False Data Injection Attacks in Control Systems

Yilin Mo, Bruno Sinopoli *†

Abstract

This paper analyzes the effects of false data injection attacks on Control System. We assume that the system, equipped with a Kalman filter and LQG controller, is used to monitor and control a discrete linear time invariant Gaussian system. We further assume that the system is equipped with a failure detector. An attacker wishes to destabilize the system by compromising a subset of sensors and sending corrupted readings to the state estimator. In order to inject fake sensor measurements without being detected the attacker needs to carefully design its inputs to fool the failure detector, since abnormal sensor measurements usually trigger an alarm from the failure detector. We will provide a necessary and sufficient condition under which the attacker could destabilize the system while successfully bypassing the failure detector. A design method for the defender to improve the resilience of the CPS against such kind of false data injection attacks is also provided.

1. Introduction

Cyber Physical Systems (CPS) refer to the embedding of widespread sensing, computation, communication and control into physical spaces [1]. Application areas are as diverse as aerospace, chemical processes, civil infrastructure, energy, manufacturing and transportation, most of which are safety-critical. The availability of cheap communication technologies such as the internet makes such infrastructures susceptible to cyber security threats, which may affect national security as some of them, such as the power grid, are vital to the normal operation of our society. Any successful attack may significantly hamper the economy, the environment or may even lead to loss of human life. As a result, security is of primary importance to guarantee safe operation of CPS. The research community has acknowledged the importance of addressing the challenge of designing secure CPS [2] [3].

The impact of attacks on the control systems is addressed in [4]. The authors consider two possible classes of attacks on the CPS: Denial of Service (DoS) attacks and deception attacks (or false data injection attacks). The DoS attack prevents the exchange of information, usually either sensor readings or control inputs between subsystems, while false data injection attack affects the data integrity of packets by modifying their payloads. A robust feedback control design against DoS attacks is further discussed in [5]. We feel that false data injection attacks can be subtler than DoS attacks as they are in principle more difficult to detect and have not been thoroughly investigated. In this paper, we want to analyze the impact of false data injection attacks on control systems.

A significant amount of research effort has been carried out to analyze, detect and handle failures in control systems. Sinopoli et al. study the impact of random packet drops on controller and estimator performance [6]. In [7], the author reviews several failure detection algorithms in dynamic systems. Results from robust control and estimation [8], a discipline that aims at designing controllers and estimators that function properly under uncertain parameters or unknown disturbances, is also applicable to some control system failures. However, a large proportion of the literature assumes that the failure is either random or benign. On the other hand, a cunning attacker can carefully design its attack strategy and deceive both detectors and robust estimators. Hence, the applicability of failure detection algorithms is questionable in the presence of a smart attacker.

Before describing our problem setup we wiah to review some of the existing literature concerning secure data aggregation over networks in the presence of compromised sensors. In [9], the author provides a general framework to evaluate how resilient the aggregation scheme is against compromised sensor data. Liu et al. study the estimation scheme in power grids and show

*Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA. Email: ymo@andrew.cmu.edu, brunos@ece.cmu.edu

†This research is supported in part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office Foundation. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of ARO, CMU, or the U.S. Government or any of its agencies.

that under some assumptions the attacker can modify the state estimate undetected [10]. However, in both studies, the authors only consider static systems with estimators that rely exclusively upon current sensor measurements. In reality, in a control system the actions taken by the attacker will not only affect the current states but also the future ones. An attacker could potentially use this fact to perform its attack over time and destabilize the system. On the other hand, the dynamics of the system could be used by the failure detector since the attack may be detected in the near future even if it results undetectable when it first occurs.

In this paper, we study the effects of false data injection attacks on control systems. We assume that the control system, which is equipped with a Kalman filter, LQG controller and failure detector, is monitoring and controlling a linear time-invariant system. The attacker's goal is to destabilize the system by compromising a subset of sensors and sending altered readings to the state estimator. The attacker also wants to guarantee that its action can bypass the failure detector. Under these assumptions, we will give a necessary and sufficient condition under which the attacker could destabilize the system without being detected.

The rest of the paper is organized as follows: In Section 2 we formulate the problem by revisiting and adapting Kalman filter, LQG controller and failure detector to our scenario. In Section 3, we define the threat model of false data injection attacks. In Section 4 we prove a necessary and sufficient condition under which the attacker could destabilize the system. We will also give some design criteria to improve the resilience of the CPS against false data injection attacks. A numerical example is provided in Section 5 to illustrate the effects of false data injection attacks on the CPS. Finally Section 6 concludes the paper.

2. Problem Formulation

In this section we model the CPS as a linear control system, which is equipped with a Kalman filter, LQG controller and failure detector.

2.1. Physical System

We assume that the physical system has Linear Time Invariant (LTI) dynamics, which take the following form:

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

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

2.2. Kalman filter

A sensor network is deployed to monitor the system described in (1). At each step all the sensor readings are collected and 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 measurements from the sensors, and $y_{k,i}$ is the measurement made by sensor i at time k. $v_k \sim \mathcal{N}(0,\, R)$ is the measurement noise independent of x_0 and w_k.

A Kalman filter is used to compute state estimation $\hat{x}_k$ from observations y_ks:

$$\hat{x}_{0|-1} = 0,\, P_{0|-1} = \Sigma, \tag{3}$$

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_k + Bu_k,\, P_{k+1|k} = AP_kA^T + Q, \\ K_k &= P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}, \\ \hat{x}_k &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}), \\ P_k &= P_{k|k-1} - K_kCP_{k|k-1}. \end{aligned} \tag{4}$$

Although the Kalman filter uses a time varying gain K_k, it is well known that this gain will converge if the system is detectable. In practice the Kalman gain usually converges in a few steps. We can safely assume the Kalman filter to be already in steady state. Let us define

$$P \triangleq \lim_{k\to\infty} P_{k|k-1},\, K \triangleq PC^T(CPC^T + R)^{-1}. \tag{5}$$

The update equations of Kalman filter are as follows:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + K\left[y_{k+1} - C(A\hat{x}_k + Bu_k)\right], \tag{6}$$

For future analysis, let us define the residue z_{k+1} at time $k+1$ to be

$$z_{k+1} \triangleq y_{k+1} - C(A\hat{x}_k + Bu_k). \tag{7}$$

(6) can be simplified as

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + Kz_{k+1}. \tag{8}$$

The estimation error e_k at time k is defined as

$$e_k \triangleq x_k - \hat{x}_k. \tag{9}$$

Manipulating (6), (7), we get the following recursive equation:

$$e_{k+1} = (A - KCA)e_k + (I - KC)w_k - Kv_k. \tag{10}$$

2.3. LQG Controller

An LQG controller is used to stabilize the system by minimizing the following objective function[1]:

$$J = \lim_{T\to\infty} \min_{u_0,\ldots,u_T} E\frac{1}{T}\left[\sum_{k=0}^{T-1}(x_k^T W x_k + u_k^T U u_k)\right], \quad (11)$$

where W, U are positive semidefinite matrices and u_k is measurable with respect to $y_0,\ldots,y_k$, i.e. u_k is a function of previous observations. It is well known that the optimal controller of the above minimization problem is a fixed gain controller, which takes the following form:

$$u_k = -(B^T S B + U)^{-1} B^T S A \hat{x}_k, \quad (12)$$

where u_k is the optimal control input and S satisfies the following Riccati equation

$$S = A^T S A + W - A^T S B (B^T S B + U)^{-1} B^T S A. \quad (13)$$

Let us define $L \triangleq -(B^T S B + U)^{-1} B^T S A$, then $u_k = L x_{k|k}$.

The systems is stable if and only if $Cov(e_k)$ and J are both bounded. In particular that implies both matrices $A - KCA$ and $A + BL$ are stable. In the rest of the paper, we will only consider stable systems. Further, we assume to be already in steady state, which means $\{x_k, y_k, \hat{x}_k\}$are stationary random processes.

2.4. Failure Detector

A failure detector is often used in control system. For example, a χ^2 failure detector computes the following quantity

$$g_k = z_k^T \mathscr{P}^{-1} z_k, \quad (14)$$

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

Other types of failure detectors have also been considered by many researchers. In [11] [12], 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. Willsky et al. A generalized likelihood ratio test to detect dynamics or sensor jump is also proposed by Willsky et al. in [13].

[1] We assume an infinite horizon LQG controller is implemented.

To make the discussion more general, we assume the detector implemented in the CPS triggers an alarm based on following event:

$$g_k > threshold, \quad (15)$$

where g_k is defined as

$$g_k \triangleq g(z_k, y_k, \hat{x}_k, \ldots, z_{k-\mathscr{T}+1}, y_{k-\mathscr{T}+1}, \hat{x}_{k-\mathscr{T}+1}). \quad (16)$$

The function g is continuous and $\mathscr{T} \in \mathbb{N}$ is the window size of the detector. It is easy to see for χ^2 detector, $g_k = z_k^T \mathscr{P}^{-1} z_k$. We further define the probability of alarm for the failure detector to be

$$\beta_k = P(g_k > threshold). \quad (17)$$

At a first glance, it seems that certain choice of g function will affect detection differently. However, since the χ^2 detector along with many other detectors performs detection by computing a certain function of $\hat{x}_k, y_k, z_k$, then none of these detectors will be able to distinguish the healthy system from the partial compromised system if, under the malicious attack, the vectors $\hat{x}_k, y_k, z_k$ have the same statistical properties as those of healthy system. In Section 4, we show how the attacker can systematically attack the system without being noticed by the failure detector if a particular algebraic condition holds.

3. False Data Injection Attacks

In this section, we assume that a malicious third party wants to compromise the integrity of the system described in Section 2. The attacker is assumed to have the following capabilities:

1. It knows the system model: We assume that the attacker knows matrices A, B, C, Q, R as described in Section 2 and the observation gain and control gain K, L.

2. It can control the readings of a subset of the sensors, denoted by S_{bad}. As a result, (2) now becomes
$$y'_k = Cx'_k + v_k + \Gamma y^a_k, \quad (18)$$
where $\Gamma = diag(\gamma_1,\ldots,\gamma_m)$ is the sensor selection matrix. γ_i is a binary variable and $\gamma_i = 1$ if and only if $i \in S_{bad}$. y^a_k is the malicious input from the attacker. Here we write the observations and states as y'_k and x'_k since they are in general different from those of the healthy system due to the malicious attack.

3. The intrusion begins at time 0. As a result, the initial conditions for the partial compromised system will be $\hat{x}'_{-1} = 0$, $Ex_0 = 0$.

Figure 1 shows the diagram of the partial compromised system.

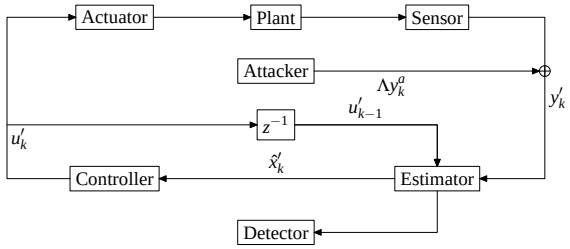

Figure 1. System Diagram

Definition 1. *An attack sequence $\mathscr{Y}$ is defined as an infinite sequence which takes the following form $y_0^a, y_1^a, \ldots$.*

It is easy to see that all the states of the partially compromised system are a function of $\mathscr{Y}$. For example, x_k' can be written as $x_k'(\mathscr{Y})$. However, in order to simplify the notation, we will use x_k' when there is no confusion. Under the previous assumptions, the new system dynamics can be written as

$$\begin{aligned} x_{k+1}' &= Ax_k' + Bu_k' + w_k, \\ y_k' &= Cx_k' + v_k + \Gamma y_k^a, \\ \hat{x}_{k+1}' &= A\hat{x}_k' + Bu_k' + K\left[y_{k+1}' - C(A\hat{x}_k' + Bu_k')\right], \\ u_k' &= L\hat{x}_k'. \end{aligned} \tag{19}$$

We can also define the new residue and estimation error respectively as

$$z_{k+1}' \triangleq y_{k+1}' - C(A\hat{x}_k' + Bu_k'),\ e_k' \triangleq x_k' - \hat{x}_k'. \tag{20}$$

Finally, the new probability of alarm is defined as

$$\beta_k' = P(g_k' > threshold), \tag{21}$$

where

$$g_k' \triangleq g(z_k', y_k', \hat{x}_k', \ldots, z_{k-\mathscr{T}+1}', y_{k-\mathscr{T}+1}', \hat{x}_{k-\mathscr{T}+1}'). \tag{22}$$

The differences between the two systems are defined as

$$\begin{aligned} \Delta x_k \triangleq\ & x_k' - x_k, \Delta\hat{x}_k \triangleq && \hat{x}_k' - \hat{x}_k, \\ \Delta u_k \triangleq\ & u_k' - u_k, \Delta y_k \triangleq && y_k' - y_k, \\ \Delta z_k \triangleq\ & z_k' - z_k, \Delta e_k \triangleq && e_k' - e_k, \Delta\beta_k = \beta_k' - \beta_k, \end{aligned} \tag{23}$$

where $x_k, \hat{x}_k, y_k, u_k, \beta_k$ are given by equations (1), (2), (6), (12), (17). $\Delta x_k, \Delta\hat{x}_k, \Delta u_k, \Delta y_k, \Delta z_k, \Delta e_k, \Delta\beta_k$ represent the differences between the partially compromised system and the healthy system.

The following definition defines what constitutes a "successful" attack.

Definition 2. *An attack sequence $\mathscr{Y}$ is (ε,α)-successful if there exists $T \in \mathbb{N}$, such that the following holds:*

$$\|\Delta x_T(\mathscr{Y})\| \geq \alpha, \Delta\beta_k(\mathscr{Y}) \leq \varepsilon, \forall k = 0, 1, \ldots, T-1.$$

The system is called (ε,α)-attackable if there exists a (ε,α)-successful attack sequence $\mathscr{Y}$ on the CPS.

Remark 1. *It is worth noticing that simply injecting a large y_k^a will result in a large Δz_k which, in turn, will induce the failure detector to trigger an alarm immediately.*

Although the definition of (ε,α)-attackable is simple, it is not so easy to verify whether a system is (ε,α)-attackable, especially when the form of g is complex. As a result, we will consider a limit case of (ε,α)-attackability.

Definition 3. *A control system is perfectly attackable if there exists an attack sequence $\mathscr{Y}$ such that the following holds:*

$$\limsup_{k\to\infty} \|\Delta x_k(\mathscr{Y})\| = \infty, \|\Delta z_k(\mathscr{Y})\| \leq 1, \forall k = 0, 1, \ldots, .$$

The next theorem shows that perfect attackability implies (ε,α)-attackability.

Theorem 1. *If a control system is perfectly attackable, then it is also (ε,α)-attackable for any $\varepsilon,\alpha > 0$,*

Proof. Since the system is perfectly attackable, there exists an attack sequence $\mathscr{Y}$, such that

$$\limsup_{k\to\infty} \|\Delta x_k(\mathscr{Y})\| = \infty, \|\Delta z_k(\mathscr{Y})\| \leq 1, k = 0, 1, \ldots \tag{24}$$

Manipulating equations (6) (12) (19), we can prove that:

$$\begin{aligned} \Delta\hat{x}_{k+1} &= (A+BL)\Delta\hat{x}_k + K\Delta z_{k+1}, \\ \Delta y_{k+1} &= \Delta z_{k+1} + C(A+BL)\Delta\hat{x}_k. \end{aligned} \tag{25}$$

Stability of $A+BL$ is guaranteed by the stability of the original system. Therefore, if $\|\Delta z_k(\mathscr{Y})\| \leq 1$ for all $k = 0, 1, \ldots$, then $\Delta\hat{x}_k(\mathscr{Y})$ and $\Delta y_k(\mathscr{Y})$ will be uniformly bounded for all k. Define the bounds to be

$$M_1 = \sup_k \|\Delta\hat{x}_k(\mathscr{Y})\|, M_2 = \sup_k \|\Delta y_k(\mathscr{Y})\|, \tag{26}$$

where $M_1, M_2 < \infty$ are constants. Due to the continuity of g, there exists $\varepsilon' > 0$ such that if $\|\Delta z_k\| \leq \varepsilon'$, $\|\Delta\hat{x}_k\| \leq \varepsilon'$, $\|\Delta y_k\| \leq \varepsilon'$, then

$$|P(g_k' > threshold) - P(g_k > threshold)| \leq \varepsilon,$$

Since $\Delta z_k(\mathscr{Y}), \Delta\hat{x}_k(\mathscr{Y}), \Delta y_k(\mathscr{Y})$ are uniformly bounded, by linearity, we can find $\delta > 0$, such that

$$\|\Delta z_k(\delta\mathscr{Y})\| \le \varepsilon', \|\Delta\hat{x}_k(\delta\mathscr{Y})\| \le \varepsilon', \|\Delta y_k(\delta\mathscr{Y})\| \le \varepsilon', \forall k.$$

By the stationarity of the random process $\{x_k, y_k, \hat{x}_k\}$, we know that

$$|P(g'_k > threshold) - P(g_k > threshold)| \le \varepsilon, \forall k.$$

Finally by linearity,

$$\limsup_{k\to\infty} \Delta x_k(\delta\mathscr{Y}) = \delta \limsup_{k\to\infty} \Delta x_k(\mathscr{Y}) = \infty.$$

Hence, $\delta\mathscr{Y}$ is an (ε,α)-successful attack sequence and the system is (ε,α)-attackable. □

In the next section, we will give a necessary and sufficient condition for a system to be perfectly attackable.

4. Main Result

In this section, we will provide an algebraic condition to identify perfectly attackable system, which is given by the following theorem:

Theorem 2. *The control system (1) is perfectly attackable if and only if A has an unstable eigenvalue and the corresponding eigenvector v satisfies:*

1. $Cv \in span(\Gamma)$, *where* $span(\Gamma)$ *is the column space of* Γ.
2. *v is a reachable state of the dynamic system* $\Delta e_{k+1} = (A - KCA)\Delta e_k - K\Gamma y^a_{k+1}$.

Before proving the theorem, we need the following lemmas:

Lemma 1. *The CPS is perfectly attackable if and only if there exists an attack sequence* $\mathscr{Y}$ *such that*

$$\limsup_{k\to\infty} \|\Delta e_k\| = \infty, \|\Delta z_k\| \le 1, k = -1, 0, \ldots \quad (27)$$

Proof. The proof follows from the boundedness of $\Delta\hat{x}_k$ and the fact that $\Delta x_k = \Delta\hat{x}_k + \Delta e_k$. Due to space limitation the complete proof will be omitted. □

Using Lemma 1, we can use Δe_k to prove that the system is perfectly attackable. The main advantages of substituting Δx_k with Δe_k is that Δe_k follows a simpler recursive equation:

$$\Delta e_{k+1} = (A - KCA)\Delta e_k - K\Gamma y^a_{k+1}. \quad (28)$$

Moreover,

$$\Delta z_{k+1} = CA\Delta e_k + \Gamma y^a_{k+1}. \quad (29)$$

Before proving Theorem 2, we need an additional lemma:

Lemma 2. *Let* $p \in \mathbb{R}^n$ *be a vector, and* $\lim_{k\to\infty} A^k p \neq 0$, *then there exists an unstable eigenvector v of matrix A, such that* $p \in span(p, A^2 p, \ldots, A^{n-1} p)$.

The proof is based on the Jordan decomposition of the A matrix and on Carley-Hamilton Theorem. The complete proof is omitted due to space limits. Now we are ready to prove Theorem 2.

Proof of Theorem 2. First we will prove the necessity. Suppose that CPS is perfectly attackable, then by Lemma 1, there exists an successful attack sequence $\mathscr{Y}$ such that

$$\limsup_{k\to\infty} \|\Delta e_k\| = \infty, \|\Delta z_k\| \le 1, k = 0, 1, \ldots.$$

A peak subsequence $\{\Delta e_{i_k}\}$ from Δe_i is defined as

$$\Delta e_{i_0} = \Delta e_0, \Delta e_{i_k} = \min\{j : \|\Delta e_j\| > \|\Delta e_{i_{k-1}}\|\}, \quad (30)$$

which means that the norm $\|\Delta e_{i_k}\|$ is larger than the norm of any preceding term in the original sequence. Since Δe_k is unbounded, there always exists such a subsequence and $\lim_{k\to\infty} \Delta e_{i_k} = \infty$. Now consider the normalized vectors defined as

$$p_k \triangleq \frac{1}{\|\Delta e_k\|} \Delta e_k. \quad (31)$$

It is trivial to see $\|p_k\|$ is bounded. As a result, there exists an index set $\{j_k\} \subset \{i_k\}$ such that all of the subsequences $\{p_{j_k}\}, \{p_{j_k-1}\}, \ldots, \{p_{j_k-n+1}\}$ converge as k goes to infinity, due to Bolzano-Weierstrass theorem. Let us define

$$q_l \triangleq \lim_{k\to\infty} p_{j_k-l}, l = 0, 1, \ldots, n-1. \quad (32)$$

In addition, since

$$\|\Delta e_{k+1}\| = \|A\Delta e_k - K\Delta z_{k+1}\| \le \|A\|\|\Delta e_k\| + \|K\|,$$

and Δe_{j_k} is unbounded, $\lim_{k\to\infty} \Delta e_{j_k-l} = \infty$ for all l from 0 to $n-1$. As a result

$$\lim_{k\to\infty} \frac{\Delta e_{j_k}}{\|\Delta e_{j_k-1}\|} = \lim_{k\to\infty} \frac{A\Delta e_{j_k-1} - K\Delta z_{j_k}}{\|\Delta e_{j_k-1}\|}$$
$$= A \lim_{k\to\infty} \frac{\Delta e_{j_k-1}}{\|\Delta e_{j_k-1}\|} = Aq_1.$$

Therefore

$$q_0 = \lim_{k\to\infty} \frac{\|\Delta e_{j_k-1}\|}{\|\Delta e_{j_k}\|} \lim_{k\to\infty} \frac{\Delta e_{j_k}}{\|\Delta e_{j_k-1}\|} = Aq_1/\|Aq_1\|.$$

Similarly, it is easy to show that $q_l = Aq_{l+1}/\|Aq_{l+1}\|$. Hence,

$$span(q_0, \ldots, q_{n-1}) = span(A^{n-1}q_{n-1}, \ldots, Aq_{n-1}, q_{n-1}).$$

By definition of $\{\Delta e_{i_k}\}$, $\|\Delta e_{j_k}\| \geq \|\Delta e_{j_k-1}\|$. Thus, $\|Aq_1\| \geq \|q_1\|$, which implies that $\lim_{k\to\infty} A^k q_{n-1} \neq 0$. From Lemma 2 it follows that there exists an unstable eigenvector v in the span of $q_0,\ldots,q_{n-1}$. Since

$$\|\frac{\Delta z_{j_k+1}}{\|\Delta e_{j_k}\|}\| = \|Cp_{j_k} + \Gamma\frac{y^a_{j_k+1}}{\|\Delta e_{j_k}\|}\| \leq \frac{1}{\|\Delta e_{j_k}\|},$$

$$Cp_{j_k} \in span(\Gamma) + B(0,(\|\Delta e_{j_k}\|)^{-1}),$$

where $B(0,(\|\Delta e_{j_k}\|)^{-1})$ is a ball center at 0 with radius $(\|\Delta e_{j_k}\|)^{-1}$. As a result

$$Cq_0 \in \bigcap_{l=1}^{\infty}\left[span(\Gamma) + B(0,(\|\Delta e_{j_k}\|)^{-1})\right] = span(\Gamma).$$

Similarly, CAq_l belongs to $span(\Gamma)$ for all l from 0 to $n-1$. As a result, $CAv \in span(CAq_0,\ldots,CAq_{n-1}) \subset span(\Gamma)$, which implies $Cv \in span(\Gamma)$.

For reachability, since Δe_k is reachable, $\alpha\Delta e_k$ is reachable for any $\alpha \in \mathbb{R}$. In particular, p_k is reachable for all k. Since the reachable subspace is closed, the limit q_l is reachable, which implies v is reachable, thus proving the necessary condition.

We now want to prove sufficiency. Since $Cv \in span(\Gamma)$, there exists y^* such that $\Gamma y^* = Cv$. Furthermore, since v is reachable, there exist $y^a_0,\ldots,y^a_{n-1}$, where n is the dimension of state space, such that $\Delta e_{n-1} = v$. Define

$$M = \max_{k=0,\ldots,n-1}\|\Delta z_k\|. \tag{33}$$

By linearity, if the attacker injects $y^a_0/M,\ldots,y^a_{n-1}/M$, then $\Delta e_{n-1} = v/M$ and $\|\Delta z_k\| \leq 1$ for $k = 0,\ldots,n-1$. As a result, the attacker could choose the attack sequence to be

$$y^a_{n+i} = y^a_i - \frac{\lambda^{i+1}}{M}y^*, i = 0,1,\ldots \tag{34}$$

One can prove that with the above attack sequence, the following equality and inequality hold :

$$\Delta e_{n+i} = \Delta e_i + \frac{\lambda^{i+1}}{M}v, i = 0,1,\ldots, \tag{35}$$

$$\|\Delta z_{n+i}\| = \|\Delta z_i\| \leq 1, i = 0,1,\ldots \tag{36}$$

Since $|\lambda| \geq 1$, $\Delta e_k \to \infty$, which implies that the system is perfectly attackable. □

Remark 2. *The attacker could use the results of Theorem 2 to design an attack sequence $\mathscr{Y}$ based on the eigendecomposition of A and the Γ matrix.*

On the other hand, the defender could also perform an eigendecomposition on A matrix, find all the unstable eigenvector v and then compute Cv. For each Cv, the non-zero elements will indicate the sensors needed by the attacker to perform a successful attack along direction v. Therefore if Cv is a sparse vector, an attacker could initiate an attack on the direction of v by compromising only a few sensors. As a result, the defender could increase the resilience of the system by installing redundant sensors to measure mode v.

5. Illustrative Examples

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

Consider a vehicle moving along the x-axis. The state space includes position x and velocity $\dot{x}$ of the vehicle. An actuator is used to control the speed of the vehicle. As a result, the system dynamics is as follows:

$$\begin{aligned}\dot{x}_{k+1} &= \dot{x}_k + u_k + w_{k,1},\\ x_{k+1} &= x_k + (\dot{x}_{k+1} + \dot{x}_k)/2 + w_{k,2}\\ &= x_k + \dot{x}_k + u_k/2 + w_{k,1}/2 + w_{k,2},\end{aligned} \tag{37}$$

which can be written in the matrix form as

$$X_{k+1} = \begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}X_k + \begin{bmatrix}1\\ 0.5\end{bmatrix}u_k + w_k, \tag{38}$$

where

$$X_k = \begin{bmatrix}\dot{x}\\ x\end{bmatrix}, w_k = \begin{bmatrix}w_{k,1}\\ w_{k,2} + 0.5w_{k,1}\end{bmatrix}. \tag{39}$$

Suppose two sensors are measuring the velocity and position respectively. Hence

$$y_k = X_k + v_k. \tag{40}$$

We assume the position sensor is compromised, i.e. $\Gamma = diag(0,1)$. We further impose the following parameters on the system

$$Q = R = W = I_2, U = 1.$$

The steady state Kalman gain and the LQG control gain under the previous assumptions are respectively

$$K = \begin{bmatrix}0.5939 & 0.0793\\ 0.0793 & 0.6944\end{bmatrix}, L = \begin{bmatrix}-1.0285 & -0.4345\end{bmatrix}.$$

Since $[01]'$ is an unstable eigenvector and is in the span of Γ and reachable, by Theorem 2, the system is perfectly attackable. Using the result we derived in Section 4, we design the attack sequence $\mathscr{Y}$ to be

$$\begin{aligned}&y^a_0 = [0,-1.000]', y^a_1 = [0,-0.367]',\\ &y^a_k = y^a_{k-2} - [0,-0.485]', k \geq 2.\end{aligned} \tag{41}$$

Figure 2 shows the evolution of the ΔX_k and Δz_k. It is easy to see that $\|\Delta z_k\|$ is always less than 1 and Δx_k goes to infinity, showing that the system is perfectly attackable.

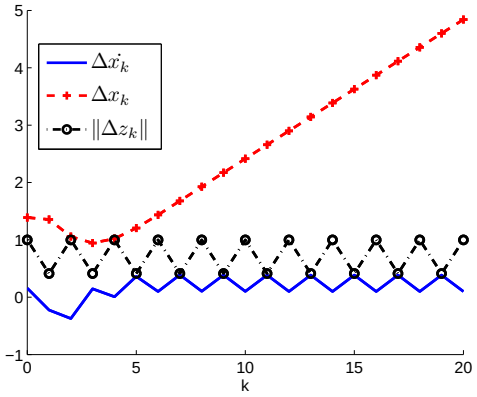

Figure 2. Evolution of $\Delta\dot{x}_k$, Δx_k, $\|\Delta z_k\|$

6. Conclusion and Future Work

This paper proposes a false data injection attack model and analyze the effects of such kind of attacks on a linear time-invariant Gaussian control system. We prove the existence of a necessary and sufficient condition under which the attack could destabilize the system while successfully bypassing a large set of possible failure detectors. We also provide a design criterion to improve the resilience of the system to false data injection attacks.

Future work will be directed toward deriving conditions under which the system is (ε, α)-attackable. We also plan to combine both the false data injection attacks and DoS attacks and study their effects on control systems.

References

[1] E. A. Lee, "Cyber physical systems: Design challenges," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2008-8, Jan 2008. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-8.html

[2] 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.

[3] 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*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–6.

[4] ——, "Secure control: Towards survivable cyber-physical systems," in *Distributed Computing Systems Workshops, 2008. ICDCS '08. 28th International Conference on*, June 2008, pp. 495–500.

[5] 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*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, April 2009, pp. 31–45. [Online]. Available: http://chess.eecs.berkeley.edu/pubs/597.html

[6] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, Jan. 2007.

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

[8] R. Stengel and L. Ryan, "Stochastic robustness of linear time-invariant control systems," *Automatic Control, IEEE Transactions on*, vol. 36, no. 1, pp. 82–87, Jan 1991.

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

[10] 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.

[11] R. V. Beard, "Failure accommodation in linear systems through self-reorganization," Man Vehicle Laborotory, Cambrideg, Massachusetts, Tech. Rep. MVT-71-1, February 1971.

[12] H. L. Jones, "Failure detection in linear systems," Ph.D. dissertation, M.I.T., Cambridge, Massachusetts, 1973.

[13] 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*, vol. 21, pp. 108–112, Feburary 1976.