49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

False Data Injection Attacks against State Estimation in Wireless Sensor Networks

Yilin Mo*, Emanuele Garone†, Alessandro Casavola†, Bruno Sinopoli*

Abstract— **In this paper we study the effect of false data injection attacks on state estimation carried over a sensor network monitoring a discrete-time linear time-invariant Gaussian system. The steady state Kalman filter is used to perform state estimation while a failure detector is employed to detect anomalies in the system. An attacker wishes to compromise the integrity of the state estimator by hijacking a subset of sensors and sending altered readings. In order to inject fake sensor measurements without being detected the attacker will need to carefully design his actions to fool the estimator as abnormal sensor measurements would result in an alarm. It is important for a designer to determine the set of all the estimation biases that an attacker can inject into the system without being detected, providing a quantitative measure of the resilience of the system to such attacks. To this end, we will provide an ellipsoidal algorithm to compute its inner and outer approximations of such set. A numerical example is presented to further illustrate the effect of false data injection attack on state estimation.**

I. Introduction

Design and analysis of systems based on wireless sensor networks (WSNs) involve cross disciplinary research spanning domains within computer science, communication systems and control theory. A WSN is composed of low-power devices that integrate computing with heterogeneous sensing and wireless communication. WSN-based systems are usually embedded in the physical world, with which they interact by collecting, processing and transmitting relevant data. In these applications, state estimators like Kalman filters are wildly used to perform model-based state estimation on the basis of lumped-parameter models of the distributed physical phenomena.

Sensor networks span a wide range of applications, including environmental monitoring and control, health care, home and office automation and traffic control [9]. Many of these applications are safety critical. Any successful attack may significantly hamper the functionality of the sensor networks and lead to economic losses. In addition, sensors in large distributed WSNs may be physically exposed, making it difficult to ensure security and availability for each and every single sensor. The research community has acknowledged the importance of addressing the challenge of designing secure estimation and control systems [5] [6].

The impact of attacks on control systems is discussed in [7]. The authors consider two possible classes of attacks on CPS: Denial of Service (DoS) and deception (or false data injection) attacks. The DoS attack prevents the exchange of information, usually either sensor readings or control inputs between subsystems, while a false data injection attack affects the data integrity of packets by modifying their payloads. A robust feedback control design against DoS attacks has been discussed in [1]. We feel that false data inject attacks can be a subtler attack than DoS as they are in principle more difficult to detect. In this paper we want to analyze the impact of false sensor data injection attacks on the state estimator.

A significant amount of research has been carried out to analyze, detect and handle failures in sensor networks. Sinopoli et al. study the impact of random packet drops on controller and estimator performance [14]. In [17], the author reviews several failure detection algorithm in dynamical systems. Robust control and estimation [15], a discipline that aims at designing controllers and estimators that work properly under uncertainty or unknown disturbances, is applicable to some sensor network failure. However, a large part of the literature assumes that the failure is either random or benign. On the other hand, a cunning attacker can carefully design his attack strategy and deceive both detectors and robust estimators. Hence, the applicability of standard failure detection algorithms is questionable in the presence of a smart attacker.

Many scholars have undertaken research for secure data aggregation over networks in the presence of compromised sensors. In [16], the author provides a general framework to evaluate how resilient the aggregation scheme is against false sensor data. Liu et al. study estimation schemes in power grids and show how the attacker could potentially modify state estimates without being detected [12]. However, in the aforementioned work, the authors only consider static estimators that generate the current estimates only on the basis of current measurements. In general control systems have dynamics, meaning that the actions led by the attacker will not only affect the current states but also the future ones. In [13], Mo and Sinopoli extend the work of Liu et al. to study the performance of Kalman filter under false data injection attacks. They prove a necessary and sufficient condition under which the attacker could make the estimation error unbounded without being detected. No analysis is provided for the case where the attacker may still be able to incur a bounded but large estimation errors. In this case the attack may still produce great damage to the system. As a result, their characterization of system vulnerability is limited.

In this paper, we study the effect of false data injection attacks on state estimation carried over sensor networks for the general class of linear Gaussian systems. We assume that the sensor network is monitoring a discrete-time linear time-invariant Gaussian system and a Kalman filter is used to

978-1-4244-7746-3/10/$26.00 ©2010 IEEE

perform state estimation. We also assume that the system is equipped with a failure detector. An attacker wishes to alter the integrity of the state estimator in steady state by compromising a subset of sensors and sending inaccurate readings to the state estimator. In this case the attacker needs to carefully design his input to fool the estimator since abnormal sensor measurements will generally trigger an alarm. We show that the attacker's strategy can be formulated as a constrained control problem. The goal of the paper is to characterize the reachable region of such constrained control problem, i.e. the set of all the estimation biases the attacker can inject in the system without being detected. We will provide an ellipsoidal algorithm in order to compute inner and outer approximations of it.

The rest of the paper is organized as follows: in Section II, we provide the problem formulation by revisiting and adapting Kalman filter and failure detector to our scenario. In Section III, we define the threat model of false data injection attacks and formulate it as a constrained control design problem. In Section IV we discuss how to approximately solve the problem and give the upper and lower bound for the reachable region. Section V provides a numerical example to illustrate the effect of a false data injection attack on the state estimator. Finally Section VI concludes the paper.

II. Problem Formulation

In this section we will introduce our assumption for the dynamical system, which we assume equipped with a Kalman filter and a failure detector.

We assume that the physical system is a discrete-time linear time-invariant (LTI) taking the following form:

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the vector of state variables at time k, $w_k \in \mathbb{R}^n$ is the process noise at time k and x_0 is the initial state. We assume w_k, x_0 are independent Gaussian random variables with $x_0 \sim \mathcal{N}(0,\ \Sigma)$, $w_k \sim \mathcal{N}(0,\ Q)$.

We assume that a sensor network is deployed to monitor the system described in (1). At each time instant all sensor readings are sent to a centralized estimator. The observation equation can be written as

$$y_k = Cx_k + v_k, \tag{2}$$

where $y_k = [y_{k,1}, \ldots, y_{k,m}]^T \in \mathbb{R}^m$ is a vector of sensor measurements, and $y_{k,i}$ is the measurement taken by sensor i at time k. It is assumed that $v_k \sim \mathcal{N}(0,\ R)$, the measurement noise, is independent of x_0 and w_k.

A Kalman filter is used to compute state estimation $\hat{x}_k$ from observations y_k, taking the following form:

$$\begin{aligned} \hat{x}_{0|-1} &= 0,\ P_{0|-1} = \Sigma, \\ \hat{x}_{k+1|k} &= A\hat{x}_k,\ P_{k+1|k} = AP_kA^T + Q, \\ K_k &= P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}, \\ \hat{x}_k &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}), \\ P_k &= P_{k|k-1} - K_kCP_{k|k-1}. \end{aligned} \tag{3}$$

Although the Kalman filter uses a time varying gain K_k, it is well known that this gain will converge to a steady state value if system (1) is detectable. In practice the Kalman filter gain usually converges in a few steps. For such a reason, in this paper we will assume that the Kalman filter gain is operating at steady state. If we define

$$P \triangleq \lim_{k\to\infty} P_{k|k-1},\ K \triangleq PC^T(CPC^T + R)^{-1}$$

then the constant Kalman filter gain satisfies

$$\hat{x}_{k+1} = A\hat{x}_k + K\left(y_{k+1} - CA\hat{x}_k\right). \tag{4}$$

Next, let us define the residue z_{k+1} at time $k+1$ as

$$z_{k+1} \triangleq y_{k+1} - CA\hat{x}_k. \tag{5}$$

This quantity allows one to simplify (4) as follows

$$\hat{x}_{k+1} = A\hat{x}_k + Kz_{k+1}. \tag{6}$$

The estimation error e_k at time k is defined as the difference between the state x_k and its estimate $\hat{x}_k$

$$e_k \triangleq x_k - \hat{x}_k. \tag{7}$$

Manipulating (4) and (5), (7) can be described by the following recursive equation:

$$e_{k+1} = (A - KCA)e_k + (I - KC)w_k - Kv_k. \tag{8}$$

We further assume a failure detection algorithm is implemented in the sensor networks. Failure detectors are used in control system for various reasons such as sensor/actuator failure detection or intrusion detection (see e.g. [4], [8] for a survey on the topic). For example, the χ^2 failure detector computes the following quantities

$$g_k = z_k^T\mathcal{P}^{-1}z_k, \tag{9}$$

where $\mathcal{P}$ is the covariance matrix of the residue z_k. Since z_k follows a Gaussian distribution, g_k is χ^2 distributed with m degrees of freedom and cannot be far away from 0. As a result, the χ^2 failure detector compares g_k with a certain threshold. If g_k is greater than the threshold, then the detector will trigger an alarm.

Other types of failure detectors have also been considered by many researchers. In [3] [11], the authors design a linear filter other than the Kalman filter to detect sensor shift or shift in matrices A and B. The gain of such filter is chosen to make the residue of the filter more sensitive to certain shift, which helps to detect a particular failure. A generalized likelihood ratio test to detect dynamics or sensor jump is also proposed by Willsky et al. in [18].

To make the discussion more general, we shall assume that the detector is instructed to trigger an alarm when:

$$g_k > threshold, \tag{10}$$

where the residue g_k is defined as

$$g_k \triangleq g(z_k,\ y_k,\ \hat{x}_k, \ldots, z_{k-\mathcal{T}+1},\ y_{k-\mathcal{T}+1},\ \hat{x}_{k-\mathcal{T}+1}). \tag{11}$$

The function g is continuous and $\mathcal{T} \in \mathbb{N}$ is the window size of the detector. It is easy to see that for a χ^2 detector,

$g_k = z_k^T \mathcal{P}^{-1} z_k$. We further define the probability of alarm for the failure detector to be

$$\beta_k = P(g_k > threshold). \tag{12}$$

III. False Data Injection Attacks

In this section we will show how an attacker can generate fake measurements for the system described in Section II without being noticed by the fault detector. As it will be clear at the end of this section, such a formulation will be instrumental to the characterization of the vulnerability of the system.

We assume that the attacker has the following capabilities:

1) It knows matrices A, C, Q, R as described in Section II and the Kalman filter gain K.
2) It can compromise the readings of a subset of the sensors, which we denote as S_{bad}. As a result, (2) now becomes
$$y'_k = Cx_k + v_k + \Gamma y^a_k, \tag{13}$$
where $\Gamma \in \mathbb{R}^{m \times |S_{bad}|}$ is defined as the non-zero rows of $diag(\gamma_1, \ldots, \gamma_m)$, where γ_i is a binary variable with $\gamma_i = 1$ if and only if $i \in S_{bad}$ and $y^a_k \in \mathbb{R}^{|S_{bad}|}$ is the malicious offset injected by the attacker. Here we write the observation as y'_k since it is different from y_k generated by the healthy system.
3) We assume that y^a_k is designed by the attacker offline. Therefore y^a_k is assumed to be independent of w_i, v_i, x_0.
4) We let the intrusion begin at time 1.

Definition 1: An attack sequence $\mathcal{Y}$ is defined as an infinite sequence of attacker's inputs of the following form $y^a_1, y^a_2, \ldots$ [1]

Notation : The state estimate $\hat{x}'_k$ of the partially compromised system is a function of $\mathcal{Y}$ that we denote hereafter as $\hat{x}'_k(\mathcal{Y})$. However, in order to improve legibility, we will use $\hat{x}'_k$ when there is no confusion. □

Because of the above assumptions, the new system dynamics can be written as

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \\ y'_k &= Cx_k + v_k + \Gamma y^a_k, \\ \hat{x}'_{k+1} &= A\hat{x}'_k + K\left(y'_{k+1} - CA\hat{x}'_k\right), \\ \hat{x}'_0 &= 0, \end{aligned} \tag{14}$$

and the new residue and estimation error become

$$\begin{aligned} z'_{k+1} &\triangleq y'_{k+1} - CA\hat{x}'_k, \\ e'_k &\triangleq x_k - \hat{x}'_k. \end{aligned} \tag{15}$$

Finally, the new probability of alarm is defined as

$$\beta'_k = P(g'_k > threshold), \tag{16}$$

where

$$g'_k \triangleq g(z'_k, y'_k, \hat{x}'_k, \ldots, z'_{k-\mathcal{T}+1}, y'_{k-\mathcal{T}+1}, \hat{x}'_{k-\mathcal{T}+1}). \tag{17}$$

[1]For simplicity here we define the attack sequence to be infinitely long. If the attack stops at time T, then we can let $y^a_k = 0$, $k \geq T$.

The differences between the compromised and the healthy system are

$$\begin{aligned} \Delta\hat{x}_k &\triangleq \hat{x}'_k - \hat{x}_k, & \Delta z_k &\triangleq z'_k - z_k, \\ \Delta e_k &\triangleq e'_k - e_k, & \Delta\beta_k &= \beta'_k - \beta_k. \end{aligned} \tag{18}$$

One can prove that the following equalities hold true

$$\Delta e_{k+1} = (A - KCA)\Delta e_k - K\Gamma y^a_{k+1},\ \Delta e_0 = 0, \tag{19}$$

$$\Delta z_k = CA\Delta e_{k-1} + y^a_k. \tag{20}$$

Since the attacker needs to fool the failure detector to avoid triggering an alarm during the attack, it would design an attack action $\mathcal{Y}$ such that $\beta'_k(\mathcal{Y})$ is lower than ε during the attack. In other words the attacker ideally would design the input so that the failure detector will trigger an alarm with no more than ε probability at each step. In principle the exact value of β'_k depends on the function g implemented by the failure detector but, typically, such information is confidential and not available to the attacker. Therefore all the attacker can do is to design the malicious injection by keeping the KL distance between z_k and z'_k small. The latter discussion leads to the following definition of a sequence the attacker can design:

Definition 2: An attack sequence $\mathcal{Y}$ is called (T, α)-feasible if the following inequality holds:

$$D(z_k||z'_k) = \|\Delta z_k(\mathcal{Y})\|_S = \sqrt{\Delta z_k^T S \Delta z_k} \leq \alpha, \tag{21}$$

for all $k = 1, \ldots, T$, where $S = \mathcal{P}^{-1}/2$ and $\mathcal{P}$ is the error covariance of z_k, and $D(z_k||z'_k)$ is the Kullback-Leibler distance between z_k and z'_k.

The following theorem states that β'_k is continuous with respect to $D(z_k||z'_k)$. Hence, β'_k will converge to β_k as $D(z_k||z'_k)$ goes to zero, which validates our definition of feasible attack.

Theorem 1: For any $\varepsilon > 0$, there exists $\alpha > 0$, such that if

$$D(z_k||z'_k) \leq \alpha,$$

for all k from 1 to T, then

$$\beta'_k \leq \beta_k + \varepsilon,$$

for all k from 1 to T.

Proof: The proof is based on the continuity of g function and the boundedness of $\Delta\hat{x}_k$. Due to the space limitation, the complete proof is omitted. The reader can refer to [13] for more details. ■

To characterize the performance of state estimator under the attack, it is sufficient to compute the distribution of the state estimation error e'_k. Since $e'_k = \Delta e_k + e_k$ and Δe_k is a function of $\mathcal{Y}$, which is independent of e_k, one can prove that the e'_k conditioned on $\mathcal{Y}$ is Gaussian and that moreover

$$\begin{aligned} E(e'_k|\mathcal{Y}) &= \Delta e_k, \\ Cov(e'_k|\mathcal{Y}) &= Cov(e_k) = \mathcal{P}. \end{aligned}$$

As a result, in order to characterize the performance of the state estimator we only need to study the reachable region of

Δe_k. To this end, let us define the (T,α)-reachable region of Δe_k as follows

$$R_{T,\alpha} \triangleq \{x \in \mathbb{R}^n : x = \Delta e_T(\mathcal{Y}),\ \mathcal{Y} \text{ is } (T,\alpha)\text{-feasible}\} \tag{22}$$

and the α-reachable region as

$$\mathcal{R}_\alpha \triangleq \bigcup_{T=1}^{\infty} R_{T,\alpha}. \tag{23}$$

By the linearity of the system, it is easy to see that

$$R_{T,\alpha} = \alpha R_{T,1},\ \mathcal{R}_\alpha = \alpha \mathcal{R}_1.$$

As a result, we only need to compute the $(T,1)$-reachable and 1-reachable regions. To simplify the notation, we will denote $R_{T,1}$ as R_T and $\mathcal{R}_1$ as $\mathcal{R}$. We further define $R_0 = \{\Delta e_0\} = \{0\}$.

In the next section, we will provide a recursive algorithm to find the inner and outer approximations of $\mathcal{R}$.

IV. An efficient algorithm to evaluate $\mathcal{R}$

In this section, we will provide an efficient algorithm to approximate $\mathcal{R}$. To this end we will first give a recursive algorithm to compute set R_k. However in practice such an algorithm may be hardly implemented in an exact way due to its high computational requirements. Then, in order to overcome such a limitation, we will relax the problem and aim at finding inner and outer approximations of R_k and $\mathcal{R}$.

By recalling (19) and (20), we can write

$$\begin{aligned}\Delta e_{k+1} &= (A - KCA)\Delta e_k - K\Gamma y^a_{k+1},\ \Delta e_0 = 0,\\ \Delta z_k &= CA\Delta e_{k-1} + y^a_k.\end{aligned}$$

Because $\mathcal{Y}$ is $(k,1)$-feasible, then it will also be $(k-1,1)$-feasible and the following proposition holds true:

Proposition 1: Given R_{k-1}, R_k can be defined recursively as

$$\begin{aligned}R_k = \{&x \in \mathbb{R}^n : \exists e \in R_{k-1},\ \exists y \in \mathbb{R}^{|S_{bad}|},\\ &x = (A - KCA)e - K\Gamma y,\ \|CAe + \Gamma y\|_S \le 1\}.\end{aligned} \tag{24}$$

Since we know that $R_0 = \{0\}$, we can recursively compute R_k and hence get R. However, the main drawback of this approach is that the shape of R_k becomes more and more complex when k increases. As a result, an analytical description will be impossible over time. For similar reasons even an exact numerical solution could take too long to complete and could require a large amount of memory [10].

For security purposes we will be satisfied with understanding the magnitude of R_k and possibly have a tight over-approximation of its exact shape. As a consequence we will propose an ellipsoidal approximation of R_k, similar to the one described in [2], that allows us to find suitable inner and outer approximations of R_k. Before describing the procedure, let us make the following mild assumptions:

1) The pair $(A - KCA, K\Gamma)$ is controllable.
2) $(A - KCA)$ is invertible.

Note that first assumption is without loss of generality since if $(A - KCA, K\Gamma)$ were not controllable, we could restrict ourselves to consider the controllable subspace. If $A - KCA$ is invertible, we can easily obtain:

$$\begin{aligned}\Delta e_{k-1} &= (A - KCA)^{-1}\Delta e_k + (A - KCA)^{-1}K\Gamma y^a_k,\\ \Delta z_k &= CA(A - KCA)^{-1}e_k\\ &\quad + [CA(A - KCA)^{-1}K + I_m]\Gamma y^a_k.\end{aligned} \tag{25}$$

To simplify notations, let us define

$$\begin{aligned}&\tilde{A} \triangleq (A - KCA)^{-1},\ \tilde{B} \triangleq (A - KCA)^{-1}K\Gamma,\\ &\tilde{C} \triangleq CA(A - KCA)^{-1},\ \tilde{D} \triangleq [CA(A - KCA)^{-1}K + I_m]\Gamma.\end{aligned}$$

R_k can then be evaluated recursively as

$$\begin{aligned}R_k = \{&x \in \mathbb{R}^n : \exists y \in \mathbb{R}^{|S_{bad}|},\\ &\tilde{A}x + \tilde{B}y \in R_{k-1}, \tilde{C}x + \tilde{D}y \in \mathcal{E}(S)\,\},\end{aligned} \tag{26}$$

where the function $\mathcal{E}$ maps a positive definite matrix X into an ellipsoid in $\mathbb{R}^n$, such that

$$\mathcal{E}(X) = \{v \in \mathbb{R}^n : v^T X v \le 1\}. \tag{27}$$

By the controllability of $(A - KCA, K\Gamma)$, the reachable region R_n must contain a neighborhood of the origin. Therefore, there exists a positive definite matrix U such that

$$R_0 = \{0\} \subset \mathcal{E}(U) \subset R_n.$$

Suppose that the ellipsoidal inner and outer approximations of R_k are respectively $\underline{R}_k$ and $\overline{R}_k$, which can be defined recursively as

$$\begin{aligned}\overline{R}_k = Out[\{&x \in \mathbb{R}^n : \exists y \in \mathbb{R}^{|S_{bad}|},\\ &\tilde{A}x + \tilde{B}y \in \mathcal{E}(U_{k-1}), \tilde{C}x + \tilde{D}y \in \mathcal{E}(S)\,\}],\ \overline{R}_0 = \mathcal{E}(U),\end{aligned} \tag{28}$$

and

$$\begin{aligned}\underline{R}_k = In[\{&x \in \mathbb{R}^n : \exists e \in \underline{R}_{k-1},\ \exists y \in \mathbb{R}^{|S_{bad}|},\\ &\tilde{A}x + \tilde{B}y \in \mathcal{E}(L_{k-1}), \tilde{C}x + \tilde{D}y \in \mathcal{E}(S)\,\}],\ \underline{R}_n = \mathcal{E}(U),\end{aligned} \tag{29}$$

where L_k, U_k are positive definite matrices such that $\underline{R}_k = \mathcal{E}(L_k)$ and $\overline{R}_k = \mathcal{E}(U_k)$, the operations $Out[\cdot]$ and $In[\cdot]$ indicate the outer and inner ellipsoidal approximation, which will be specified later. We know that the following theorem holds.

Theorem 2:

$$\begin{aligned}&R_k \subseteq \overline{R}_k,\ k = 0,\ 1, \dots\\ &\underline{R}_k \subseteq R_k,\ k = n,\ n+1, \dots\end{aligned} \tag{30}$$

Proof: The theorem can be proved by induction. ■

Then, we can finally embed $\mathcal{R}$ by using its outer and inner approximations as:

Corollary 1:

$$\underline{\mathcal{R}} = \bigcup_{k=n}^{\infty} \underline{R}_k \subseteq \mathcal{R} \subseteq \bigcup_{k=0}^{\infty} \overline{R}_k = \overline{\mathcal{R}}. \tag{31}$$

Proof: The proof follows from the definition of $\mathcal{R}$. ■

Next we will show how to compute $\underline{R}_k$ and $\overline{R}_k$. First, let us focus on the outer ellipsoid approximation. To this end, let us extend the space from $\mathbb{R}^n$ to $\mathbb{R}^{n+|S_{bad}|}$. By (28), the

vector $[x^T,\, y^T]^T$ will be in the intersection of the following ellipsoids in the extended space:

$$\begin{aligned}&\left\{[x^T, y^T]\begin{bmatrix}\tilde{A}^T U_k \tilde{A} & \tilde{A}^T U_k \tilde{B}\\ \tilde{B}^T U_k \tilde{A} & \tilde{B}^T U_k \tilde{B}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \le 1\right\}\\ &\bigcap\left\{[x^T, y^T]\begin{bmatrix}\tilde{C}^T S \tilde{C} & \tilde{C}^T S \tilde{D}\\ \tilde{D}^T S \tilde{C} & \tilde{D}^T S \tilde{D}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \le 1\right\}.\end{aligned} \tag{32}$$

The minimum volume ellipsoidal outer approximation of the above intersection is given by the following Theorem.

Theorem 3: The minimum volume ellipsoidal outer approximation of the region

$$\{x^T S_1 x \le 1\} \bigcap \{x^T S_2 x \le 1\},$$

where S_1, S_2 are positive semidefinite, is

$$\{x^T S_3 x \le 1\},$$

where S_3 is the solution of the following convex optimization problem:

$$\begin{aligned}\max_{\alpha} \quad & logdet(S_3)\\ & S_3 = \alpha S_1 + \beta S_2\\ & \alpha + \beta = 1,\ \alpha,\ \beta \ge 0.\end{aligned} \tag{33}$$

Proof: The proof is omitted due to the space constraints. ■

As a result, we can find α_k, β_k, such that the outer approximation of (32) is given by

$$\overline{R}'_{k+1} = \left\{[x^T, y^T] U'_{k+1}\begin{bmatrix}x\\ y\end{bmatrix} \le 1\right\}, \tag{34}$$

where

$$U'_{k+1} = \alpha_k \begin{bmatrix}\tilde{A}^T U_k \tilde{A} & \tilde{A}^T U_k \tilde{B}\\ \tilde{B}^T U_k \tilde{A} & \tilde{B}^T U_k \tilde{B}\end{bmatrix} + \beta_k \begin{bmatrix}\tilde{C}^T S \tilde{C} & \tilde{C}^T S \tilde{D}\\ \tilde{D}^T S \tilde{C} & \tilde{D}^T S \tilde{D}\end{bmatrix}. \tag{35}$$

Once such an a high dimensional ellipsoid $\overline{R}'_{k+1}$ is obtained, in order to project it into the subspace of x we will use the following result.

Theorem 4:

$$\begin{aligned}&Proj_x\left\{[x^T, y^T]\begin{bmatrix}S_1 & S_2\\ S_2^T & S_3\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \le 1\right\}\\ &= \{x^T (S_1 - S_2 S_3^+ S_2^T) x \le 1\},\end{aligned} \tag{36}$$

where $\begin{bmatrix}S_1 & S_2\\ S_2^T & S_3\end{bmatrix}$ is positive semidefinite and S_3^+ is the Moore-Penrose inverse.

As a consequence of the above theorem, we have

$$\overline{R}_{k+1} = \{x^T U_{k+1} x \le 1\}, \tag{37}$$

where

$$\begin{aligned}U_{k+1} = {} & \alpha_k \tilde{A}^T U_k \tilde{A} + \beta_k \tilde{C}^T S \tilde{C} - (\alpha_k \tilde{A}^T U_k \tilde{B} + \beta_k \tilde{C}^T S \tilde{D})\\ & \times (\alpha_k \tilde{B}^T U_k \tilde{B} + \beta_k \tilde{D}^T S \tilde{D})^+ (\alpha_k \tilde{A}^T U_k \tilde{B} + \beta_k \tilde{C}^T S \tilde{D})^T.\end{aligned} \tag{38}$$

Finally, since we already know that $U_0 = U$, a recursive algorithm for the outer approximation of R_k is obtained.

For what regards the inner approximation algorithm, the procedure is similar to the outer approximation. First we extend the space. By (29), vector $[x^T, y^T]^T$ is in the intersection of two ellipsoids

$$\begin{aligned}&\left\{[x^T, y^T]\begin{bmatrix}\tilde{A}^T L_k \tilde{A} & \tilde{A}^T L_k \tilde{B}\\ \tilde{B}^T L_k \tilde{A} & \tilde{B}^T L_k \tilde{B}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \le 1\right\}\\ &\bigcap\left\{[x^T, y^T]\begin{bmatrix}\tilde{C}^T S \tilde{C} & \tilde{C}^T S \tilde{D}\\ \tilde{D}^T S \tilde{C} & \tilde{D}^T S \tilde{D}\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} \le 1\right\}.\end{aligned} \tag{39}$$

Then, the maximum volume inner ellipsoidal approximation is given by the following theorem.

Theorem 5: Consider the intersection of the two ellipsoids

$$\{x^T S_1 x \le 1\} \bigcap \{x^T S_2 x \le 1\}, \tag{40}$$

where S_1, S_2 are positive semidefinite. Then S_1 and S_2 can be diagonalized at the same time, which means that

$$S_1 = X \Lambda_1 X^T,\ S_2 = X \Lambda_2 X^T, \tag{41}$$

where X is invertible and Λ_1, Λ_2 are diagonal. The maximum volume ellipsoid inner approximation of the region defined by (40) is given by

$$\{x^T S_3 x \le 1\},$$

where

$$S_3 = X \max(\Lambda_1,\, \Lambda_2) X^T, \tag{42}$$

and max means elementwise maximum.

Proof: Due to space limit, we will omit the proof. ■

Finally, once computed the maximum volume inner approximation of (39), we can project it to get $\underline{R}_{k+1}$.

V. Illustrative Examples

In this section, we will provide a numerical example to illustrate the effects of false data injection attacks on a WSN.

Consider a vehicle which is moving along the x-axis. The state space includes position x and velocity $\dot{x}$ of the vehicle. As a result, the discrete-time system dynamics are as follows:

$$\begin{aligned}\dot{x}_{k+1} &= \dot{x}_{k+1} + w_{k,1},\\ x_{k+1} &= x_k + \dot{x}_k + w_{k,2},\end{aligned} \tag{43}$$

which can be written in the matrix form as

$$X_{k+1} = \begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix} X_k + w_k, \tag{44}$$

where

$$X_k = \begin{bmatrix}\dot{x}\\ x\end{bmatrix},\ w_k = \begin{bmatrix}w_{k,1}\\ w_{k,2}\end{bmatrix}. \tag{45}$$

Suppose two sensors are measuring velocity and position respectively. Hence

$$y_k = X_k + v_k. \tag{46}$$

We consider two cases, where either the velocity sensor or the position sensor is compromised, i.e. $\Gamma = [1, 0]'$ or $\Gamma = [0, 1]'$. We assume that the covariance of the noise is $Q = R = I$. The steady state Kalman gain in this case is

$$K = \begin{bmatrix}0.5939 & 0.0793\\ 0.0793 & 0.6944\end{bmatrix}. \tag{47}$$

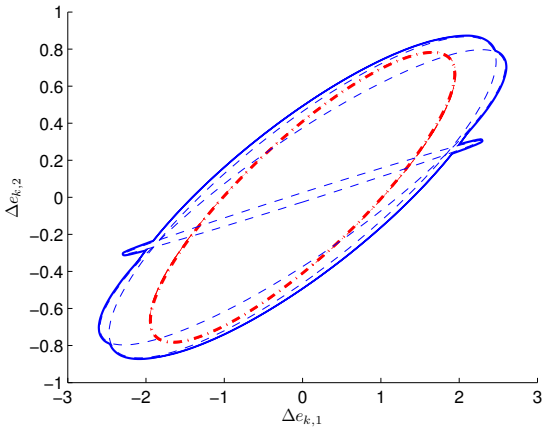

Fig. 1. Inner and Outer Approximation When the Velocity Sensor is Compromised

Figure 1 shows the inner and outer approximation $\underline{\mathcal{R}}$ and $\overline{\mathcal{R}}$ of the reachable region $\mathcal{R}$ when the velocity sensor is compromised. The blue solid line is the union of all $\overline{R}_k$ and red dashed line is the union of the inner approximation $\underline{R}_k$, both taken from $k = 0$ to 20. From the simulation we can conclude that the inner and outer approximation $\underline{\mathcal{R}}$ and $\overline{\mathcal{R}}$ are bounded. The blue dash line is $\overline{R}_k$ for $k = 2, 4, 8$. It can be seen that the inner and outer approximations are quite tight.

Figure 2 shows the inner and outer approximation $\underline{R}_k$ and $\overline{R}_k$ of the reachable region $\mathcal{R}$ when the position sensor is compromised. The thick blue line is $\overline{R}_{10}$ and the other dashed blue lines represent $\overline{R}_0$ to $\overline{R}_9$ respectively. Similarly, the thick red line is $\underline{R}_{12}$ and the other red lines are $\underline{R}_2$ to $\underline{R}_{11}$. By continuing the computation of $\underline{R}_k$ and $\overline{R}_k$ it results that the region grows over time. The reason behind this behavior is that the system is not detectable when using only the velocity sensor and, as a result, the attacker would make arbitrary large errors on the position estimation. For a more detailed discussion on this special case, please refer to [13].

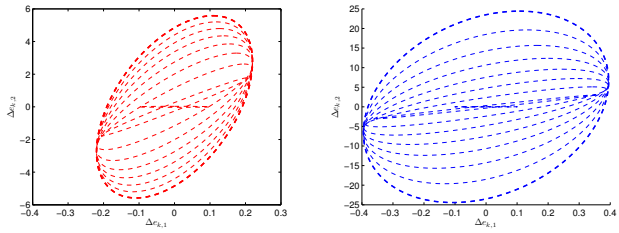

Fig. 2. Inner(left) and Outer(right) Approximation When Position Sensor is Compromised

VI. Conclusion

In this paper, we study the effect of false data injection attacks on state estimation carried over sensor networks. We formulate the false data injection attack as a constrained control problem and provide an algorithm, based on ellipsoidal approximation, to compute the inner and outer approximations for the reachable region of the constrained control problem. We also present a numerical example to further illustrate the effect of false data injection attacks on state estimation.

Future work will include characterizing the shape and size of the inner and outer ellipsoidal approximation and consider the case where both denial of service attacks and false data injection attacks may happen at the same time.

References

[1] S. Amin, A. Cardenas, and S. S. Sastry. Safe and secure networked control systems under denial-of-service attacks. In *Hybrid Systems: Computation and Control*, pages 31–45. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, April 2009.

[2] D. Angeli, A. Casavola, G. Franzè, and E. Mosca. An ellipsoidal off-line mpc scheme for uncertain polytopic discrete-time systems. *Automatica*, 44(12):3113–3119, December 2008.

[3] R. V. Beard. Failure accommodation in linear systems through self-reorganization. Technical Report MVT-71-1, Man Vehicle Laborotory, Cambrideg, Massachusetts, February 1971.

[4] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder. *Diagnosis and Fault-Tolerant Control*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[5] E. Byres and J. Lowe. The myths and facts behind cyber security risks for industrial control systems. In *Proceedings of the VDE Kongress*. VDE Congress, 2004.

[6] A. A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *HOTSEC'08: Proceedings of the 3rd conference on Hot topics in security*, pages 1–6, Berkeley, CA, USA, 2008. USENIX Association.

[7] A. A. Cárdenas, S. Amin, and S. Sastry. Secure control: Towards survivable cyber-physical systems. In *Distributed Computing Systems Workshops, 2008. ICDCS '08. 28th International Conference on*, pages 495–500, June 2008.

[8] J. Chen and R. J. Patton. *Robust model-based fault diagnosis for dynamic systems*. Kluwer Academic Publishers, Norwell, MA, USA, 1999.

[9] N. P. M. (Ed.). *Sensor Networks and Configuration*. Springer, 2007.

[10] T. A. Johansen. Approximate explicit receding horizon control of constrained nonlinear systems. *Automatica*, 40(2):293 – 300, 2004.

[11] H. L. Jones. *Failure Detection in Linear Systems*. PhD thesis, M.I.T., Cambridge, Massachusetts, 1973 1973.

[12] Y. Liu, P. Ning, and M. Reiter. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, November 2009.

[13] Y. Mo and B. Sinopoli. False data injection attacks in cyber physical systems. In *First Workshop on Secure Control Systems*, Stockholm, Sweden, April 2010.

[14] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163–187, Jan. 2007.

[15] R. Stengel and L. Ryan. Stochastic robustness of linear time-invariant control systems. *Automatic Control, IEEE Transactions on*, 36(1):82–87, Jan 1991.

[16] D. Wagner. Resilient aggreagion in sensor networks. In *ACM Workshop on Security of Ad Hoc and Sensor Networks*, Oct 25 2004.

[17] A. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12:601–611, Nov 1976.

[18] A. S. Willsky and H. L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21:108–112, Feburary 1976.