||\boldsymbol{\theta}^{(j)}(t)||^2 + \lambda(t-1)\tilde{\boldsymbol{z}}_j(t)'\boldsymbol{\Sigma}_j^{-1}\tilde{\boldsymbol{z}}_j(t) + \xi||\boldsymbol{M}_k^+||_F^2 | \mathcal{F}_{t-1}, \{\boldsymbol{T}_k^+, \boldsymbol{M}_k^+, \boldsymbol{d}_k^+\}_{1\leq k\leq N} \Bigg)$$

using (7) and (8) under $\{\boldsymbol{T}_k^+, \boldsymbol{M}_k^+, \boldsymbol{d}_k^+\}_{1\leq k\leq N}$. The attacker computes κ_t^- in a similar way using $\{\boldsymbol{T}_k^-, \boldsymbol{M}_k^-, \boldsymbol{d}_k^-\}_{1\leq k\leq N}$.
4) The attacker updates each element (i, j) of $\boldsymbol{T}_k(t-1)$, $\boldsymbol{M}_k(t-1)$ and $\boldsymbol{d}_k(t-1)$ for all $1 \leq k \leq N$ as follows:

$$\begin{aligned}
\boldsymbol{T}_k(t)(i,j) &= \left[\boldsymbol{T}_k(t-1)(i,j) - a(t) \times \frac{(\kappa_t^+ - \kappa_t^-)}{2c(t)\boldsymbol{\Delta}_{(i,j)}^{(k)}(t)} \right]_{-A_0}^{A_0} \\
\boldsymbol{M}_k(t)(i,j) &= \left[\boldsymbol{M}_k(t-1)(i,j) - a(t) \times \frac{(\kappa_t^+ - \kappa_t^-)}{2c(t)\boldsymbol{\Pi}_{(i,j)}^{(k)}(t)} \right]_{-A_0}^{A_0} \\
\boldsymbol{d}_k(t)(i,1) &= \left[\boldsymbol{d}_k(t-1)(i,1) - a(t) \times \frac{(\kappa_t^+ - \kappa_t^-)}{2c(t)\boldsymbol{\beta}_{(i,1)}^{(k)}(t)} \right]_{-A_0}^{A_0}
\end{aligned} \tag{16}$$

5) The sensors make observations $\{\boldsymbol{y}_k(t)\}_{1\leq k\leq N}$, which are accessed by the attacker.
6) The attacker calculates $\boldsymbol{z}_k(t) = \boldsymbol{y}_k(t) - \boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)$ for all $k \in \{1, 2, \cdots, N\}$.
7) The attacker calculates $\tilde{\boldsymbol{z}}_{\boldsymbol{k}}(t) = \boldsymbol{T}_k(t)\boldsymbol{z}_k(t) + \boldsymbol{b}_k(t)$ for all $k \in \{1, 2, \cdots, N\}$, where $\boldsymbol{b}_k(t) = \boldsymbol{M}_k(t)\boldsymbol{\theta}^{(k)}(t-1) + \boldsymbol{d}_k(t)$. The observations are accordingly modified as $\tilde{\boldsymbol{y}}_k(t) = \tilde{\boldsymbol{z}}_k(t) + \boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)$ and sent to the agent nodes.
8) The attacker updates the Lagrange multiplier as follows:
If η is known to attacker: OLAADE-SPSA-1

$$\lambda(t) = [\lambda(t-1) + b(t)(\sum_{k=1}^{N} \tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t) - \frac{\alpha\eta}{J})]_0^{A_0} \tag{17}$$

If η is unknown to attacker but alarms are observable: OLAADE-SPSA-2

$$\lambda(t) = [\lambda(t-1) + b(t)(I_t - \alpha)]_0^{A_0} \tag{18}$$

9) The agent nodes compute the estimates locally, using (3) and the modified $\{\tilde{\boldsymbol{y}}_k(t)\}_{1\leq k\leq N}$. The agent nodes broadcast their estimates to their neighboring nodes.

end

regular hexagon and the line topology. The underlying process is q-dimensional, with $q = 2$, while the observations recorded at each node $\boldsymbol{y}_k(t) \in \mathbb{R}^{3\times 1}$. The system parameters $\boldsymbol{A}, \boldsymbol{Q}, \{\boldsymbol{R}_k\}_{1\leq k\leq 6}, \{\boldsymbol{H}_k\}_{1\leq k\leq 6}$ are chosen randomly and independently for the two different topologies. The KCF parameters $\{\boldsymbol{G}_k, \boldsymbol{C}_k\}_{1\leq k\leq 6}$ are computed using a technique from [5], and $\{\boldsymbol{\Sigma}_k\}_{1\leq k\leq 6}$ are computed by simulating the KCF under no attack.

For FDI attack, we set $\boldsymbol{x}^* = [2, 2]'$, $\eta = 300$, χ^2 window size $J = 10$ and $\lambda(0) = 4$ and regularization constant $\xi = 0.5$. To maintain the convexity of the problem, we fix $\boldsymbol{T}_k(t) = \boldsymbol{I}$, $\forall\ 1 \leq k \leq 6$ and $\forall\ t \geq 1$. We then allow the algorithm to run until convergence of $\lambda(t)$. The χ^2 detector raises alarms for FDI.

For the attack variants OLAADE-KKT-1 and OLAADE-SPSA-1, the adversary does not have access to the alarms. In this case, we notice that the Markov inequality based upper bound to the detection probability P_d as in (5) is too loose in practice, which in turn leads to a higher than necessary penalty in λ update equation (14). To alleviate this problem, we introduce a hyper-parameter c to be multiplied to the term $\frac{\alpha\eta}{J}$, which is tuned to get closer to the detection probability upper bound. For the KKT-2 and SPSA-2 variants, since the attacker has access to alarm triggers at the nodes, such a hyper-parameter is not required.

Motivated by the ADAM algorithm [44], we implement an adaptive step size optimization variant for $\lambda(t)$ for faster convergence. However, to be able to reasonably observe the

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

effect of changing λ on the detection probability, we update λ at a lower timescale of $0.1\times$, i.e., for each iterative update of λ, we let the underlying process be simulated for 10 iterations before the next update.

A. OLAADE-KKT

Let us recall that, for OLAADE-KKT, we seek to obtain the value of λ for optimizing the MSE from target vs detection probability trade-off. Once the $\lambda(t)$ iterate converges to λ^*, we simulate multiple sample paths under this fixed λ^*, and calculate the deviation from target, i.e., $\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{N}||\hat{\boldsymbol{x}}^{(k)}(t) - \boldsymbol{x}^*||^2$ for each sample path.

In Fig 1, we demonstrate the effectiveness of the attack along one sample path, by plotting the deviation of the state estimates from the specified target across the nodes, under attack and no attack cases. The broader simulation results for OLAAD-KKT-1-LC are summarized in Fig 2 and Fig 3. The mean performance of 10 sample runs is reported, and standard deviation values are highlighted as error bars. Similar results for OLAAD-KKT-2-LC are reported in tabular form in Appendix E. For OLAAD-KKT-1-LC, we report the results for that particular choice of hyper-parameter which allowed us to achieve the detection probability closest to the target, based on a grid-search.

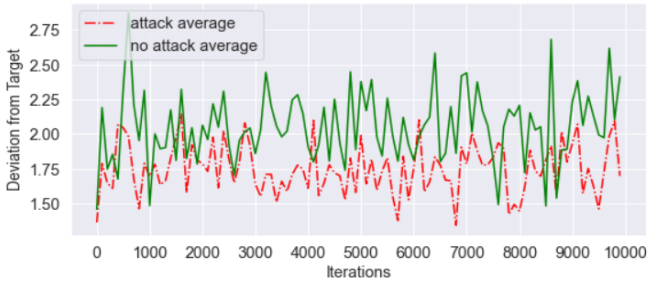


Fig. 1: OLAADE-KKT-1-LC: Average MSE from x^*, 3-regular topology, $\alpha = 0.3$

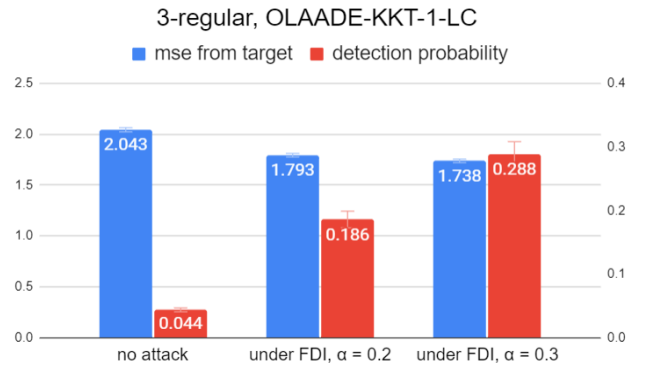


Fig. 2: $N = 6$, 3-regular topology, OLAADE-KKT-1-LC

As mentioned previously, it is important to note that the underlying process parameters were different for the two topologies. This can be seen from the fact that the detection probability under the no-attack case varies slightly for the two settings. In fact, the nature of these underlying parameters often determines how well the attack can drive the estimates to the target value, while keeping the detection rate under α.

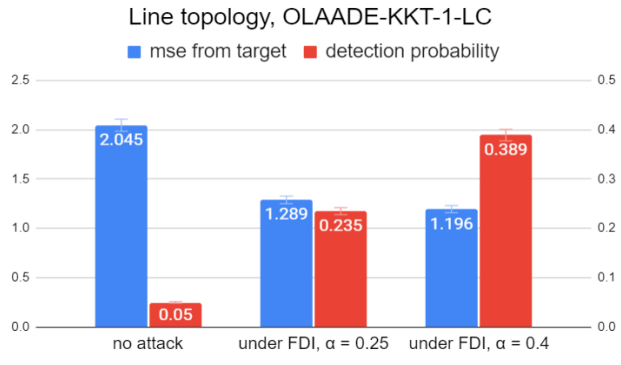


Fig. 3: $N = 6$, line topology, OLAADE-KKT-1-LC

B. OLAADE-SPSA

We repeat the same set of experiments, with the same set of attack parameters for the OLAADE-SPSA attack scheme. Note that in this case, we want to estimate the values of M, d for mounting an effective attack. As before, we report the mean performance of the attack, averaged over ten sample runs. It is again observed that OLAADE-SPSA is able to push all estimates closer to the target, while respecting the detection constraint.

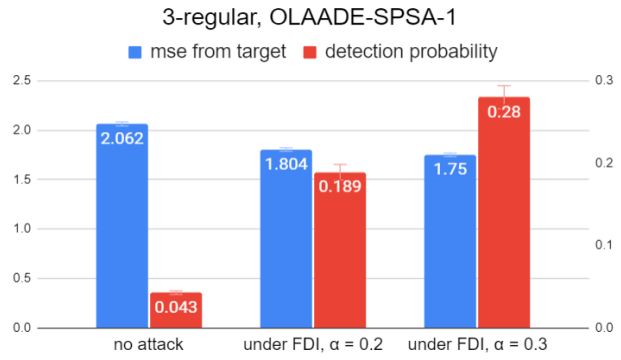


Fig. 4: $N = 6$, 3-regular topology, OLAADE-SPSA-1

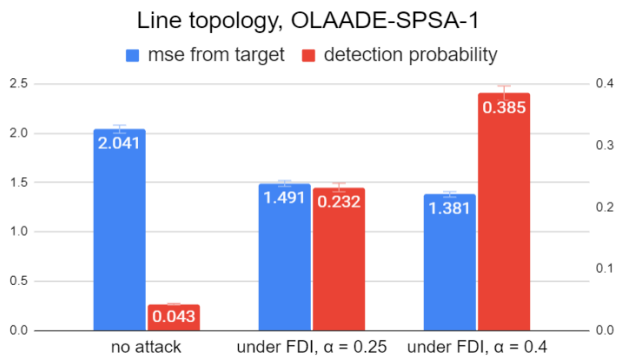


Fig. 5: $N = 6$, line topology, OLAADE-SPSA-1

C. Scalability of the algorithms for large N

It is to be noted that the faster timescale optimization in OLAADE-KKT can be decomposed into multiple sub-problems, one for each agent node. The KKT conditions (10)-(13) can be solved for each node $k \in \mathcal{N}$ separately. Hence, the computational complexity per slot encountered by the attacker in the faster timescale is $O(N)$, and the $\lambda(t)$ update in the slower timescale requires a summation of N terms. Similar complexity analysis applies to OLAADE-SPSA. However, large N may result in very slow convergence for these algorithms. In the numerical work, we consider $N = 6$ only to demonstrate the efficacy of the proposed algorithms and not the convergence rates.

D. Comparison with a naive Alternative: attack with constrained energy

While the literature lacks a competing algorithm with which we can compare the performance of our algorithms, we propose a reasonable alternative. In order to highlight the importance of the chi-squared penalization term in the cost objective, we consider an alternative attack strategy that constrains the energy budget of the injected error. More explicitly, we consider a myopic attack which, at each time t, solves the following problem:

$$\min_{\{\boldsymbol{e}_k(t)\}_{1\leq k\leq N}} \sum_{k=1}^{N} \mathbb{E}(||\hat{\boldsymbol{x}}^{(k)}(t) - \boldsymbol{x}^*||^2 + \zeta||\boldsymbol{e}_k(t)||^2|\mathcal{F}_{t-1}) \quad (19)$$

The possible advantages of such a formulation include a low-complexity closed form solution and no memory costs for parameters. The multiplier ζ can be tuned to control the energy budget as well as the attack detection probability. However, such a formulation fails to produce effective attacks for feasible values of detection probability by the detector, as demonstrated in Figure 6. This justifies the chi-squared penalty term.

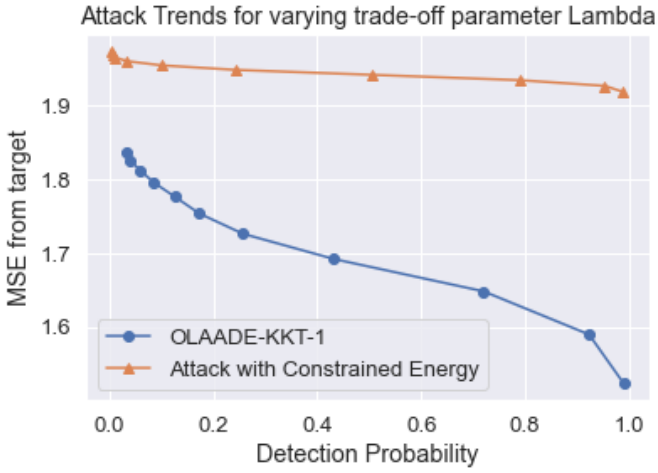


Fig. 6: Comparison of OLAADE-KKT-1 with Energy Constraint formulation for varying ζ.

E. Discussion

We highlight some key takeaways from the simulation results. Firstly, the OLAADE-KKT attack variants are always at least as good or better than their OLAADE-SPSA counterparts, depending on the underlying process parameters. This matches our intuition, since the KKT variants are indeed provably optimal for the convex formulation. However, it is important to note that the KKT algorithms require us to solve a family of matrix equations at each iteration, which requires matrix inversion; this makes the computational complexity of the KKT variants per slot higher than that of the SPSA variants.

The second observation is that, the performance of the respective variants of KKT and SPSA when the adversary does not have direct access to alarms does not alter much even if access is made available. In practice, however, this will seldom be the case, since the true values of η, J and α are not directly available to the attacker apriori, and will therefore need to be assumed. Therefore, any conservative attacker without access to alarms would tend to lower the estimate for the detection threshold in order to avoid detection, and consequently, the performance of the attack without access to alarms will be worse.

VII. Conclusions

In this paper, we designed an optimal linear attack for distributed cyber-physical systems. The problem was posed an a constrained optimization problem. The parameters of the attack scheme were learnt and optimized on-line, using tools from KKT, two-timescale stochastic approximation and SPSA. Numerical results demonstrated the efficacy of each of the proposed attack scheme. It is important to note that, OLAADE-KKT based attacks require an active adversary in the sense that, while the attack parameters converge in distribution, they have to be updated in each iteration to remain effective. On the other hand, while OLAADE-SPSA does not have that particular bottleneck, it can often require more effort to tune its parameters for convergence.

It is to be noted that we have only proved convergence of the proposed algorithms, and these results do not depend on the network topology so long as we have a connected graph. Of course, convergence rates will depend on the topology as well as the Kalman and consensus gain matrices, and characterizing the impact of the topology and gain matrices on the convergence rate is an interesting research problem for future. Also, in future, we seek to extend this work for unknown process and observation dynamics.

Appendix A
Proof of Theorem 1

Under this FDI attack, we have:

$$\begin{aligned}
&\hat{\boldsymbol{x}}^{(k)}(t) \\
=\ &\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1) + \boldsymbol{G}_k\tilde{\boldsymbol{z}}_k(t) + \boldsymbol{C}_k \sum_{j\in\mathcal{N}_k} (\bar{\boldsymbol{x}}^{(j)}(t) - \bar{\boldsymbol{x}}^{(k)}(t)) \\
=\ &\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1) + \boldsymbol{G}_k(\boldsymbol{T}_k(\boldsymbol{y}_k(t) - \boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)) + \boldsymbol{b}_k(t)) \\
+\ &\boldsymbol{C}_k\boldsymbol{A} \sum_{j\in\mathcal{N}_k} (\hat{\boldsymbol{x}}^{(j)}(t-1) - \hat{\boldsymbol{x}}^{(k)}(t-1)) \qquad (20)
\end{aligned}$$

Now,

$$\begin{aligned}
&\boldsymbol{\theta}^{(k)}(t)\\
=\;&(\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A})\hat{\boldsymbol{x}}^{(k)}(t-1)\\
&+\boldsymbol{G}_k\boldsymbol{T}_k\underbrace{\boldsymbol{y}_k(t)}_{\doteq\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{x}(t-1)+\boldsymbol{H}_k\boldsymbol{w}(t-1)+\boldsymbol{v}_k(t)}+\boldsymbol{G}_k\boldsymbol{b}_k(t)\\
&+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}(\hat{\boldsymbol{x}}^{(j)}(t-1)-\hat{\boldsymbol{x}}^{(k)}(t-1))-\boldsymbol{x}^*\\
=\;&(\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{\phi}(t-1)\\
&+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}(\boldsymbol{\theta}^{(j)}(t-1)-\boldsymbol{\theta}^{(k)}(t-1))-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*\\
&+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{w}(t-1)+\boldsymbol{G}_k\boldsymbol{b}_k(t)+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{v}_k(t)\\
=\;&(\boldsymbol{A}-\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}-N_k\boldsymbol{C}_k\boldsymbol{A})\boldsymbol{\theta}^{(k)}(t-1)\\
&+\boldsymbol{C}_k\boldsymbol{A}\sum_{j\in\mathcal{N}_k}\boldsymbol{\theta}^{(j)}(t-1)-(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}^*+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{\phi}(t-1)\\
&+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{w}(t-1)+\boldsymbol{G}_k\boldsymbol{b}_k(t)+\boldsymbol{G}_k\boldsymbol{T}_k\boldsymbol{v}_k(t) \qquad (21)
\end{aligned}$$

Clearly, $\mathbb{E}(||\boldsymbol{\theta}^{(k)}(t)||^2|\mathcal{F}_{t-1})$ can be expressed as (7); in this expression, we have used the fact that, for a column vector $\boldsymbol{a}$, $||\boldsymbol{a}||_2^2=\text{Tr}(\boldsymbol{a}\boldsymbol{a}')$ where $\boldsymbol{a}'$ is the transpose of $\boldsymbol{a}$.

On the other hand, given $\mathcal{F}_{t-1}$, $\boldsymbol{x}(t-1)\sim\mathcal{N}(\hat{\boldsymbol{x}}(t-1),\boldsymbol{R}(t-1))$ where $(\hat{\boldsymbol{x}}(t-1),\boldsymbol{R}(t-1))$ can be computed by a standard Kalman filter. Now,

$$\begin{aligned}
\tilde{\boldsymbol{z}}_k(t)&=\boldsymbol{T}_k\boldsymbol{z}_k(t)+\boldsymbol{b}_k(t)\\
&=\boldsymbol{T}_k\boldsymbol{y}_k(t)-\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{b}_k(t)\\
&=\boldsymbol{T}_k(\boldsymbol{H}_k\boldsymbol{x}(t)+\boldsymbol{v}_k(t))-\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{b}_k(t)\\
&=\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{x}(t-1)+\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{w}(t-1)+\boldsymbol{T}_k\boldsymbol{v}_k(t)\\
&\quad-\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{b}_k(t) \qquad (22)
\end{aligned}$$

which, given $\mathcal{F}_{t-1}$, is distributed as $\mathcal{N}(\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}(t-1)-\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\hat{\boldsymbol{x}}^{(k)}(t-1)+\boldsymbol{M}_k\boldsymbol{\theta}^{(k)}(t-1)+\boldsymbol{d}_k,\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{Q}\boldsymbol{H}_k'\boldsymbol{T}_k'+\boldsymbol{T}_k\boldsymbol{R}_k\boldsymbol{T}_k'+\boldsymbol{T}_k\boldsymbol{H}_k\boldsymbol{A}\boldsymbol{R}(t-1)\boldsymbol{A}'\boldsymbol{H}_k'\boldsymbol{T}_k'+\boldsymbol{S}_k)$. Hence, $\mathbb{E}(\tilde{\boldsymbol{z}}_k(t)'\boldsymbol{\Sigma}_k^{-1}\tilde{\boldsymbol{z}}_k(t)|\mathcal{F}_{t-1})$ is given by (8).

Appendix B
Proof of Lemma 1

The proof uses the fact that the function $||\sum_{i=1}^n c_iv_i+c||_2^2$ for any arbitrary real known coefficients $\{c_i\}_{1\le i\le n}$ and c and scalar variables $\{v_i\}_{1\le i\le n}$ is convex in $\{v_i\}_{1\le i\le n}$, since Hessian of this function will be $[c_1,c_2,\cdots,c_n]'[c_1,c_2,\cdots,c_n]$ which is a positive semi-definite matrix. Hence, the first term in the R.H.S. of (7) is convex in the arguments. Just as another example, let us consider another term $\text{Tr}(\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\boldsymbol{S}_k\boldsymbol{\Sigma}_k^{-\frac{1}{2}})$ from (8); this can be rewritten as $\text{Tr}(\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\boldsymbol{U}_k\boldsymbol{U}_k'\boldsymbol{\Sigma}_k^{-\frac{1}{2}})=||\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\boldsymbol{U}_k||_F^2$ which is convex in $\boldsymbol{U}_k$ since $\boldsymbol{\Sigma}_k^{-\frac{1}{2}}\boldsymbol{U}_k$ is a linear function of $\boldsymbol{U}_k$. Convexity of other terms can be proven in a similar way.

Appendix C
Proof of Lemma 2

Let us consider the evolution of $\boldsymbol{\theta}^{(k)}(t)$ in (21), and let $\boldsymbol{\theta}(t)$ be the vertical concatenation of the column vectors $\{\boldsymbol{\theta}^{(k)}(t)\}_{1\le k\le N}$. Hence, the evolution of $\boldsymbol{\theta}(t)$ is given by: $\boldsymbol{\theta}(t)=\boldsymbol{M}\boldsymbol{\theta}(t-1)+\boldsymbol{\zeta}_t$ where $\boldsymbol{\zeta}_t$ is a stable Gaussian proces since $\boldsymbol{\phi}(t)$ is a stable process. Hence, $\{\boldsymbol{\theta}(t)\}_{t\ge0}$ is a stable process if the spectral radius of $\boldsymbol{M}$ is less than 1.

Appendix D
Proof of Theorem 2

Note that, the $\{\boldsymbol{M}_k(t),\boldsymbol{d}_k(t)\}_{1\le k\le N}$ update and hence the evolution of $g_t(\cdot)$ runs in a faster timescale, while the $\lambda(t)$ update runs in a slower timescale. Also $g_{t,\lambda}(\cdot)$ and $g_\lambda^*(\cdot)$ are continuously differentiable in λ over a compact interval $[0,A_0]$, and hence are Lipschitz continuous. Clearly, by an argument similar to [7, Chapter 6, Lemma 1], we claim that $\lim_{t\to\infty}||g_t(\cdot)-g_{t,\lambda(t)}(\cdot)||_{TV}=0$ almost surely. This proves convergence in faster timescale.

Now we will prove convergence in the slower timescale. Note that, using the fact that $\lambda(t)\in[0,A_0]$ for all $t\ge0$, and using Assumption 1 and Lemma 3, we can easily say that $\{\tilde{\tilde{\boldsymbol{z}}}_k(t)\}_{1\le k\le N}$ is stable under $\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}$. Also, note that $\{\boldsymbol{x}(t),\hat{\boldsymbol{x}}^{(k)}(t),\boldsymbol{y}_k(t),\tilde{\tilde{\boldsymbol{z}}}_k(t),\boldsymbol{M}_k(t),\boldsymbol{d}_k(t)\}_{1\le k\le N}\}_{t\ge0}$ is a stable Markov chain under any $\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}$ with $\lambda(t-1)\in[0,A_0]$. Hence, the $\lambda(t)$ iteration can be written as:

$$\begin{aligned}
\lambda(t+1)&=[\lambda(t-1)+b(t)(\sum_{k=1}^N\mathbb{E}_{\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}}\\
&\quad\left(\tilde{\tilde{\boldsymbol{z}}}_k(t)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(t)\right)-\frac{\alpha J}{\eta}+\zeta_1(t))]_0^{A_0}
\end{aligned}$$

where $\zeta_1(t)\doteq\sum_{k=1}^N\tilde{\tilde{\boldsymbol{z}}}_k(t)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(t)-\sum_{k=1}^N\mathbb{E}_{\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}}\left(\tilde{\tilde{\boldsymbol{z}}}_k(t)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(t)\right)$ is a zero-mean Martingale difference noise. Now,

$$\begin{aligned}
&\sum_{k=1}^N\mathbb{E}_{\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}}\left(\tilde{\tilde{\boldsymbol{z}}}_k(t)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(t)\right)\\
=\;&\lim_{\tau\to\infty}\sum_{k=1}^N\mathbb{E}_{\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}}\\
&\left(\tilde{\tilde{\boldsymbol{z}}}_k(\tau)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(\tau)\right)+o(1)\\
=\;&\sum_{k=1}^N\mathbb{E}_{\mu_{\lambda(t-1),\{\boldsymbol{M}_k(t-1),\boldsymbol{d}_k(t-1)\}_{1\le k\le N}}}\left(\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\right)+o(1)\\
=\;&\sum_{k=1}^N\mathbb{E}_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}\sim g_{t,\lambda(t-1)}(\cdot)}\mathbb{E}_{\mu_{\lambda(t-1),\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}}}\\
&\left(\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\right)+o(1)+\zeta_2(t)\\
=\;&\sum_{k=1}^N\mathbb{E}_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}\sim g_{\lambda(t-1)}^*(\cdot)}\mathbb{E}_{\lambda(t-1),\mu_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\le k\le N}}}\\
&\left(\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\right)+o(1)+o(1)+\zeta_2(t)
\end{aligned}$$

where the first equality follows from the stability of the above Markov chain, and the second equality follows from the dominated convergence theorem. The third equality uses the fact that $X=\mathbb{E}(X)+X-\mathbb{E}(X)$, with $\zeta_2(t)$ being a Martingale difference noise. The fourth equality follows from the fact that $\lim_{t\to\infty}||g_{t,\lambda}(\cdot)-g_\lambda^*(\cdot)||_{TV}=0$ and the dominated convergence theorem.

Hence, the $\lambda(t)$ iteration can be rewritten as:

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/TIFS.2022.3146078, IEEE Transactions on Information Forensics and Security

$$\lambda(t+1) = \Bigg[\lambda(t-1) + b(t)\Big(\sum_{k=1}^{N} \mathbb{E}_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}\sim g^*_{\lambda(t-1)}(\cdot)} \mathbb{E}_{\mu_{\lambda(t-1)},\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}} \Big(\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\Big) + \zeta_1(t) + \zeta_2(t) + o(1)\Bigg]_0^{A_0}$$

Now, since $g^*_\lambda(\cdot)$ is continuous in λ, we can say that

$$\mathbb{E}_{\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{k=1}^N\sim g^*_{\lambda(t-1)}(\cdot)} \mathbb{E}_{\mu_{\lambda(t-1)},\{\boldsymbol{M}_k,\boldsymbol{d}_k\}_{1\leq k\leq N}} (\tilde{\tilde{\boldsymbol{z}}}_k(\infty)\Sigma_k^{-1}\tilde{\tilde{\boldsymbol{z}}}_k(\infty))$$

is continuously differentiable in $\lambda(t-1) \in [0, A_0]$ and hence Lipschitz continuous. Also, the offset $o(1)$ goes to 0 as $t \to \infty$. Hence, by the theory of basic stochastic approximation [7, Chapter 2], two-timescale stochastic approximation [7, Chapter 6] and projected stochastic approximation [7, Chapter 5], we can say that $\lambda(t) \to \Lambda$ almost surely.

APPENDIX E
MORE SIMULATION RESULTS

Permissible detection probability (α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.2	0.044 +/- 0.003	0.186 +/- 0.01	2.062 +/- 0.002	1.793 +/- 0.002
0.3	0.044 +/- 0.005	0.286 +/- 0.011	2.063 +/- 0.004	1.738 +/- 0.003

TABLE I: $N = 6$, 3-regular topology, OLAADE-KKT-1-LC

Permissible detection probability(α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.25	0.047 +/- 0.004	0.235 +/- 0.009	2.045 +/- 0.007	1.289 +/- 0.003
0.4	0.05 +/- 0.005	0.389 +/- 0.013	2.038 +/- 0.008	1.196 +/- 0.003

TABLE II: $N = 6$, Line topology, OLAADE-KKT-1-LC

Permissible detection probability (α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.2	0.046 +/- 0.003	0.178 +/- 0.008	2.063 +/- 0.005	1.799 +/- 0.004
0.3	0.044 +/- 0.003	0.287 +/- 0.013	2.062 +/- 0.002	1.741 +/- 0.002

TABLE III: $N = 6$, 3-regular topology, OLAADE-KKT-2-LC

Permissible detection probability(α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.25	0.05 +/- 0.004	0.223 +/- 0.01	2.048 +/- 0.012	1.305 +/- 0.005
0.4	0.049 +/- 0.003	0.355 +/- 0.016	2.046 +/- 0.009	1.211 +/- 0.004

TABLE IV: $N = 6$, Line topology, OLAADE-KKT-2-LC

Permissible detection probability (α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.2	0.043 +/- 0.005	0.189 +/- 0.012	2.062 +/- 0.003	1.804 +/- 0.003
0.3	0.044 +/- 0.005	0.28 +/- 0.013	2.062 +/- 0.002	1.75 +/- 0.002

TABLE V: $N = 6$, 3-regular topology, OLAADE-SPSA-1

Permissible detection probability(α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.25	0.052 +/- 0.006	0.232 +/- 0.014	2.042 +/- 0.014	1.491 +/- 0.010
0.4	0.049 +/- 0.006	0.385 +/- 0.008	2.041 +/- 0.009	1.381 +/- 0.005

TABLE VI: $N = 6$, Line topology, OLAADE-SPSA-1

Permissible detection probability (α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.2	0.043 +/- 0.004	0.184 +/- 0.01	2.064 +/- 0.003	1.805 +/- 0.002
0.3	0.045 +/- 0.006	0.292 +/- 0.013	2.061 +/- 0.004	1.746 +/- 0.004

TABLE VII: $N = 6$, 3-regular topology, OLAADE-SPSA-2

Permissible detection probability(α)	Detection probability (no attack)	Detection probability under FDI	Deviation from $\boldsymbol{x}^*$ (no attack)	Deviation from $\boldsymbol{x}^*$ under FDI
0.25	0.049 +/- 0.006	0.234 +/- 0.011	2.04 +/- 0.009	1.426 +/- 0.005
0.4	0.054 +/- 0.007	0.385 +/- 0.012	2.042 +/- 0.012	1.323 +/- 0.007

TABLE VIII: $N = 6, q = 2$, Line topology, OLAADE-SPSA-2

REFERENCES

[1] Moulik Choraria, Arpan Chattopadhyay, Urbashi Mitra, and Erik Strom. Optimal deception attack on networked vehicular cyber physical systems. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1131–1135. IEEE, 2019.

[2] Yilin Mo and Bruno Sinopoli. Secure control against replay attacks. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 911–918. IEEE, 2009.

[3] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on scada systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, 2014.

[4] Yanpeng Guan and Xiaohua Ge. Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. *IEEE Transactions on Signal and Information Processing over Networks*, 4(1):48–59, 2018.

[5] R. Olfati-Saber. Kalman-consensus filter : Optimality, stability, and performance. In *Conference on Decision and Control*, pages 7036–7042. IEEE, 2009.

[6] Ziyang Guo, Dawei Shi, Karl Henrik Johansson, and Ling Shi. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, 4(1):4–13, 2017.

[7] Vivek S. Borkar. *Stochastic approximation: a dynamical systems viewpoint.* Cambridge University Press, 2008.

[8] J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

[9] Derui Ding, Qing-Long Han, Xiaohua Ge, and Jun Wang. Secure state estimation and control of cyber-physical systems: A survey. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):176–190, 2020.

[10] Yuan Chen, Soummya Kar, and José MF Moura. Optimal attack strategies subject to detection constraints against cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 2017.

[11] Jian Sun Guangyu Wu and Jie Chen. Optimal data injection attacks in cyber-physical systems. *IEEE Transactions on Cybernatics*, 48(12):3302–3312, 2018.

[12] Nam N Tran, Hemanshu R Pota, Quang N Tran, Xuefei Yin, and Jiankun Hu. Designing false data injection attacks penetrating ac-based bad data detection system and fdi dataset generation. *Concurrency and Computation: Practice and Experience*, page e5956, 2020.

[13] Yuan Chen, Soummya Kar, and José MF Moura. Cyber physical attacks with control objectives and detection constraints. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 1125–1130. IEEE, 2016.

[14] Lin Liu and Zhiyu Xi. False data injection attack sequence design against quantized networked control systems. In *2019 IEEE International Conference on Unmanned Systems (ICUS)*, pages 542–547. IEEE, 2019.

[15] Fabio Pasqualetti, Florian Dörfler, and Francesco Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[16] Fei Miao, Quanyan Zhu, Miroslav Pajic, and George J Pappas. Coding schemes for securing cyber-physical systems against stealthy data injection attacks. *IEEE Transactions on Control of Network Systems*, 4(1):106–117, 2017.

[17] Yuzhe Li, Ling Shi, and Tongwen Chen. Detection against linear deception attacks on multi-sensor remote state estimation. *IEEE Transactions on Control of Network Systems*, 2017.

[18] Shaunak Mishra, Yasser Shoukry, Nikhil Karamchandani, Suhas N Diggavi, and Paulo Tabuada. Secure state estimation against sensor attacks in the presence of noise. *IEEE Transactions on Control of Network Systems*, 4(1):49–59, 2017.

[19] Zhengen Zhao, Yimin Huang, Ziyang Zhen, and Yuzhe Li. Data-driven false data-injection attack design and detection in cyber-physical systems. *IEEE transactions on cybernetics*, 2020.

[20] Arman Sargolzaei, Kasra Yazdani, Alireza Abbaspour, Carl D Crane III, and Warren E Dixon. Detection and mitigation of false data injection attacks in networked control systems. *IEEE Transactions on Industrial Informatics*, 16(6):4281–4292, 2019.

[21] Jiangfan Zhang and Xiaodong Wang. Quickest detection of time-varying false data injection attacks in dynamic linear regression models. *arXiv preprint arXiv:1811.05423*, 2018.

[22] Arash Golabi, Abdelkarim Erradi, Ashraf Tantawy, and Khaled Shaban. Detecting false data injection attacks in linear parameter varying cyber-physical systems. In *2019 International Conference on Cyber Security for Emerging Technologies (CSET)*, pages 1–8. IEEE, 2019.

[23] Ziyang Guo, Dawei Shi, Daniel E Quevedo, and Ling Shi. Secure state estimation against integrity attacks: A gaussian mixture model approach. *IEEE Transactions on Signal Processing*, 67(1):194–207, 2018.

[24] Akanshu Gupta, Abhinava Sikdar, and Arpan Chattopadhyay. Quickest detection of false data injection attack in remote state estimation. *arXiv preprint arXiv:2010.15785, accepted in IEEE International Symposium on Information Theory (ISIT)*, 2021.

[25] Miroslav Pajic, Insup Lee, and George J Pappas. Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, 4(1):82–92, 2017.

[26] Arpan Chattopadhyay and Urbashi Mitra. Attack detection and secure estimation under false data injection attack in cyber-physical systems. In *Information Sciences and Systems (CISS), 2018 52nd Annual Conference on*, pages 1–6. IEEE, 2018.

[27] Arpan Chattopadhyay, Urbashi Mitra, and Erik G Ström. Secure estimation in v2x networks with injection and packet drop attacks. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–6. IEEE, 2018.

[28] Arpan Chattopadhyay and Urbashi Mitra. Security against false data injection attack in cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 2019.

[29] Chensheng Liu, Jing Wu, Chengnian Long, and Yebin Wang. Dynamic state recovery for cyber-physical systems under switching location attacks. *IEEE Transactions on Control of Network Systems*, 4(1):14–22, 2017.

[30] Kebina Manandhar, Xiaojun Cao, Fei Hu, and Yao Liu. Detection of faults and attacks including false data injection attack in smart grid using kalman filter. *IEEE transactions on control of network systems*, 1(4):370–379, 2014.

[31] Gaoqi Liang, Junhua Zhao, Fengji Luo, Steven R Weller, and Zhao Yang Dong. A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, 8(4):1630–1638, 2017.

[32] Qie Hu, Dariush Fooladivanda, Young Hwan Chang, and Claire J Tomlin. Secure state estimation and control for cyber security of the nonlinear power systems. *IEEE Transactions on Control of Network Systems*, 2017.

[33] Yorie Nakahira and Yilin Mo. Attack-resilient h2, h-infinity, and l1 state estimator. *IEEE Transactions on Automatic Control*, 2018.

[34] Hamza Fawzi, Paulo Tabuada, and Suhas Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[35] Cheng-Zong Bai, Vijay Gupta, and Fabio Pasqualetti. On kalman filtering with compromised sensors: Attack stealthiness and performance bounds. *IEEE Transactions on Automatic Control*, 62(12):6641–6648, 2017.

[36] Yanpeng Guan and Xiaohua Ge. Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. *IEEE Transactions on Signal and Information Processing over Networks*, 4(1):48–59, 2017.

[37] Bharadwaj Satchidanandan and Panganamala R Kumar. Dynamic watermarking: Active defense of networked cyber–physical systems. *Proceedings of the IEEE*, 105(2):219–240, 2016.

[38] Xiaohua Ge, Qing-Long Han, Maiying Zhong, and Xian-Ming Zhang. Distributed krein space-based attack detection over sensor networks under deception attacks. *Automatica*, 109:108557, 2019.

[39] Florian Dörfler, Fabio Pasqualetti, and Francesco Bullo. Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1486–1491. IEEE, 2011.

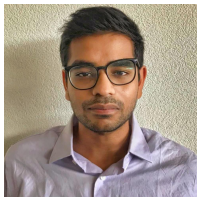
[40] Ashkan Moradi, Naveen KD Venkategowda, and Stefan Werner. Coordinated data-falsification attacks in consensus-based distributed kalman filtering. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 495–499. IEEE, 2019.

[41] An-Yang Lu and Guang-Hong Yang. Malicious attacks on state estimation against distributed control systems. *IEEE Transactions on Automatic Control*, 65(9):3911–3918, 2019.

[42] Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[43] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. II*. Athena Scientific, 2007.

[44] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, (ICLR) 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.



Moulik Choraria is a graduate student and Sarwate Fellow in the Department of Electrical and Computer Engineering, UIUC. He obtained his M.S., specializing in Communication Systems, at the École Polytechnique Fédérale de Lausanne (EPFL) in 2021 with his thesis entitled "The Inductive Bias of Polynomial Neural Networks", and received a B.Tech. degree in Electrical Engineering from the Indian Institute of Technology (IIT) Delhi in 2019. His research interests include theoretical machine learning and information theory.

Arpan Chattopadhyay obtained his B.E. in Electronics and Telecommunication from Jadavpur University, Kolkata, India in 2008, and M.E. and Ph.D in Telecommunication from Indian Institute of Science, Bangalore, India, in 2010 and 2015, respectively. He is currently working as an assistant professor in the Electrical Engineering department, IIT Delhi. Previously, he held postdoc positions in the Electrical Engineering department, University of Southern California, Los Angeles, USA, and INRIA/ENS Paris, France. His research interests include wireless communication and networks, cyber-physical systems, networked estimation and control, reinforcement learning, etc.

Urbashi Mitra (Fellow) received the B.S. and the M.S. degrees from the University of California, Berkeley, CA, USA, and the Ph.D. degree from Princeton University, Princeton, NJ, USA. She is currently the Gordon S. Marshall Professor in Engineering with the University of Southern California, Los 1461 Angeles, CA. USA, with appointments in Electrical and Computer Engineering and Computer Science. She was the Inaugural Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS. She was a Member of the IEEE Information Theory Society's Board of Governors (2002-2007, 2012-2017), the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications and Networks (2012-2016), the IEEE Signal Processing Society's awards Board (2017-2018), and the Chair/Vice-Chair of the IEEE Communication Theory Technical Committee (2017-2020). She was the recipient of the 2021 USC Viterbi School of Engineering Senior Research Award, the 2017 IEEE Women in Communications Engineering Technical Achievement Award, a 2015 U.K. Royal Academy of Engineering Distinguished Visiting Professorship, a 2015 U.S. Fulbright Scholar Award, a 2015-2016 U.K. Leverhulme Trust Visiting Professorship, IEEE Communications Society Distinguished Lecturer, 2012 Globecom Signal Processing for Communications Symposium Best Paper Award, 2012 U.S. National Academy of Engineering Lillian Gilbreth Lectureship, the 2009 DCOSS Applications Systems Best Paper Award, 2001 Okawa Foundation Award, 2000 Ohio State University's College of Engineering Lumley Award for Research, and a 1996 National Science Foundation CAREER Award. She has been an Associate Editor for the following IEEE publications: TRANSACTIONS ON SIGNAL PROCESSING, TRANSACTIONS ON INFORMATION THEORY, JOURNAL OF OCEANIC ENGINEERING, and TRANSACTIONS ON COMMUNICATIONS. She is current the Area Editor for Communications of the IEEE TRANSACTIONS ON INFORMATION THEORY. Dr. Mitra has held visiting appointments with King's College, London, U.K., Imperial College, London, U.K., the Delft University of Technology, Delft, the Netherlands, Stanford University, Stanford, CA, USA, Rice University, Houston, TX, USA, and the Eurecom Institute, Biot, France.

Erik G. Ström (S'93–M'95–SM'01–F'21) received the M.S. degree from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1990, and the Ph.D. degree from the University of Florida, Gainesville, in 1994, both in electrical engineering. He accepted a postdoctoral position at the Department of Signals, Sensors, and Systems at KTH in 1995. In February 1996, he was appointed Assistant Professor at KTH, and in June 1996 he joined Chalmers University of Technology, Göteborg, Sweden, where he is now a Professor in Communication Systems since June 2003. Dr. Ström currently heads the Division of Communications, Antennas, and Optical Networks, is the director of ChaseOn, a Vinnova Competence Center focused on antenna system, and the director of Chalmers' Area-of-Advance Information and Communication Technology. His research interests include signal processing and communication theory in general, and constellation labelings, channel estimation, synchronization, multiple access, medium access, multiuser detection, wireless positioning, and vehicular communications in particular. Since 1990, he has acted as a consultant for the Educational Group for Individual Development, Stockholm, Sweden. He is a senior editor of *IEEE Transaction on Intelligent Transport Systems*, a contributing author and associate editor for Roy. Admiralty Publishers FesGas-series, and was a co-guest editor for the *Proceedings of the IEEE* special issue on Vehicular Communications (2011) and the *IEEE Journal on Selected Areas in Communications* special issues on Signal Synchronization in Digital Transmission Systems (2001) and on Multiuser Detection for Advanced Communication Systems and Networks (2008). Dr. Ström was a member of the board of the IEEE VT/COM Swedish Chapter 2000–2006. He received the Chalmers Pedagogical Prize in 1998, the Chalmers Ph.D. Supervisor of the Year award in 2009, and the Chalmers Area of Advance Award in 2020.