Secure Estimation in V2X Networks with Injection and Packet Drop Attacks

Arpan Chattopadhyay, Urbashi Mitra, and Erik G. Ström

***Abstract*—Vehicle-to-anything (V2X) communications are essential for facilitating cooperative intelligent transport system (C-ITS) components such as traffic safety and traffic efficiency applications. Integral to proper functioning of C-ITS systems is sensing and telemetery. To this end, this paper examines how to ensure security in sensing systems for V2X networks. In particular, secure remote estimation of a Gauss-Markov process based on measurements done by a set of vehicles is considered. The measurements are collected by the individual vehicles and are communicated via wireless links to the central fusion center. The system is attacked by malicious or compromised vehicles with the goal of increasing the estimation error. The attack is achieved by two mechanisms: false data injection (FDI) and garbage packet injection. This paper extends a previously proposed adaptive filtering algorithm for tackling FDI to accommodate both FDI and garbage packet injection, by filtering out malicious observations and thus enabling secure estimates. The efficacy of the proposed filter is demonstrated numerically.**

***Index Terms*—Secure remote estimation,V2X, V2V, false data injection, packet drop attack, Kalman filter, stochastic approximation.**

I. Introduction

Vehicle-to-vehicle (V2V), vehicle-to-road-infrastructure (V2I), and vehicular-to-network (V2N), collectively gathered under the name vehicular-to-anything (V2X) communication, are enablers for so-called cooperative intelligent transport system (C-ITS) applications. C-ITS applications can be broadly categorized into traffic safety and traffic efficiency applications. V2X networks exhibit characteristics that differ from classical cellular communication networks, largely due to the potential high mobility of the vehicles. For example, V2V networks have unique reliability and latency requirements necessitating new resource allocation strategies [1], [2]. Additionally, V2V channels have unique channel characteristics that enable the design of channel estimation strategies which directly exploit these features, such as group and element-wise sparsity coupled with distinct patters of diffusive multipath in the delay-Doppler plane [3].

In this paper, we consider a C-ITS application in which a number of vehicles make noisy observations of a random process (*e.g.*, the current road conditions: temperature, friction, *etc.*) and transmit the observations to a central data fusion node, *e.g.*, an edge computing device, see Figure 1. The objective is to leverage all vehicle sensors to perform high-precision estimation of the realization of the random process. This can subsequently be used to improve safety and efficiency, *e.g.*, by controlling the inter-vehicle distances for automated vehicles and thereby ensuring safe and fuel-efficient travel.

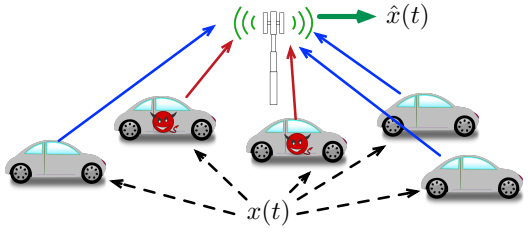

Figure 1. Illustration of the remote estimation problem.

Like any cyber-physical system (CPS), C-ITS applications are subject to attacks to achieve various malicious goals. Needless to say, such attacks can be life-threatening and cause major damage to equipment. An unfortunate drawback with V2X communication is that it opens up the possibility of an attack on the vehicle systems, *e.g.*, its sensors, from a distance [4], or an attack on the overall application. Attacks could either be denial-of-service attacks (*e.g.*, jamming attack where a jammer tries to block the wireless bandwidth, see [5], [6]), or deception or integrity attacks where the attacker attempts to modify the information sent via data packets. Examples of integrity attacks could be a replay attack (see [7], [8]) and the false data injection (FDI) attack considered in this paper. An FDI attack modifies the data sent by the sensors to the remote estimator, either by breaking into the cryptography or by external manipulation of the sensors (*e.g.*, placing a heater near a temperature sensor or by attacking the internal bus structure in the vehicle). FDI attacks on vehicle positioning or velocity estimation schemes can have catastrophic impact such as accidents in a vehicular network. Hence, secure estimation in presence of such attacks is crucial for vehicle safety.

Herein, we propose a new attack type that has the capability of incorporating unique features of the V2X channel. In particular, we define a *packet drop attack* (PDA), which could fall into either of these categories (jamming or FDI). Under PDA, the malicious sensor sends a garbage packet instead of the observation packet to the fusion center, with the intention of making the remote estimator believe that the received packet is corrupted by noise and not due to any malicious attacker. Packet loss can also occur due to interference from a jammer. PDA is harmful because the remote estimator misses important observations that could have been used in the estimation process. Another challenge is that PDAs can be more challenging to disambiguate from actual packet losses that are inherent in V2X networks.

As we define the PDA herein, such attacks have not been

previously examined in the literature, to our knowledge. In contrast, there has been recent active attention paid to FDI attacks. In [9], the conditions for an undetectable attack are developed and the minimum number of attacked sensors is determined to ensure undetectability. A linear deception attack scheme is proposed in [10] to fool the popular χ^2 detector; later, a new algorithm was designed in [11] to tackle this linear deception attack. The algorithm in [11] requires the knowledge of a few *safe* sensors; this requirement was later obviated in [12] wherein an adaptive filter using a stochastic approximation technique is proposed.

Coding the sensor output for attack detection via a χ^2 detector is examined in [13]; this scheme is vulnerable to breaches in cryptographic security. Centralized and decentralized attack detection schemes for noiseless systems are developed in [14]. Attack-resilient state estimation with *bounded* noise was discussed in [15]. Attack detection and secure estimation for linear Gaussian systems was also considered in [16], but this detection scheme, which uses Kalman innovation sequence based detection, is not resilient against the linear deception attack of [10]. An optimal attack strategy with a control objective subject to a constraint on the attack detection probability is designed in [17]. Sparsity models to characterize the switching location attack in a *noiseless* linear system are designed in [18] , while considering state recovery constraints for various attack modes. Security against FDI attack in power systems is discussed in [19], [20].

There has been some limited work investigating attacks in specific vehicular networks. For example, secure estimation in presence of an FDI attack is investigated in [21], and the performance of the proposed algorithm is tested on an unmanned aerial vehicle (UAV) networks. Alternatively, [22] does design FDI attack detection scheme for vehicular networks exploiting additional (noisy) side information such as angle-of-arrival. Neither work above considers mobility as captured by the PDA model we provide here; furthermore, our methods do not require location side information. Instead, we adapt the approach of [12] wherein malicious observations are filtered without the need for location attributes or so-called safe sensors.

In this paper, we make the following contributions. (i) We extend the adaptive filter of [12] (designed to tackle FDI attacks alone), in order to address PDA and FDI attacks simultaneously. The key idea is to minimize a linear combination of three terms: the error under no attack, anomalies among estimates returned by various sensor subsets, and the deviation of the observed packet loss probability from the channel's packet error probability. We develop a learning algorithm to learn a linear filter over time, in order to minimize the mentioned objective function. This is the first work that considers secure estimation under a combination of PDA and FDI. (ii) Numerical results show that the proposed adaptive filter offers an error performance reasonably close to a Kalman filter that knows the attacked sensor set and hence can completely ignore the malicious observations. The proposed adaptive filter offers an error performance much better than the traditional Kalman filter under FDI and PDA attacks.

The rest of the paper is organized as follows. The system model is provided in Section II. Secure estimation under PDA and FDI attack is presented in Section III. Numerical results are provided in Section IV, followed by the conclusion in Section V.

II. System Model

Throughout this paper, matrices and vectors will be denoted by bold capital and bold small letters, and sets will be denoted by bold letters with caligraphic font.

A. Sensing and estimation model

We consider a set of vehicles $\mathcal{N} := \{1, 2, \cdots, N\}$ that measures a discrete-time multi-dimensional stochastic process $\{\mathbf{x}(t)\}_{t\geq 0}$ that evolves as

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{w}(t), \tag{1}$$

where the process noise $\mathbf{w}(t) \sim N(\mathbf{0}, \mathbf{Q})$ has covariance matrix $\mathbf{Q}$, and is i.i.d. across t. The observation made by vehicle i at time t is

$$\mathbf{y}_i(t) = \mathbf{C}_i\mathbf{x}(t) + \mathbf{v}_i(t), \tag{2}$$

where $\mathbf{C}_i$ is the observation matrix of vehicle i, and the observation noise $\mathbf{v}_i(t) \sim N(\mathbf{0}, \mathbf{R}_i)$. The observation noise $\mathbf{v}_i(t)$ is assumed to be independent across i and i.i.d. across t. The pair $(\mathbf{A}, \mathbf{Q}^{\frac{1}{2}})$ is assumed to be stabilizable and $(\mathbf{A}, \mathbf{C}_i)$ is assumed to be detectable for all $i \in \mathcal{N}$. The process and observation models are known at the remote estimator.

The vehicles send their observation to a remote estimator (*e.g.*, a road-side unit containing an edge computing device) via wireless links. The remote estimator estimates $\hat{\mathbf{x}}(t)$ at each time t based on the received observations[1]. Due to fading in wireless channels and receiver noise, the packet containing the observation $\mathbf{y}_i(t)$ is lost with a *probability* $p_i > 0$ *known to the fusion center*; the packet loss process is assumed to be independent across vehicles and i.i.d. across time. The assumption of known p_i can be motivated as follows: packets are tagged with the vehicle position and speed, and the fusion center can learn the packet error probability of these vehicles by observing data traffic over time. An example how the packet error probability varies as a function of average SNR (distance) and speed of vehicles for 802.11p-based transmission is found in [23].

The remote estimator seeks to minimize the time-average mean squared error (MSE):

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}||\mathbf{x}(t) - \hat{\mathbf{x}}(t)||^2. \tag{3}$$

[1]For simplicity, we assume that observations are time-synchronized and ignore transmission and processing delays. We can relax these assumptions, but only at the cost of significantly more complicated notation and with no real new insights.

If there is no packet error, the sensing and observation models can be rewritten as:

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t), \end{aligned} \tag{4}$$

where $\mathbf{y}(t) \in \mathbb{R}^{m\times 1}$ and $\mathbf{v}(t) \in \mathbb{R}^{m\times 1}$ are found by stacking $\{\mathbf{y}_i(t)\}_{1\le i\le N}$ and $\{\mathbf{v}_i(t)\}_{1\le i\le N}$, respectively. Hence, $\mathbf{C} = (\mathbf{C}_1' : \mathbf{C}_2' : \cdots : \mathbf{C}_N')'$ (where $\mathbf{C}_1'$ is the transpose of $\mathbf{C}_1$) and $\mathbf{v}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, where $\mathbf{R} = \mathrm{diag}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N)$. The minimum mean-squared error (MMSE) estimator in this case is the Kalman filter (see [24]):

$$\begin{aligned} \hat{\mathbf{x}}_{t+1|t} &= \mathbf{A}\hat{\mathbf{x}}_t \\ \mathbf{P}_{t+1|t} &= \mathbf{A}\mathbf{P}_t\mathbf{A}' + \mathbf{Q} \\ \mathbf{K}_{t+1} &= \mathbf{P}_{t+1|t}\mathbf{C}'(\mathbf{C}\mathbf{P}_{t+1|t}\mathbf{C}' + \mathbf{R})^{-1} \\ \hat{\mathbf{x}}_{t+1} &= \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}_{t+1}(\mathbf{y}(t+1) - \mathbf{C}\hat{\mathbf{x}}_{t+1|t}) \\ \mathbf{P}_{t+1} &= (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{C})\mathbf{P}_{t+1|t}, \end{aligned} \tag{5}$$

where $\hat{\mathbf{x}}_{t+1}$ is the estimate, and $\mathbf{P}_{t+1}$ is the the error covariance matrix for $\hat{\mathbf{x}}_{t+1}$, given that $\hat{\mathbf{x}}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_0)$. It can be shown that $\lim_{t\to\infty} \mathbf{P}_{t+1|t} = \bar{\mathbf{P}}$ exists and that it is the unique fixed point to the $\mathbf{P}_{t+1|t}$ iteration called the *Riccati equation* [24]. The vector sequence $\mathbf{z}_t := \mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}_{t|t-1}$ is called the *innovation sequence*; it was proved in [24] that $\{\mathbf{z}_t\}_{t\ge 1}$ is a zero-mean Gaussian sequence, pairwise independent across t, and its steady-state covariance matrix is $\mathbf{\Sigma}_z := (\mathbf{C}\bar{\mathbf{P}}\mathbf{C}' + \mathbf{R})$.

B. Attack Model

We assume that an *unknown* subset $\mathcal{A} \subset \mathcal{N}$ of vehicles can be attacked by an external attacker (with the number of attacked vehicles bounded as $|\mathcal{A}| \le n_0 < N$). If vehicle i is under attack, then, at a given time t, it either introduces an error $\mathbf{e}_i(t)$ into the observation with an unknown probability $(1-q_i)$, or sends a garbage packet to the remote estimator with probability q_i. Let $\mathbb{I}_{i,g}(t)$ be the indicator that vehicle i's sensor sends a garbage packet and $\mathbb{I}_{i,l}(t)$ be the indicator that a packet is lost due to fading or receiver noise. Let $\mathbb{I}_{i,g\oplus l}(t) = \mathbb{I}_{i,g}(t) \oplus \mathbb{I}_{i,l}(t)$, where $\oplus$ stands for the logical OR operation, and let $\emptyset$ denote any garbage packet or lost packet.

The received observation at the remote estimator at time t from vehicle i is

$$\mathbf{y}_i(t) = \begin{cases} \mathbf{C}_i\mathbf{x}(t) + \mathbf{e}_i(t) + \mathbf{v}_i(t), & i \in \mathcal{A}, \mathbb{I}_{i,g\oplus l}(t) = 0 \\ \emptyset, & i \in \mathcal{A}, \mathbb{I}_{i,g\oplus l}(t) = 1 \\ \mathbf{C}_i\mathbf{x}(t) + \mathbf{v}_i(t), & i \notin \mathcal{A}, \mathbb{I}_{i,l}(t) = 0 \\ \emptyset, & i \notin \mathcal{A}, \mathbb{I}_{i,l}(t) = 1 \end{cases} \tag{6}$$

The attacker injects the error sequence $\{\mathbf{e}_i(t)\}_{t\ge 1}$ for all $i \in \mathcal{A}$ and sometimes sends garbage packets in order to maximize the MSE given by (3).

III. Secure estimation

Here we will extend the secure estimation algorithm of [12] to the case where both FDI and PDA attacks can be present simultaneously, under the situation where the remote estimator does not have additional side information such as the identity of the attacked sensors or any safe sensor subset. This algorithm does not detect any attack, so this will be mostly useful for the situation where the system administrator cannot take necessary measures even if an attack is detected. Note that, if an attack detector is employed, then the remote estimator has the liberty to ignore the observations coming from the sensors that are declared malicious by the detector; but this approach will result in a loss of important information revealed by the malicious sensor observations. Hence, our proposed algorithm fuses *all* sensor observations at the remote estimator. The basic idea behind the proposed algorithm is to choose the filter gain matrix in a way so that the estimate anomalies from various vehicle sensor subsets are close to each other. However, since the observation noise at various sensors have different statistics, the objective function should also acount for the MSE when there is no attack. When a PDA attack is present, the packet loss rate for sensor $i \in \mathcal{A}$ will be higher than p_i; hence, the deviation of the observed packet loss rate from p_i also needs to be optimized.

In order to perform the optimization, we focus on the class of linear estimators similar to the Kalman filter in (5); but here the gain matrix $\mathbf{K}_{t+1}$ is learned via stochastic gradient descent. The goal is to minimize the time-average cost function provided in (7) (next page) over the gain matrix sequence $\{\mathbf{K}_t\}_{t\ge 0}$.

In (7), the single stage cost $c(t)$ has various components. The first term $Tr(\mathbf{P}_t)$ is the MSE when there is no attack; this is a random variable where the randomness comes from randomized $\{\mathbf{K}_t\}_{t\ge 0}$ update and channel errors. The second term captures the anomaly in estimates $\hat{\mathbf{x}}_{\mathcal{B}}(t)$ and $\hat{\mathbf{x}}_{\mathcal{B}^c}(t)$ coming from vehicle sensor subsets $\mathcal{B}$ of size n_0 and its complement, maximized over various subsets of size n_0. The number of times a garbage packet is received from vehicle sensor i due to channel error or due to attack is denoted by $N_i(t)$; under no attack, $N_i(t)/t$ converges to p_i almost surely, but converges to $1-(1-p_i)(1-q_i)$ when there is an attack. The term $(1+\xi\sum_{i\in\mathcal{B}}\max\{N_i(t)/t - p_i, 0\})$ puts a penalty for a vehicle sensor subset from which we observe more packet drops than expected. The multipliers $\lambda > 0$ and $\xi > 0$ capture the weights of various cost components.

A. The proposed algorithm

Since we do not have any closed form expression of the objective function in (7), direct computation of the gradient w.r.t. $\mathbf{K}_t$ is difficult. We will minimize the cost function (7) by iteratively learning an optimal gain matrix $\mathbf{K}^*$ via stochastic gradient descent (we update $\mathbf{K}_t$ in an on-line fashion), as more observations are gathered. Hence, we use simultaneous perturbation stochastic approximation (SPSA, see [25]) for gradient estimation, where, all elements of $\mathbf{K}_t$ are perturbed simultaneously by a random vector in two opposite directions in order to obtain $\mathbf{K}_t^+$ and $\mathbf{K}_t^-$, and a noisy gradient estimate of the single stage cost is obtained from the single stage cost values evaluated at these two perturbed gain matrices $\mathbf{K}_t^+$ and $\mathbf{K}_t^-$. This noisy gradient estimate is used in stochastic gradient descent.

$$\limsup_{\tau\uparrow\infty}\frac{1}{\tau}\sum_{t=0}^{\tau}\underbrace{\mathbb{E}\left[\mathrm{Tr}(\mathbf{P}_t)+\lambda\max_{\mathcal{B}\in 2^{\mathcal{N}}:|\mathcal{B}|=n_0}||\hat{\mathbf{x}}_{\mathcal{B}}(t)-\hat{\mathbf{x}}_{\mathcal{B}^c}(t)||^2\left(1+\xi\sum_{i\in\mathcal{B}}\max\{N_i(t)/t-p_i,0\}\right)\right]}_{:=c(t)} \quad (7)$$

The algorithm uses two non-negative sequences $\{a(t)\}_{t\geq 0}$ and $\{b(t)\}_{t\geq 0}$ satisfying the following conditions:

1) $\sum_{t=0}^{\infty} a(t)=\infty$,
2) $\sum_{t=0}^{\infty} a^2(t)<\infty$,
3) $\lim_{t\to\infty} b(t)=0$,
4) $\lim_{t\to\infty} a^2(t)/b^2(t)<\infty$ with $a(t)=t^{-k_1}, b(t)=t^{-k_2}$, and
5) $k_1\in(\frac{1}{2},1], k_2\in(0,\frac{1}{2})$.

The first two conditions are standard for stochastic approximation (see [26]). The third condition ensures that the gradient estimate is asymptotically unbiased. The fourth condition is required for the convergence of SPSA (see [25]). The fifth condition is required for the convergence of the asynchronous stochastic approximation used in this algorithm.

The algorithm also requires a large constant $l>0$.

Let us pick a small number $\delta>0$, and define $\mathcal{K}:=\{\mathbf{K}\in\mathbb{R}^{q\times m}: ||\lambda_{\max}(\mathbf{I}-\mathbf{KC})||\leq 1-\delta\}$. Recalling that $\{\mathbf{K}_t\}_{t\geq 0}$ is updated iteratively, let $\mathbf{G}_t$, $\mathbf{G}_t^+$, $\mathbf{G}_t^-$ and $\mathbf{C}_t$ denote the restrictions of $\mathbf{K}_t$, $\mathbf{K}_t^+$, $\mathbf{K}_t^-$ and $\mathbf{C}$ to the vehicle sensors whose observations are received properly by the remote estimator at time t; hence, if $\mathbf{K}_t$ is known, the fusion center can compute $\mathbf{G}_t$ by looking at the packets received from all sensors at time t. Let $\mathbf{G}_{t,\mathcal{B}}$, $\mathbf{G}_{t,\mathcal{B}}^+$ and $\mathbf{G}_{t,\mathcal{B}}^-$ denote the restrictions of $\mathbf{K}_t$, $\mathbf{K}_t^+$ and $\mathbf{K}_t^-$ to the vehicle sensors belonging to the sensor subset $\mathcal{B}$ whose observations are received properly by the remote estimator at time t. Let us denote the sensor corresponding to the j-th column of $\mathbf{K}_t$ by s_j. The proposed SECEST (abbreviation for *secure estimation*) algorithm is given below.

The SECEST Algorithm

Input: l, δ, λ, ξ, n_0, $\{p_i\}_{1\leq i\leq N}$.
Initialization: $\mathbf{K}_1$, $\mathbf{P}_0$, $\hat{\mathbf{x}}(0)=\mathbf{0}$.
For $t=1,2,3,\ldots$

1) Collect $\mathbf{y}_t$ from **all** sensors (excluding garbage packets).
2) Declare the estimate $\hat{\mathbf{x}}(t)=\mathbf{A}\hat{\mathbf{x}}(t-1)+\mathbf{G}_t(\mathbf{y}_t-\mathbf{C}_t\mathbf{A}\hat{\mathbf{x}}(t-1))$.
3) Compute the error covariance matrix $\mathbf{P}_t:=(\mathbf{I}-\mathbf{G}_t\mathbf{C}_t)(\mathbf{A}\mathbf{P}_{t-1}\mathbf{A}'+\mathbf{Q})(\mathbf{I}-\mathbf{G}_t\mathbf{C}_t)'+\mathbf{G}_t\mathbf{R}(\mathbf{G}_t)'$.
4) Pick a random matrix $\mathbf{\Delta}_t\in\{-1,1\}^{q\times m}$ such that each entry of $\mathbf{\Delta}_t$ is chosen independently with probability $\frac{1}{2}$.
5) Compute $\mathbf{K}_t^+:=\mathbf{K}_t+b(t)\mathbf{\Delta}_t$ and $\mathbf{K}_t^-:=\mathbf{K}_t-b(t)\mathbf{\Delta}_t$.
6) Compute the pseudo-estimates $\hat{\mathbf{x}}_{\mathcal{B}}^+(\mathbf{t})=\mathbf{A}\hat{\mathbf{x}}(t-1)+\mathbf{G}_{t,\mathcal{B}}^+(\mathbf{y}_t-\mathbf{C}_t\mathbf{A}\hat{\mathbf{x}}(t-1))$ and $\hat{\mathbf{x}}_{\mathcal{B}}^-(\mathbf{t})=\mathbf{A}\hat{\mathbf{x}}(t-1)+\mathbf{G}_{t,\mathcal{B}}^-(\mathbf{y}_t-\mathbf{C}_t\mathbf{A}\hat{\mathbf{x}}(t-1))$, and similarly $\hat{\mathbf{x}}_{\mathcal{B}^c}^+(\mathbf{t})$ and $\hat{\mathbf{x}}_{\mathcal{B}^c}^-(\mathbf{t})$ for all sensor subsets $\mathcal{B}$ of size n_0.
7) Compute the pseudo error covariance matrix $\mathbf{P}_t^+:=(\mathbf{I}-\mathbf{G}_t^+\mathbf{C}_t)(\mathbf{A}\mathbf{P}_{t-1}\mathbf{A}'+\mathbf{Q})(\mathbf{I}-\mathbf{G}_t^+\mathbf{C}_t)'+\mathbf{G}_t^+\mathbf{R}(\mathbf{G}_t^+)'$ and similarly compute $\mathbf{P}_t^-$.
8) Compute $c^+(t):=\mathrm{Tr}(\mathbf{P}_t^+)+\lambda(\max_{\mathcal{B}\in 2^{\mathcal{N}}:|\mathcal{B}|=n_0}||\hat{\mathbf{x}}_{\mathcal{B}}^+(t)-\hat{\mathbf{x}}_{\mathcal{B}^c}^+(t)||^2(1+\xi\sum_{i\in\mathcal{B}}\max\{\frac{N_i(t)}{t}-p_i,0\}))$, and similarly $c^-(t)$.
9) *SPSA update:* For all (i,j) such that $\mathbb{I}_{s_j,g\oplus l}=0$:

$$\hat{\mathbf{K}}_{t+1}(i,j)=\left[\mathbf{K}_t(i,j)-a(t-N_{s_j}(t))\times\frac{c^+(t)-c^-(t)}{2b(t)\mathbf{\Delta}_t(i,j)}\right]_{-l}^{l} \quad (8)$$

Project $\hat{\mathbf{K}}_{t+1}$ on $\mathcal{K}$ to obtain $\mathbf{K}_{t+1}$.

end

Observations:

- Equation (8) is a stochastic gradient descent algorithm. A noisy estimate of the gradient of $\mathbb{E}c(t)$ w.r.t. $\mathbf{K}_t(i,j)$ is used, which is $\frac{c^+(t)-c^-(t)}{2b(t)\mathbf{\Delta}_t(i,j)}$.
- (8) is an asynchronous stochastic approximation iteration (see [26]).
- We project $\hat{\mathbf{K}}_{t+1}(i,j)$ onto a compact interval $[-l,l]$ to ensure stability of the (8) iteration. Also, $|\lambda_{max}(\mathbf{I}-\mathbf{K}_{t+1}\mathbf{C}\leq(1-\delta)$ is maintained, since this ensures that the $\mathbf{P}_t$ iteration remains bounded in case there is no packet loss (see [12]); ensuring stability and convergence of SECEST under packet loss and injection attack is left for future research.
- The error covariance matrix evolves as:

$$\mathbf{P}_t:=(\mathbf{I}-\mathbf{G}_t\mathbf{C}_t)(\mathbf{A}\mathbf{P}_{t-1}\mathbf{A}'+\mathbf{Q})(\mathbf{I}-\mathbf{G}_t\mathbf{C}_t)'+\mathbf{G}_t\mathbf{R}(\mathbf{G}_t)' \quad (9)$$

where $\mathbf{K}_t$ is not updated optimally according to (5). This motivates step 3 and step 7.

IV. Numerical results

Here we numerically demonstrate the efficacy of SECEST for secure remote estimation. In [12], the authors compared a similar algorithm with the algorithm of [11], but [11] is not designed to handle packet errors. Hence, we compare our SECEST algorithm against a Kalman filter that perfectly knows at any given time instant which sensors are under attack; this benchmark algorithm is called GENIE. In GENIE, the Kalman filter ignores the observations coming from malicious sensors. We also compare the MSE performance of SECEST with a standard Kalman filter which is unaware of any attack; this algorithm is called KALMAN.

We consider a static attack (where the attacked sensor subset does not vary with time) as well as a switching location attack (where the attacked sensor subset varies with time). In each case, we consider an independent realization of a problem instance with the following parameter values. The state transition matrix $\mathbf{A}$ is taken as 0.5 times a randomly generated $q\times q$ stochastic matrix. Matrix $\mathbf{Q}$ is chosen to be a positive semidefinite (PSD) matrix, whose square root is 0.1 times a $q\times q$ random matrix whose entires uniformly chosen from $[-1,1]$. The matrix $\mathbf{R}$ is also generated similarly.

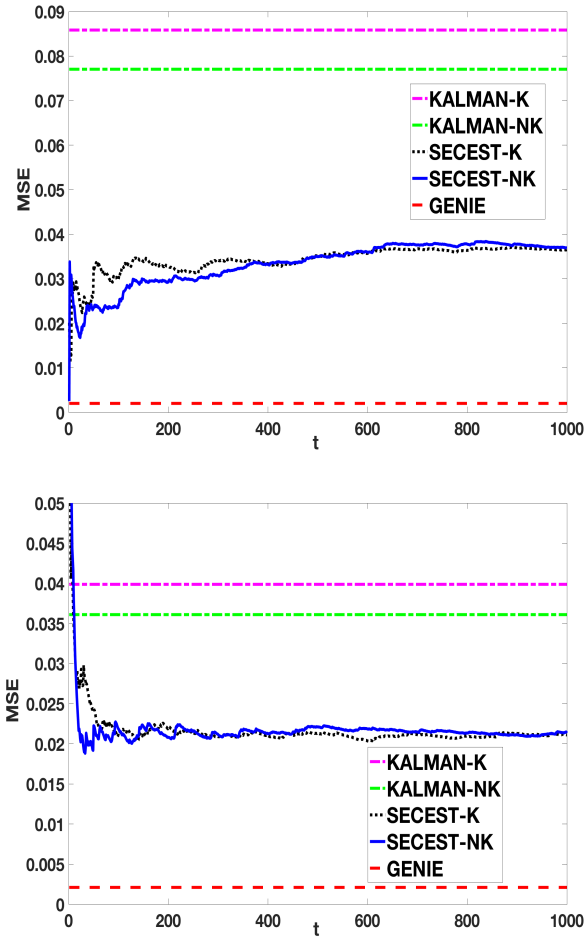

Figure 2. Performance comparison of SECEST against GENIE and KALMAN, under static attack (top) and switching location attack (bottom). $N = 8$, $n_0 = 3$, $k = 3$, $q = 3$, $\lambda = 1$, $\xi = 1$. $T = 20$ for switching location attack. Different realizations for system model are considered in the two plots. GENIE-K and GENIE-NK have very close MSE values, hence a single plot is provided for them.

Observation matrix $\mathbf{C} \in \mathbb{R}^{kN \times q}$ is chosen randomly from $[0,1]^{kN \times q}$; the observation made by each sensor is a k-dimensional column vector. The maximum number of sensors that can be under attack at a time is denoted by n_0.

For all simulation instances, the channel error probability is chosen to be 0.05, and the value of q_i for each $i \in \{1, 2, \cdots, N\}$ is chosen independently and uniformly from $[0, 0.1]$. For ease of computation, it is assumed that if a sensor i is under attack, then the packet sent by that sensor is lost with probability $0.05 + q_i$; this is a valid approximation since these probabilities are small.

The attacker always changes the sign of the innovation vectors coming from malicious sensors; this is the the worst possible linear attack [10] when the remote estimator uses a Kalman filter for estimation. In this case, let the true observation made at the sensors be $\mathbf{y}(t)$ (ignoring lost packets), and let the restriction of $\mathbf{C}$ to the sensors whose observations are not lost at time t be $\mathbf{C}(t)$. Some of the received packets will contain false observations. Let us consider another matrix $\mathbf{C}_a(t)$ which is same as $\mathbf{C}(t)$, except that the entries corresponding to the benign (not malicious) sensors whose observations are received successfully are set to 0. Similarly, let $\mathbf{y}_a(t)$ be same as $\mathbf{y}(t)$, except that the entries corresponding to the benign sensors whose observations are received successfully are 0.

We consider the situations where the attacker knows the estimate made by the remote estimator, and where the attacker can only run a Kalman filter in order to guess the estimate at the remote estimator. When the attacker knows the estimate made by the remote estimator, the received observation at time t at the remote estimator becomes $\tilde{\mathbf{y}}(t) = \mathbf{y}(t) + 2\mathbf{C}_a(t)A\hat{\mathbf{x}}(t-1) - 2\mathbf{y}_a(t)$; this is equivalent to inverting the sign of the innovation vector. We call the corresponding variants of SECEST, KALMAN and GENIE by SECEST-K, KALMAN-K and GENIE-K (with *knowledge* of the estimate).

However, if $\hat{\mathbf{x}}(t-1)$ is not known to the attacker, then the attacker can run a Kalman filter on the received observations at the estimator, in order to maintain a proxy $\hat{\mathbf{x}}_{kalman}(t-1)$ for $\hat{\mathbf{x}}(t-1)$. In this case, the the received observation at time t at the remote estimator is $\tilde{\mathbf{y}}(t) = \mathbf{y}(t) + 2\mathbf{C}_a(t)A\hat{\mathbf{x}}_{kalman}(t-1) - 2\mathbf{y}_a(t)$. We call the corresponding variants of SECEST, KALMAN and GENIE by SECEST-NK, KALMAN-NK and GENIE-NK (*no knowledge* of estimate).

Static attack: Here we assume that a fixed subset of n_0 sensors are under attack, and compare the time-average MSE of GENIE-K, GENIE-NK, KALMAN-K and KALMAN-NK with the time-variation of SECEST-K and SECEST-NK along different sample paths (due to the ergodicity of the processes) for the same problem instance; we ran simulation for multiple problem instances to verify the trend in our findings. Since GENIE-K and GENIE-NK yield very small MSE and they are close to each other, we provide one single curve for both of them. The top plot in Figure 2 shows that the time-average MSE of SECEST converges to a value reasonably close to the average MSE of GENIE for both case where the estimate is or is not available to the attacker. Also, SECEST performs much better than KALMAN in both cases; in fact, the margin of improvement over KALMAN is observed to be much higher in many other problem instances.

Switching location attack: At time $t = 1, T+1, 2T+1, \cdots$ (with $T = 20$), a random sensor subset of size n_0 is chosen in an i.i.d. fashion, and this subset is attacked over the next T slots for PDA attack and FDI via sign inversion of the innovation sequence. The probability of attacking a sensor i is proportional to $\frac{1}{i^2}$. The bottom plot of Figure 2 again shows that SECEST provides a much better MSE than KALMAN.

It is important to note that, the performance of SECEST can be further improved by choosing ξ and λ optimally; the numerical results provided here are only for one particular (λ, ξ) pair. On-line learning of these parameter values is left for future research endeavours.

V. Conclusions

Herein, we propose a generalized model for attacks in a V2X network with application to C-ITS systems. As C-ITS systems rely heavily on telemetry (sensor) information in order to provide traffic management and vehicular safety, attack mitigation is critical. The new attack model incorporates the classical false data injection attack as well as the newly defined *packet drop attack* which has the capability of capturing V2X channel characteristics induced by mobility. We have developed a secure estimation algorithm (SECEST) under both false data injection and packet drop attacks. The proposed SECEST algorithm extends the *simultaneous perturbation stochastic approximation* based techniques provided in [12] for secure estimation against false data injection attacks alone, to the case where the packet drop attack is present along with the injection attack. Numerical results demonstrate the efficacy of the proposed algorithm. In particular, SECEST outperforms a standard Kalman filter significantly, and performs reasonably close to a strategy which has side information about the identity of the attacked sensors.

References

[1] W. Sun, E. G. Ström, F. Brännström, K. C. Sou, and Y. Sui, "Radio resource management for d2d-based v2v communication," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6636–6650, 2016.

[2] E. Strom, H. Hartenstein, P. Santi, and W. Wiesbeck, "Vehicular communications: Ubiquitous networks for sustainable mobility [point of view]," *Proceedings of the IEEE*, vol. 98, no. 7, pp. 1111–1112, 2010.

[3] S. Beygi, U. Mitra, and E. G. Ström, "Nested sparse approximation: Structured estimation of v2v channels using geometry-based stochastic channel model," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4940–4955, 2015.

[4] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno, *et al.*, "Comprehensive experimental analyses of automotive attack surfaces." in *USENIX Security Symposium*. San Francisco, 2011.

[5] M. K. Hanawal, M. J. Abdel-Rahman, and M. Krunz, "Game theoretic anti-jamming dynamic frequency hopping and rate adaptation in wireless systems," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on*. IEEE, 2014, pp. 247–254.

[6] Y. Guan and X. 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*, 2017.

[7] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 911–918.

[8] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2014.

[9] Y. Chen, S. Kar, and J. M. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Transactions on Control of Network Systems*, 2017.

[10] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 4–13, 2017.

[11] Y. Li, L. Shi, and T. Chen, "Detection against linear deception attacks on multi-sensor remote state estimation," *IEEE Transactions on Control of Network Systems*, 2017.

[12] A. Chattopadhyay and U. Mitra, "Attack detection and secure estimation under false data injection attack in cyber-physical systems," in *Conference on Information Sciences and Systems (CISS), 2018*.

[13] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 106–117, 2017.

[14] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[15] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2017.

[16] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 49–59, 2017.

[17] Y. Chen, S. Kar, and J. M. Moura, "Cyber physical attacks with control objectives and detection constraints," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 1125–1130.

[18] C. Liu, J. Wu, C. Long, and Y. Wang, "Dynamic state recovery for cyber-physical systems under switching location attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 14–22, 2017.

[19] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using Kalman filter," *IEEE transactions on control of network systems*, vol. 1, no. 4, pp. 370–379, 2014.

[20] G. Liang, J. Zhao, F. Luo, S. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, 2017.

[21] Y. H. Chang, Q. Hu, and C. J. Tomlin, "Secure estimation based Kalman filter for cyber-physical systems against adversarial attacks," *arXiv preprint arXiv:1512.03853*, 2015.

[22] M. Sun, M. Li, and R. Gerdes, "A data trust framework for VANETs enabling false data detection and secure vehicle tracking," in *Communications and Network Security (CNS), 2017 IEEE Conference on*. IEEE, 2017, pp. 1–9.

[23] E. G. Ström, "On 20 MHz channel spacing for V2X communication based on 802.11 OFDM," in *Proc. Annual Conf. IEEE Ind. Electron. Soc.*, Nov. 2013.

[24] B. D. Anderson and J. B. Moore, "Optimal filtering," *Englewood Cliffs*, vol. 21, pp. 22–95, 1979.

[25] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332–341, 1992.

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