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 wish 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...
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:

\[ x_{k+1} = Ax_k + Bu_k + w_k, \]

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

\[ y_k = Cx_k + v_k, \]

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 \( y_k \):

\[
\hat{x}_{k|k-1} = 0, \quad P_{0|0} = \Sigma, \\
\hat{x}_{k+1|k} = A\hat{x}_k + Bu_k, \quad P_{k+1|k} = AP_kA^T + Q, \\
K_k = P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}, \\
\hat{x}_k = \hat{x}_{k|k} + K_k(y_k - C\hat{x}_{k|k-1}), \\
P_k = P_{k+1|k} - K_kCP_{k+1|k}.
\]

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

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

The update equations of Kalman filter are as follows:

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

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

\[ z_{k+1} \triangleq y_{k+1} - C(A\hat{x}_k + Bu_k). \]

(6) can be simplified as

\[ \hat{x}_{k+1} = A\hat{x}_k + Bu_k + Kz_{k+1}. \]

The estimation error \( e_k \) at time \( k \) is defined as

\[ e_k \triangleq x_k - \hat{x}_k. \]

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

\[ e_{k+1} = (A - KCA)e_k + (I - KC)w_k - Kv_k. \]
2.3. LQG Controller

An LQG controller is used to stabilize the system by minimizing the following objective function:\footnote{We assume an infinite horizon LQG controller is implemented.}

\[
J = \lim_{T \to \infty} \min U \frac{1}{T} \left[ \sum_{k=0}^{T-1} \left( x_k^T W x_k + u_k^T U u_k \right) \right],
\]

(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:

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

(12)

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

\[
S = A^T S A + W - A^T S B (B^T S B + U)^{-1} B^T S A.
\]

(13)

Let us define \( L \triangleq -(B^T S B + U)^{-1}B^T S A \), then \( u_k = L \hat{x}_k \).

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 \( \chi^2 \) failure detector computes the following quantity

\[
g_k \triangleq z_k^T \mathcal{P}^{-1} z_k,
\]

(14)

where \( \mathcal{P} \) is the covariance matrix of the residue \( z_k \). Since \( z_k \) is Gaussian distributed, \( g_k \) is \( \chi^2 \) distributed with \( m \) degrees of freedom. As a result, \( g_k \) cannot be far away from 0. The \( \chi^2 \) 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].

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

\[
g_k > \text{threshold},
\]

(15)

where \( g_k \) is defined as

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

(16)

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

\[
b_k = P(g_k > \text{threshold}).
\]

(17)

At a first glance, it seems that certain choice of \( g \) function will affect detection differently. However, since the \( \chi^2 \) 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, \( y_k \) now becomes

\[
y_k' = C x_k' + v_k + \Gamma y_k' \triangleq y_k^e,
\]

(18)

where \( \Gamma = diag(\gamma_1, \ldots, \gamma_m) \) is the sensor selection matrix. \( \gamma_i \) is a binary variable and \( \gamma_i = 1 \) if and only if \( i \in S_{bad} \). \( y_k^e \) is the malicious input from the attacker. Here we write the observations and states as \( y_k^e \) 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 \( x_{-1}' = 0, E x_0 = 0. \)
Figure 1 shows the diagram of the partial compromised system.

![Diagram](attachment:image.png)

**Figure 1. System Diagram**

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

It is easy to see that all the states of the partially compromised system are a function of $\mathcal{Y}$. For example, $x'_k$ can be written as $x'_k(\mathcal{Y})$. 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

$$
\begin{align*}
 x'_{k+1} &= A x'_k + B u'_k + w_k, \\
 y'_k &= C x'_k + v_k + \Gamma y'_k, \\
 \hat{x}'_{k+1} &= A \hat{x}'_k + B u'_k + K [y'_k + C (A x'_k + Bu'_k)] , \\
 u'_k &= L \hat{x}'_k.
\end{align*}
$$

(19)

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

$$
\begin{align*}
 z'_k &= y'_k - C(A x'_k + Bu'_k), \\
 e'_k &= x'_k - \hat{x}'_k.
\end{align*}
$$

(20)

Finally, the new probability of alarm is defined as

$$
P(g'_k > \text{threshold}),
$$

(21)

where

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

(22)

The differences between the two systems are defined as

$$
\begin{align*}
 \Delta x_k &= x'_k - x_k, \\
 \Delta \hat{x}_k &= \hat{x}'_k - \hat{x}_k, \\
 \Delta u_k &= u'_k - u_k, \\
 \Delta y_k &= y'_k - y_k, \\
 \Delta z_k &= z'_k - z_k, \\
 \Delta e_k &= e'_k - e_k, \\
 \Delta \beta_k &= \beta'_k - \beta_k.
\end{align*}
$$

(23)

where $x_k, \hat{x}_k, u_k, \beta_k$ 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 $\mathcal{Y}$ is $(\varepsilon, \alpha)$-successful if there exists $T \in \mathbb{N}$, such that the following holds:

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

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

**Remark 1.** It is worth noticing that simply injecting a large $y'_k$ will result in a large $\Delta y_k$ which, in turn, will induce the failure detector to trigger an alarm immediately.

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

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

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

(24)

The next theorem shows that perfect attackability implies $(\varepsilon, \alpha)$-attackability.

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

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

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

(25)

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

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

(26)

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

$$
\begin{align*}
 M_1 &= \sup_{k} \|\Delta \hat{x}_k(\mathcal{Y})\|, \\
 M_2 &= \sup_{k} \|\Delta y_k(\mathcal{Y})\|,
\end{align*}
$$

(27)

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

$$
|P(g'_k > \text{threshold}) - P(g_k > \text{threshold})| \leq \varepsilon,
$$

(28)
Since $\Delta z_k(\mathcal{Y}), \Delta y_k(\mathcal{Y}), \Delta y_k(\mathcal{Y})$ are uniformly bounded, by linearity, we can find $\delta > 0$, such that
\[ \|\Delta z_k(\mathcal{Y})\| \leq \epsilon', \|\Delta y_k(\mathcal{Y})\| \leq \epsilon', \|\Delta y_k(\mathcal{Y})\| \leq \epsilon', \forall k. \]
By the stationarity of the random process $\{x_k, y_k, \dot{x}_k\}$, we know that
\[ |P(g_k > \text{threshold}) - P(g_k > \text{threshold})| \leq \epsilon, \forall k. \]
Finally, by linearity,
\[ \limsup_{k \to \infty} \Delta x_k(\mathcal{Y}) = \delta \limsup_{k \to \infty} \Delta x_k(\mathcal{Y}) = \infty. \]
Hence, $\mathcal{Y}$ is an $(\epsilon, \alpha)$-successful attack sequence and the system is $(\epsilon, \alpha)$-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. $v \in \text{span}(\Gamma)$, where $\text{span}(\Gamma)$ is the column space of $\Gamma$.
2. $v$ is a reachable state of the dynamic system
\[ \Delta e_{k+1} = (A - KCA)\Delta e_k - KT\gamma_k. \]

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 $\mathcal{Y}$ such that
\[ \limsup_{k \to \infty} \|\Delta e_k\| = \infty, \|\Delta z_k\| \leq 1, k = 0, 1, \ldots \]  (27)

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

Using Lemma 1, we can use $\Delta e_k$ to prove that the system is perfectly attackable. The main advantages of substituting $\Delta x_k$ with $\Delta e_k$ is that $\Delta e_k$ follows a simpler recursive equation:
\[ \Delta e_{k+1} = (A - KCA)\Delta e_k - KT\gamma_k. \]  (28)

Moreover, $\Delta z_{k+1} = C\Delta e_k + \Gamma y_k$.

Before proving Theorem 2, we need an additional lemma:

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

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 $\mathcal{Y}$ such that
\[ \limsup_{k \to \infty} \|\Delta e_k\| = \infty, \|\Delta z_k\| \leq 1, k = 0, 1, \ldots \]

A peak subsequence $\{\Delta e_{i_k}\}$ from $\Delta e_i$ 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 $\Delta e_k$ 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
\[ p_k \triangleq \frac{1}{\|\Delta e_k\|} \Delta e_k. \]  (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 $\{p_{j_k}\}, \{p_{j_k-1}\}, \ldots, \{p_{j_1}\}$ converge as $k$ goes to infinity, due to Bolzano-Weierstrass theorem. Let us define
\[ q_l \triangleq \lim_{k \to \infty} p_{j_k-l}, l = 0, 1, \ldots, n - 1. \]  (32)

In addition, since
\[ \|\Delta e_{k+1}\| = \|A\Delta e_k - K\Delta z_k\| \leq \|A\|\|\Delta e_k\| + \|K\|, \]
and $\Delta e_{j_k}$ is unbounded, $\lim_{k \to \infty} \Delta e_{j_k-l} = \infty$ for all $l$ from 0 to $n - 1$. As a result
\[ \lim_{k \to \infty} \frac{\Delta e_{j_k}}{\|\Delta e_{j_k-l}\|} = \lim_{k \to \infty} \frac{A\Delta e_{j_k-l} - K\Delta z_{j_k}}{\|\Delta e_{j_k-l}\|} = Aq_1. \]
Therefore
\[ q_0 = \lim_{k \to \infty} \frac{\Delta e_{j_k-1}}{\|\Delta e_{j_k-1}\|} = \frac{\Delta e_{j_k}}{\|\Delta e_{j_k-1}\|} = Aq_1/\|Aq_1\|. \]

Similarly, it is easy to show that $q_l = Aq_{l+1}/\|Aq_{l+1}\|$. Hence,
\[ \text{span}(q_0, \ldots, q_{n-1}) = \text{span}(A^{n-1}q_{n-1}, \ldots, Aq_{n-1}, q_{n-1}). \]
By definition of \( \{ \Delta e_k \} \), \( \|\Delta e_k\| \geq \|\Delta e_{k-1}\| \). Thus, \( \|Aq_i\| \geq \|q_i\| \), 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
\[
\left\| \frac{\Delta e_{k+1}}{\|\Delta e_k\|} \right\| = \| Cp_{k+1} + \frac{y_{k+1}}{\|\Delta e_k\|} \| \leq \frac{1}{\|\Delta e_k\|},
\]
\( C p_{k+1} \in \text{span} (\Gamma) + B (0, (\|\Delta e_k\|)^{-1}) \), where \( B (0, (\|\Delta e_k\|)^{-1}) \) is a ball center at 0 with radius \((\|\Delta e_k\|)^{-1}\). As a result
\[
C q_0 \in \bigcap_{l=1}^{\infty} \left[ \text{span} (\Gamma) + B (0, (\|\Delta e_k\|)^{-1}) \right] = \text{span} (\Gamma).
\]

Similarly, \( C A q_l \) belongs to \( \text{span} (\Gamma) \) for all \( l \) from 0 to \( n-1 \). As a result, \( C A v \in \text{span} (C A q_0, \ldots, C A q_{n-1}) \subset \text{span} (\Gamma) \), which implies \( C v \in \text{span} (\Gamma) \).

For reachability, since \( \Delta e_k \) is reachable, \( \Delta e_k \) 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 \( q_i \) is reachable, which implies \( v \) is reachable, thus proving the necessary condition.

We now want to prove sufficiency. Since \( C v \in \text{span} (\Gamma) \), there exists \( y^* \) such that \( \Gamma y^* = C v \). Furthermore, since \( v \) is reachable, there exist \( y_0^*, \ldots, y_{n-1}^* \), where \( n \) is the dimension of state space, such that \( \Delta e_{n-1} = v \). Define
\[
M = \max_{k=0, \ldots, n-1} \| \Delta z_k \|. \quad (33)
\]

By linearity, if the attacker injects \( y_0^*/M, \ldots, y_{n-1}^*/M \), 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
\[
y_{n+i}^* = y_i^* - \frac{\lambda^{i+1}}{M} y^*, \quad i = 0, 1, \ldots \quad (34)
\]
One can prove that with the above attack sequence, the following equality and inequality hold:
\[
\Delta e_{n+i} = \Delta e_i + \frac{\lambda^{i+1}}{M} v, \quad i = 0, 1, \ldots \quad (35)
\]
\[
\|\Delta z_{n+i}\| = \|\Delta z_i\| \leq 1, \quad i = 0, 1, \ldots \quad (36)
\]
Since \( |\lambda| \geq 1 \), \( \Delta e_k \to \infty \), which implies that the system is perfectly attackable. \( \square \)

**Remark 2.** The attacker could use the results of Theorem 2 to design an attack sequence \( \mathcal{A} \) based on the eigendecomposition of \( A \) and the \( \Gamma \) matrix.

On the other hand, the defender could also perform an eigendecomposition on \( A \) matrix, find all the unstable eigenvector \( v \) and then compute \( C v \). For each \( C v \), the non-zero elements will indicate the sensors needed by the attacker to perform a successful attack along direction \( v \). Therefore if \( C v \) 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:
\[
\dot{x}_k+1 = \dot{x}_k + u_k + w_k, \quad x_k+1 = x_k + (\dot{x}_k+1 + \dot{x}_k)/2 + w_k, \quad (37)
\]
which can be written in the matrix form as
\[
X_{k+1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} X_k + \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} u_k + w_k, \quad (38)
\]
where
\[
X_k = \begin{bmatrix} \dot{x}_k \\ x_k \end{bmatrix}, \quad w_k = \begin{bmatrix} w_{k,1} \\ w_{k,2} + 0.5 w_{k,1} \end{bmatrix}. \quad (39)
\]
Suppose two sensors are measuring the velocity and position respectively. Hence
\[
y_k = X_k + v_k. \quad (40)
\]
We assume the position sensor is compromised, i.e. \( \Gamma = \text{diag}(0,1) \). We further impose the following parameters on the system
\[
Q = R = W = I_2, \quad U = 1.
\]
The steady state Kalman gain and the LQG control gain under the previous assumptions are respectively
\[
K = \begin{bmatrix} 0.5939 & 0.0793 \\ 0.0793 & 0.6944 \end{bmatrix}, \quad L = \begin{bmatrix} -1.0285 \\ -0.4345 \end{bmatrix}.
\]
Since \( [01]' \) is an unstable eigenvector and is in the span of \( \Gamma \) and reachable, by Theorem 2, the system is perfectly attackable. Using the result we derived in Section 4, we design the attack sequence \( \mathcal{A} \) to be
\[
y_0^* = [0, -1.0000]', \quad y_1^* = [0, -0.367]', \quad (41)
\]
\[
y_1^* = y_{k-2}' - [0, -0.485]', \quad k \geq 2.
\]
Figure 2 shows the evolution of the \( \Delta x_k \) and \( \Delta z_k \). It is easy to see that \( \|\Delta z_k\| \) is always less than 1 and \( \Delta z_k \) goes to infinity, showing that the system is perfectly attackable.
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 $(\epsilon, \alpha)$-attackable. We also plan to combine both the false data injection attacks and DoS attacks and study their effects on control systems.

References


