AIAA 2013-4653

Guidance, Navigation, and Control and Co-located Conferences
August 19-22, 2013, Boston, MA
AIAA Infotech@Aerospace (I@A) Conference

Analysis of Optimal False Data Injection Attacks in Unmanned Aerial Systems

Andrew Shull [*] and Inseok Hwang [†]

Purdue University, West Lafayette, Indiana, 47907, United States

With the increasing power and convenience offered by the use of embedded systems in control applications, such systems will undoubtedly continue to be developed and deployed. Recently, however, a focus on data-centric systems and developing network-enabled control systems has emerged, allowing for greater performance, safety, and resource allocation in systems such as unmanned aerial systems (UASs). However, this increase in connectivity also introduces vulnerabilities into these systems, potentially providing access to malicious parties seeking to disrupt the operation of those systems or to cause damage. In this analysis, stealthy false data injection attacks against linear feedback systems of that are commonly used in UASs are considered. The identification of these attacks is formed as an optimization problem constrained by the ability of monitoring systems to detect the attack. The optimal attack input is then determined for an example application so that the worst case system performance can be identified and, if needed, improved.

Nomenclature

$\mathbf{A}$	System evolution matrix	$\mathbf{B}$	System input matrix
$\mathbf{B}_{\mathrm{d}}$	System disturbance input matrix	$\mathbf{C}$	System output matrix
$\mathbf{D}$	Observation noise input matrix	$\mathbf{F}$	Attack availability matrix
$\mathbf{L}[k]$	Kalman filter gain	$\mathbf{L}$	Steady-state Kalman filter gain
$\mathbf{K}$	Feedback control gain	$\mathbf{W}$	Error Weighting matrix
$\mathbf{Q}[k]$	Estimate error covariance	$\mathbf{Q}_{\infty}$	Steady-state Kalman filter covariance
$\mathbf{Q}_{\mathrm{r}}$	Nominal residual covariance	$\mathbf{Q}_{\mathrm{a}}[k]$	Attacked estimate error covariance
$\mathbf{x}[k]$	System state	$\hat{\mathbf{x}}[k]$	State estimate
$\mathbf{e}[k]$	State estimate error	$\mathbf{r}[k]$	Measurement residual vector
$g[k]$	Fault detection test signal	$\bar{g}[k]$	Expected fault detection test signal
$\mathbf{w}[k]$	Process noise	$\mathbf{v}[k]$	Observation noise
$\mathbf{a}[k]$	Attack input vector	$\bar{\mathbf{e}}[k]$	Expected state estimate error
$\Delta\mathbf{x}[k]$	Nominal and attacked state difference	$\Delta\hat{\mathbf{x}}[k]$	Nominal and attacked estimate difference
$\mathbf{I}_n$	$n \times n$ Identity matrix	$\mathbf{0}_n$	$n \times n$ matrix with elements equal to 0
$\mathbf{0}_{n,m}$	$n \times m$ matrix with elements equal to 0	$\mathbf{X}$	Optimization variable
$\mathbf{W}, \mathbf{W}'$	State error weighting matrices	$\mathbf{S}$	Optimization variable index selection matrix
$\mathbf{\Gamma}$	Optimization affine constraint matrix	$\mathbf{\Lambda}_k$	Residual constraint matrix at time k
$\theta(\mathbf{X})$	Optimization objective function	$\mathbf{\Psi}(\mathbf{X})$	Residual constraint function
$\lambda,\ \mu$	Lagrange multipliers	$g(\lambda,\ \mu)$	Dual problem function
Subscripts			
a	Attacked value	n	nominal value
Operators			
$\mathrm{sgn}(*)$	Sign function	$\mathrm{diag}(*)$	Diagonal matrix
$\mathrm{blkdiag}(*)$	Block diagonal matrix	$\mathrm{blktoeplitz}(*,*)$	Block toeplitz matrix
$\mathcal{N}(\mu,\sigma^2)$	Gaussian distribution		

[*] Systems Engineer, Raytheon Missile Systems, andrew.shull@raytheon.com, AIAA Member.

[†] Associate Professor, Aeronautics and Astronautics, ihwang@purdue.edu, AIAA Associate Fellow.

I. Introduction

A. Problem Definition

A traditional computer science approach to cybersecurity focuses on techniques such as restricting system access and encrypting sensitive data.[1] While this is a key component to ensuring the security of cyber-physical systems, it only protects the computing resources of the system. It is important that in addition to this, the overall system is designed so as to minimize the risk should these traditional security approaches fail to restrict access to the system. To that end, the problem of system cybersecurity is considered here from a complimentary control systems perspective.

This analysis investigates false data injection attacks on linear Gaussian systems such as those commonly found in unmanned aerial systems (UASs) in components such as GPS, ADS-B, ground station observation systems, and formation control.[2,3,4,5,6,7,8,9,10] A false data injection attack is an attack in which the attacker is able to arbitrarily change some or all of the measurements made by the system. This type of attack has been identified as particularly interesting in part because of the comparative ease with which it can be performed. An attacker may not be able to directly manipulate system configurations such as controller gains or state estimates due to access restrictions, memory protection, or other successful computer security policies. Measurement data, however, is often gathered from an external source and may not be as easily protected. This is especially true for remote sensors which report measurements over a data network. This configuration lends itself to network-based attacks, direct physical manipulation of a remote sensor, or manipulation of the environment to change the measurement. Given the noise inherent in sensor measurements, it can also be difficult to distinguish valid measurements from modified measurements.

The formation of optimal false data attacks on linear systems that use Kalman filter state estimates to implement feedback is investigated. This is done by formulating an optimization problem that seeks to determine the attack input that maximizes the error of the system state relative to the nominal state trajectory. This optimal attack input is constrained by the requirement that the attack should not cause monitoring systems to identify a fault condition, as doing so would alert the user to the presence of the attack and permit the implementation of a mitigation policy.

In this analysis, the focus is on identifying attacks which are able to either degrade system performance or introduce failure while attempting to avoid being detected by mitigation systems. Detection permits the implementation of a mitigation policy and alerts the user to the presence of the attack, both of which are considered to be detrimental to the attacker's objectives. Identifying and denying these stealthy attacks is critical to securing the sensitive components of UASs and ensuring their safe operation, especially as these systems become increasingly prevalent, both in military and civilian applications.

II. Previous Research

Much of the analysis of control system security in cyber-physical systems has focused on the effects of manipulating sensor measurements to the detriment of the system, with particular emphasis on the sensors and state estimators in electric power grids and smart grid systems. This analysis was initially motivated by the work of Liu et al. who noted that by intelligently choosing which remote terminals to compromise in an electric power grid, an unobservable attack on the weighted least squares estimate of the grid state estimator could be formed. The feasibility of this attack is a function of how many sensors have been corrupted, with it being shown that if sufficiently many sensors are compromised, an attack is guaranteed to exist.[11,12]

This work motivated significant research into false data attacks, with much of it focused on false data injection systems on weighted least squares estimators in power grid systems, including attack formation,[13,14,15,16] protection schemes,[17] and how such attacks can be used to manipulate the electricity commodity market for profit.[18] In addition to analysis of these attacks on power grids, there is analysis of this type of attack on general sensor networks,[19,20] and general feedback systems.[21]

The analysis of false data injection attacks has also been extended to the problem of multi-agent consensus, in which there can be malicious or faulty members of vehicle network that will report false data to other members of the network. The case of a malicious agent inserting data is an example of the Byzantine general problem.[22]Investigations into the security of such networks, with a primary focus on the identification and exclusion of offending agents, have been performed.[4,5,6,7,8]

In this paper, we extend the work of Liu et al. on false data injection attacks on air traffic control systems[2] to false data injection attacks in general linear systems that use Kalman filter estimates to feedback on the system. False data injection attacks have been identified as a particularly interesting attack in part because of the comparative ease with which they can be performed. While the internal components of an autopilot can be directly monitored and protected, sensors are often external and may not be as easily protected. This is especially true for remote sensors which depend on data networks for communication, creating the possibility network-based attacks such as man-in-the-middle or replay attacks.[23] Direct physical manipulation of a remote sensor, which may be less monitored or protected, is also possible, as well as manipulations of the environment to change the sensor readings, as in GPS or Automatic Dependent Surveillance-Broadcast (ADS-B) spoofing.[9,10] Given the noise inherent in sensor measurements, it can also be difficult to distinguish valid measurements from modified measurements. A sophisticated attacker can make use of this to formulate a false data injection attack that is undetectable by vehicle monitoring systems and therefore does not trigger the enforcement of an attack mitigation policy which may be contrary to the attacker's objectives.

III. System Model

A. Nominal System

The dynamics of the closed-loop system are modeled as in (1), with the subscript n used to denote the nominal (no attack) case.

$$\begin{aligned}\mathbf{x}_{\mathrm{n}}[k+1] &= \mathbf{A}\mathbf{x}_{\mathrm{n}}[k] + \mathbf{B}\mathbf{K}\hat{\mathbf{x}}_{\mathrm{n}}[k] + \mathbf{B}_{\mathrm{d}}\mathbf{w}[k] \\ \hat{\mathbf{x}}_{\mathrm{n}}[k+1] &= \mathbf{A}\hat{\mathbf{x}}_{\mathrm{n}}[k] + \mathbf{B}\mathbf{K}\hat{\mathbf{x}}_{\mathrm{n}}[k] + \mathbf{L}[k]\,\mathbf{r}_{\mathrm{n}}[k] \\ \mathbf{r}_{\mathrm{n}}[k] &= \mathbf{C}\mathbf{x}_{\mathrm{n}}[k] + \mathbf{D}\mathbf{v}[k] - \mathbf{C}\hat{\mathbf{x}}_{\mathrm{n}}[k]\end{aligned} \tag{1}$$

In this model, $\mathbf{x}[k] \in \mathbb{R}^{\mathrm{n}}$ is the system state vector, $\hat{\mathbf{x}}[k] \in \mathbb{R}^{\mathrm{n}}$ is the system state estimate, $\mathbf{r}[k] \in \mathbb{R}^{\mathrm{m}}$ is the estimator residual vector, and $\mathbf{u}[k] \in \mathbb{R}^{\mathrm{p}}$ is the system input vector. The process and observation noises, $\mathbf{w}[k] \in \mathbb{R}^{\mathrm{p}}$ and $\mathbf{v}[k] \in \mathbb{R}^{\mathrm{m}}$, respectively, are assumed to be independent, zero-mean white noise processes such that $\mathbb{E}\left[\mathbf{w}[k]\,\mathbf{w}^{\top}[j]\right] = \delta_{ij} = \mathbb{E}\left[\mathbf{v}[k]\,\mathbf{v}^{\top}[j]\right]$, where δ_{ij} is the Kronecker delta. $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the discrete time system transition matrix, $\mathbf{B} \in \mathbb{R}^{n\times p}$ is the system input matrix, $\mathbf{C} \in \mathbb{R}^{m\times n}$ is the system output matrix, and $\mathbf{B}_{\mathrm{d}}$ and $\mathbf{D}$ are real-valued matrices of appropriate dimension to map $\mathbf{w}[k]$ into $\mathbb{R}^{n}$ and $\mathbf{v}[k]$ into $\mathbb{R}^{m}$. $\mathbf{L}[k] \in \mathbb{R}^{n\times m}$ is the Kalman filter gain, as in (2) where $\mathbf{Q}[k] \in \mathbb{R}^{n\times n}$ is the covariance of the state estimate. In the closed-loop configuration, $\mathbf{K} \in \mathbb{R}^{p\times n}$ is an arbitrary feedback gain matrix.

$$\begin{aligned}\mathbf{Q}[k+1] &= \mathbf{A}\mathbf{Q}[k]\,\mathbf{A}^{\top} + \mathbf{B}_{\mathrm{d}}\mathbf{B}_{\mathrm{d}}^{\top} \\ &\qquad - \mathbf{A}\mathbf{Q}[k]\,\mathbf{C}^{\top}\left(\mathbf{C}\mathbf{Q}[k]\,\mathbf{C}^{\top} + \mathbf{D}\mathbf{D}^{\top}\right)^{\text{-1}}\mathbf{C}\mathbf{Q}[k]\,\mathbf{A}^{\top} \\ \mathbf{L}[k] &= \mathbf{A}\mathbf{Q}[k]\,\mathbf{C}^{\top}\left(\mathbf{C}\mathbf{Q}[k]\,\mathbf{C}^{\top} + \mathbf{D}\mathbf{D}^{\top}\right)^{\text{-1}}\end{aligned} \tag{2}$$

In this analysis, the Kalman filter is assumed to have entered the steady state, $\mathbf{L}$ and $\mathbf{Q}_{\infty}$, as in (3) prior to the onset of the attack. Because the error covariance and the gain are dependent on the assumptions of the system model and not on the measurements, the estimator gain and calculated error covariance will not change from these steady state values throughout the attack.

$$\begin{gathered}\mathbf{Q}_{\infty} = \mathbf{A}\mathbf{Q}_{\infty}\mathbf{A}^{\top} + \mathbf{B}_{\mathrm{d}}\mathbf{B}_{\mathrm{d}}^{\top} - \mathbf{A}\mathbf{Q}_{\infty}\mathbf{C}^{\top}\left(\mathbf{C}\mathbf{Q}_{\infty}\mathbf{C}^{\top} + \mathbf{D}\mathbf{D}^{\top}\right)^{\text{-1}}\mathbf{C}\mathbf{Q}_{\infty}\mathbf{A}^{\top} \\ \mathbf{L} = \mathbf{A}\mathbf{Q}_{\infty}\mathbf{C}^{\top}\left(\mathbf{C}\mathbf{Q}_{\infty}\mathbf{C}^{\top} + \mathbf{D}\mathbf{D}^{\top}\right)^{\text{-1}}\end{gathered} \tag{3}$$

The instantaneous error in the state estimate is given by the difference between the actual state value and the corresponding estimate, $\mathbf{e}[k] = \mathbf{x}[k] - \hat{\mathbf{x}}[k]$. Because $\mathbb{E}[\hat{\mathbf{x}}[k]] = \mathbf{x}[k]$ in the nominal case, the expected estimate error, $\bar{\mathbf{e}}[k]$, is 0.

System fault detection is performed using the χ^2 fault detector in which the Kalman filter residuals, $\mathbf{r}[k]$, are statistically tested to determine if they are normally distributed with zero mean and covariance $\mathbf{Q}_{\mathrm{r}} = \left(\mathbf{C}\mathbf{Q}_{\infty}\mathbf{C}^{\top} + \mathbf{D}\mathbf{D}^{\top}\right)$ as is expected under correct operations. If this is the case, the residual test function $g[k] = \mathbf{r}^{\top}[k]\,\mathbf{Q}_{\mathrm{r}}^{\text{-1}}\mathbf{r}[k]$ will have a χ^2 distribution. This distribution can be tested by comparing $g[k]$ to a detection threshold, τ, which can be chosen experimentally or can be derived to give desired detection probabilities. If $g[k]$ is greater than this threshold, the hypothesis that it has a χ^2 distribution is rejected, and a fault is determined to have occurred.[24,25]

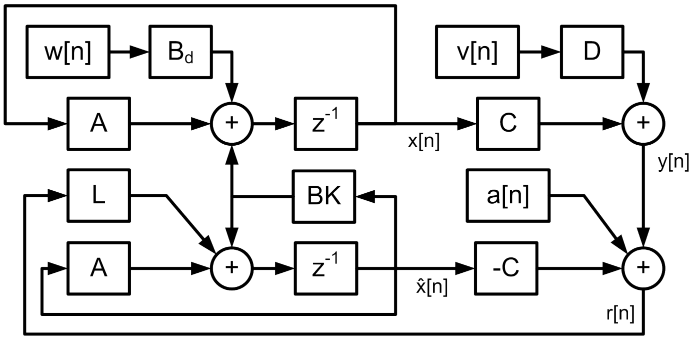


Figure 1: Block diagram of system with attack input

B. Attacked System

For a false data injection attack, the following assumptions are made:

1. The attacker has the ability to arbitrarily modify some or all of the system measurements. This ability will be reflected in the attack availability matrix as defined in (4).

$$\mathbf{F} = \operatorname{diag}\left(\begin{bmatrix} i_1 & i_2 & \dots & i_m \end{bmatrix}\right)$$
$$i_k = \begin{cases} 1, & \text{if measurements from the } k\text{th sensor are modifiable} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

   To simplify the notation, this availability is not incorporated into the system dynamics. Instead, the constraint $(\mathbf{I}_m - \mathbf{F})\,\mathbf{a}\,[k] = 0$ is enforced at every time step during the attack formulation.

2. The objective of the attacker is to formulate the attack input to optimally degrade some aspect of the system performance while not triggering a fault state. Triggering a fault state permits the implementation of a mitigation policy and alerts the user to the presence of the attack, both of which are assumed to be detrimental to the attacker's objectives.

3. The attacker has perfect knowledge of the system dynamics $(\mathbf{A}, \mathbf{B}, \mathbf{C})$, controller and estimator parameters $(\mathbf{L}, \mathbf{Q}_\infty, \mathbf{K})$, and noise properties, $(\mathbf{B}_\mathrm{d}, \mathbf{D})$.

4. The attack occurs over a finite time interval from 0 to T_a.

These assumptions may be invalid for attacks that might be seen in the real world, particularly the assumption that the attacker has perfect knowledge of the system. These broad assumptions are made both so that some initial results can be determined before more specific cases are considered and so that the absolute worst case attack can be identified. This worst case attack can be useful in determining worst case system performance and then determining if the performance is degraded to an extent that would warrant more careful analysis.

The attack input, $\mathbf{a}\,[k] \in \mathbb{R}^\mathrm{m}$ is modeled as an offset to the measurements made by the system as shown in the attacked system dynamics in (5). The system components in this attacked case are denoted by the subscript a.

$$\begin{aligned} \mathbf{x}_\mathrm{a}\,[k+1] &= \mathbf{A}\mathbf{x}_\mathrm{a}\,[k] + \mathbf{B}\mathbf{K}\hat{\mathbf{x}}_\mathrm{a}\,[k] + \mathbf{B}_\mathrm{d}\mathbf{w}\,[k] \\ \hat{\mathbf{x}}_\mathrm{a}\,[k+1] &= \mathbf{A}\hat{\mathbf{x}}_\mathrm{a}\,[k] + \mathbf{B}\mathbf{K}\hat{\mathbf{x}}_\mathrm{a}\,[k] + \mathbf{L}\mathbf{r}_\mathrm{a}\,[k] \\ \mathbf{r}_\mathrm{a}\,[k] &= \mathbf{C}\mathbf{x}_\mathrm{a}\,[k] + \mathbf{D}\mathbf{v}\,[k] + \mathbf{a}\,[k] - \mathbf{C}\hat{\mathbf{x}}_\mathrm{a}\,[k] \end{aligned} \tag{5}$$

The full model of the system, including the attack input, is shown in Figure 1.

IV. Optimization Problem Formulation

In seeking to develop the attack input as an optimization problem that conforms to these assumptions, there are three system error functions that are of note. The state trajectory, which is the difference between the nominal and attacked system states and is necessary for the evaluation of the objective function, is given in (6). Similarly, the estimate trajectory error, which is needed to evaluate the time evolution of the state trajectory error, is defined in (7). Finally, the estimate error, used in the evaluation of the expected residual test signal, is the difference between the system state and the state estimate, and is shown in (8).

$$\Delta\mathbf{x}\left[k\right] = \mathbf{x}_{\mathrm{n}}\left[k\right] - \mathbf{x}_{\mathrm{a}}\left[k\right] \tag{6}$$

$$\Delta\hat{\mathbf{x}}\left[k\right] = \hat{\mathbf{x}}_{\mathrm{n}}\left[k\right] - \hat{\mathbf{x}}_{\mathrm{a}}\left[k\right] \tag{7}$$

$$\mathbf{e}\left[k\right] = \mathbf{x}\left[k\right] - \hat{\mathbf{x}}\left[k\right] \tag{8}$$

The general optimization problem for determining the optimal attack input that maximizes the sum of the square of the state trajectory error is shown in (9), where $\mathbf{W}'$ is an error weighting matrix and T_a is the ending time of the finite duration attack input.

$$\begin{array}{rl} \textbf{Maximize:} & \theta\left(\mathbf{X}\right) = \sum_{k=0}^{T_a} \Delta\mathbf{x}^{\top}\left[k\right]\mathbf{W}'\Delta\mathbf{x}\left[k\right] \\ & \text{1. Expected Estimate Error Dynamics} \\ & \text{2. State Trajectory Difference Dynamics} \\ \textbf{Subject to:} & \text{3. Estimate Trajectory Difference Dynamics} \\ & \text{4. Attack Availability} \\ & \text{5. Detectability Constraint} \end{array} \tag{9}$$

A. Objective Function

In setting up this optimization problem, it is notationally useful to create a single variable that incorporates all of the system parameters over the entire attack horizon that are needed in the optimization. This optimization variable, $\mathbf{X}$, is defined as in (10), where $\boldsymbol{\nu}\left(k\right) \in \mathbb{R}^{\mathrm{3n+m}}$, $\mathbf{X} \in \mathbb{R}^{(T_a+1)(3n+m)}$, and n and m are the dimension of the system state and measurements, respectively.

$$\begin{aligned} \boldsymbol{\nu}\left(k\right) &= \begin{bmatrix}\Delta\mathbf{x}^{\top}\left[k\right] & \Delta\hat{\mathbf{x}}^{\top}\left[k\right] & \mathbf{e}^{\top}\left[k\right] & \mathbf{a}^{\top}\left[k\right]\end{bmatrix} \\ \mathbf{X} &= \begin{bmatrix}\boldsymbol{\nu}\left(0\right) & \boldsymbol{\nu}\left(1\right) & \dots & \boldsymbol{\nu}\left(T_a\right)\end{bmatrix}^{\top} \end{aligned} \tag{10}$$

To formulate the objective function as a function of $\mathbf{X}$, a selection matrix, $\mathbf{S}$, is defined to pick out the state trajectory error over the attack window, where $\mathbf{I}_j$ is an $j \times j$ Identity matrix and $\mathbf{0}_{j,k}$ is a $j \times k$ matrix whose elements are zero.

$$\mathbf{S} = \mathrm{blkdiag}\left(\begin{bmatrix}\mathbf{I}_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,m}\end{bmatrix}\cdots\begin{bmatrix}\mathbf{I}_n & \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \mathbf{0}_{n,m}\end{bmatrix}\right) \tag{11}$$

An expanded weighting matrix $\mathbf{W} = \mathrm{blkdiag}\left(\begin{bmatrix}\mathbf{W}' & \dots & \mathbf{W}'\end{bmatrix}\right)$ is also defined. The objective function for this optimization problem, θ, can then be defined as in (12)

$$\theta\left(\mathbf{X}\right) = \mathbf{X}^{\top}\mathbf{S}^{\top}\mathbf{W}\mathbf{S}\mathbf{X} \tag{12}$$

It is noted that if $\mathbf{W}'$ is chosen to be positive semidefinite, this is a convex function.

B. Error Constraints

The evolution of the expected estimate error is given by (13).

$$\bar{\mathbf{e}}\left[k+1\right] = \mathbf{A}\bar{\mathbf{e}}\left[k\right] - \mathbf{L}\mathbf{a}\left[k\right] \tag{13}$$

The expected difference dynamics are given in Equations (14) and (15) where $\mathbf{A}_{\mathrm{C}} = \mathbf{A} + \mathbf{B}\mathbf{K} - \mathbf{L}\mathbf{C}$. It is noted that the observation and process noises are not dependent on the attack input and provide the same contribution to the attacked trajectory as they would to the nominal trajectory and therefore do not have an effect on the difference. These terms are accordingly not random variables. It is also noted that if the contribution of these noises are not identical, the expectation of the differences will give the same result.

$$\begin{aligned}\Delta\mathbf{x}\left[k+1\right] &= \mathbf{A}\Delta\mathbf{x}\left[k\right] + \mathbf{B}\mathbf{K}\Delta\hat{\mathbf{x}}\left[k\right] && (14)\\ \Delta\hat{\mathbf{x}}\left[k+1\right] &= \mathbf{A}_{\mathrm{C}}\Delta\hat{\mathbf{x}}\left[k\right] + \mathbf{L}\mathbf{C}\Delta\mathbf{x}\left[k\right] - \mathbf{L}\mathbf{a}\left[k\right] && (15)\end{aligned}$$

These dynamics, along with the attack availability constraint, can be incorporated into the problem formulation by defining the constraint matrix $\mathbf{\Gamma}$ in (16).

$$\begin{gathered}\mathbf{\Gamma}_1 = \begin{bmatrix} -\mathbf{I}_n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_m - \mathbf{F} \end{bmatrix}, \ \mathbf{\Gamma}_2 = \begin{bmatrix} \mathbf{A} & \mathbf{B}\mathbf{K} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}\mathbf{C} & \mathbf{A}_{\mathrm{C}} & \mathbf{0} & -\mathbf{L} \\ \mathbf{0} & \mathbf{0} & \mathbf{A} & -\mathbf{L} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \\ \mathbf{\Gamma} = \text{blktoeplitz}\left(\begin{bmatrix}\mathbf{\Gamma}_1 & \mathbf{\Gamma}_2 & \mathbf{0} & \dots & \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{\Gamma}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0}\end{bmatrix}\right)\end{gathered} \tag{16}$$

By applying the affine constraint in (17) during the attack determination, the state and state estimate trajectory error dynamics, expected estimate error dynamics, and attack availability constraints can all be enforced.

$$\mathbf{\Gamma}\mathbf{X} = 0 \tag{17}$$

C. Residual Constraint

To reduce the likelihood of the attack being detected, the attack input should be constrained so as to maintain a minimal profile. Because the χ^2 fault detection scheme uses a threshold comparison test on the residual test signal, $g\left[k\right] = \mathbf{r}_{\mathrm{a}}^{\top}\left[k\right]\mathbf{Q}_{\mathrm{r}}^{-1}\mathbf{r}_{\mathrm{a}}\left[k\right]$, the expected value of the signal, $\bar{g}\left[k\right]$ should be constrained to be below a design threshold, τ' during the attack formation. This design threshold should nominally be less than the threshold value used in the actual fault detection system, τ. This will ensure that the expected operation of the system under attack will not trigger a fault state, although it is still possible for variations in the value of the $g\left[k\right]$ to exceed the detection threshold and trigger a fault in actual operations.

In evaluating the expected value of this residual test signal, the time evolution of the estimate error covariance is one term that is needed. This parameter is given in (18).

$$\begin{aligned}\mathbf{Q}_{\mathrm{a}}\left[k+1\right] = \ & \mathbf{A}_{\mathrm{L}}\mathbf{Q}_{\mathrm{a}}\left[k\right]\mathbf{A}_{\mathrm{L}}^{\top} + \mathbf{B}_{\mathrm{d}}\mathbf{B}_{\mathrm{d}}^{\top} - \mathbf{L}\mathbf{D}\mathbf{D}^{\top}\mathbf{L}^{\top} - \mathbf{L}\mathbf{C}\bar{\mathbf{e}}\left[k\right]\mathbf{a}^{\top}\left[k\right]\mathbf{L}^{\top} \\ & - \mathbf{L}\mathbf{a}\left[k\right]\bar{\mathbf{e}}^{\top}\left[k\right]\mathbf{C}^{\top}\mathbf{L}^{\top} + \mathbf{L}\mathbf{a}\left[k\right]\mathbf{a}^{\top}\left[k\right]\mathbf{L}^{\top} \\ = \ & \mathbf{Q}_{\infty} + \bar{\mathbf{e}}\left[k+1\right]\bar{\mathbf{e}}^{\top}\left[k+1\right]\end{aligned} \tag{18}$$

The residual test signal then can be formulated as a function of the attack input as in (19), where m is again the dimension of the system measurement vector.

$$\begin{aligned}\bar{g}\left[k\right] \ = \ & \operatorname{tr}\left(\mathbb{E}\left[\sqrt{\mathbf{Q}_{\mathrm{r}}^{-1}}\left(\mathbf{C}\mathbf{e}\left[k\right] + \mathbf{D}\mathbf{v}\left[k\right] + \mathbf{a}\left[k\right]\right)\left(\mathbf{C}\mathbf{e}\left[k\right] + \mathbf{D}\mathbf{v}\left[k\right] + \mathbf{a}\left[k\right]\right)^{\top}\sqrt{\mathbf{Q}_{\mathrm{r}}^{-1}}\right]\right) \\ = \ & \operatorname{tr}\left(\mathbf{Q}_{\mathrm{r}}^{-1}\left(\mathbf{C}\bar{\mathbf{e}}\left[k\right]\bar{\mathbf{e}}^{\top}\left[k\right]\mathbf{C}^{\top} + \mathbf{C}\bar{\mathbf{e}}\left[k\right]\mathbf{a}^{\top}\left[k\right] + \mathbf{a}\left[k\right]\bar{\mathbf{e}}^{\top}\left[k\right]\mathbf{C}^{\top} + \mathbf{a}\left[k\right]\mathbf{a}^{\top}\left[k\right]\right)\right) \\ & + \operatorname{tr}\left(\mathbf{Q}_{\mathrm{r}}^{-1}\left(\mathbf{C}\mathbf{Q}_{\infty}\mathbf{C}^{\top} + \mathbf{D}\mathbf{D}^{\top}\right)\right) \\ = \ & \begin{bmatrix}\bar{\mathbf{e}}\left[k\right] \\ \mathbf{a}\left[k\right]\end{bmatrix}^{\top}\begin{bmatrix}\mathbf{C}^{\top}\mathbf{Q}_{\mathrm{r}}^{-1}\mathbf{C} & \mathbf{C}^{\top}\mathbf{Q}_{\mathrm{r}}^{-1} \\ \mathbf{Q}_{\mathrm{r}}^{-1}\mathbf{C} & \mathbf{Q}_{\mathrm{r}}^{-1}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{e}}\left[k\right] \\ \mathbf{a}\left[k\right]\end{bmatrix} + m\end{aligned} \tag{19}$$

A sparse, positive semidefinite matrix, $\mathbf{\Lambda}_k$, can be defined for every $k \in [0,\ T_a]$ such that $\bar{g}\left[k\right] = \mathbf{X}^{\top}\mathbf{\Lambda}_k\mathbf{X} + m$. Making use of the Schur compliment and noting that $\mathbf{Q}_{\mathrm{r}}^{-1}$ is positive definite and $\mathbf{C}^{\top}\mathbf{Q}_{\mathrm{r}}^{-1}\mathbf{C} - \mathbf{C}^{\top}\mathbf{Q}_{\mathrm{r}}^{-1}\mathbf{Q}_{\mathrm{r}}\mathbf{Q}_{\mathrm{r}}^{-1}\mathbf{C}$ is positive semidefinite, $\mathbf{\Lambda}_k$ can be shown to be positive semidefinite and $\bar{g}\left[k\right]$ is therefore a convex function.[26] By forming a vector containing this residual test signal at each time instant, $\mathbf{\Psi}\left(\mathbf{X}\right)$, and constraining each element of that vector to be less than τ in the opimtimization, as in (20), these residual constraints can be enforced.

$$\mathbf{\Psi}\left(\mathbf{X}\right) \coloneqq \begin{bmatrix}\bar{g}\left[0\right] - \tau \\ \bar{g}\left[1\right] - \tau \\ \vdots \\ \bar{g}\left[T_a\right] - \tau\end{bmatrix}, \ \mathbf{\Psi}\left(\mathbf{X}\right) \leq 0 \tag{20}$$

D. Quadratically Constrained Quadratic Program

The functions defined in these previous sections can be combined to form the quadratically constrained quadratic program problem given in (21).

$$\begin{array}{rl} \textbf{Minimize:} & -\theta\left(\mathbf{X}\right) \\ \textbf{Subject to:} & \begin{array}{l} 1.\ \mathbf{\Psi}\left(\mathbf{X}\right) \leq 0 \\ 2.\ \mathbf{\Gamma}\mathbf{X} = 0 \end{array} \end{array} \tag{21}$$

This the minimization of a concave objective function subject to quadratic constraints. Without further development of the problem, a local minimum is therefore not, in general, a global minimum. If $(\mathbf{X}^*, \lambda^*, \mu^*)$ is a solution to the dual problem of the primal problem in (21) as stated in (22), then that solution satisfies the KKT conditions in (23) in addition to satisfying the constraints of the primal problem.[27, 26]

$$g\left(\lambda, \mu\right) = \inf\left(-\theta\left(\mathbf{X}\right) + \lambda\mathbf{\Psi}\left(\mathbf{X}\right) + \mu\mathbf{\Gamma}\mathbf{X}\right) \tag{22}$$

Because this problem generalizes similarly to that investigated by Liu et al., these requirements are of the same form.[2]

$$\begin{gathered} \lambda^* \geq 0 \\ \lambda^{*\top}\mathbf{\Psi}\left(\mathbf{X}^*\right) = 0 \\ -\nabla\theta\left(\mathbf{X}^*\right) + \lambda^*\nabla\mathbf{\Psi}\left(\mathbf{X}^*\right) + \mu^*\nabla\mathbf{\Gamma}\mathbf{X}^* = 0 \end{gathered} \tag{23}$$

Therefore, by solving (23), we can design the optimal attack that maximizes the state error while not being detected.

V. Simulated Results

This analysis was applied to an example continuous time system with eigenvalues at $-.1 \pm j$ and $-.25 \pm 2j$ which was sampled at 10 Hz to give the discrete dynamics in (24).

$$\begin{gathered} \mathbf{A} = \begin{bmatrix} 1.0000 & 0.1000 & 0.0050 & 0.0002 \\ -0.0007 & 0.9998 & 0.0991 & 0.0049 \\ -0.0200 & -0.0071 & 0.9746 & 0.0957 \\ -0.3929 & -0.1461 & -0.5023 & 0.9076 \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} 0.0050 & 0.0000 \\ 0.1000 & 0.0001 \\ -0.0002 & 0.0024 \\ -0.0071 & 0.0479 \end{bmatrix} \\ \mathbf{B}_{\mathrm{d}} = \begin{bmatrix} 0.0002 & 0.0000 \\ 0.0050 & 0.0000 \\ -0.0000 & 0.0001 \\ -0.0004 & 0.0024 \end{bmatrix},\ \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},\ \mathbf{D} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \end{gathered} \tag{24}$$

The steady-state Kalman filter gain was used for this system, as well as the LQR feedback gain with unit state and control costs.

$$\mathbf{L} = \begin{bmatrix} 0.0358 & -0.0368 \\ 0.0094 & -0.0061 \\ -0.0376 & 0.0443 \\ -0.0143 & 0.0088 \end{bmatrix},\ \mathbf{K} = \begin{bmatrix} -1.9434 & -2.0965 & -1.7853 & -0.2128 \\ 0.3209 & -0.0277 & 0.0447 & -0.3600 \end{bmatrix} \tag{25}$$

The first state sensor is available for attack and is also the state of interest for error insertion.

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ \mathbf{W}' = \mathrm{diag}\left(\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}\right) \tag{26}$$

It is noted that because this system does not have an unstable eigenvalue, it does not satisfy a necessary condition for a perfectly attackable system.[21] There will therefore not exist an attack sequence $\mathcal{A}$ with a bounded attack input at every instant that will introduce arbitrarily large state error as the attack horizon becomes infinitely large.

$$\nexists \left\{ \mathcal{A} \;\middle|\; \|\mathbf{a}\,[k]\| \le 1 \;\forall\; k \in [0,\ T_a]\,,\ \ \lim_{T_a \to \infty} \sup\left(\|\Delta\mathbf{x}\,[T_a]\|\right) = \infty \right\} \tag{27}$$

The goal of this attack will accordingly be to degrade system performance and not to cause unbounded state error.

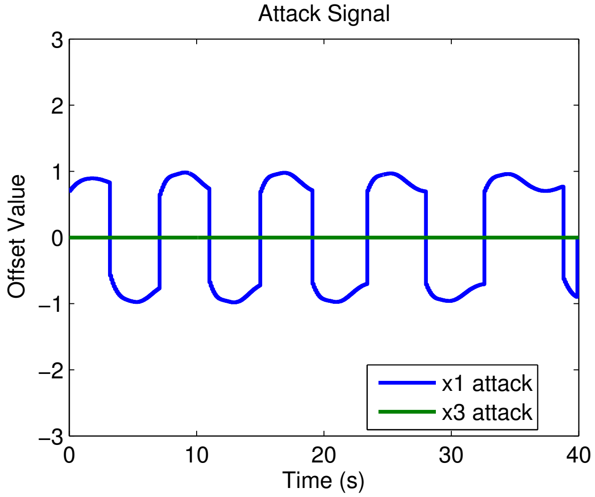
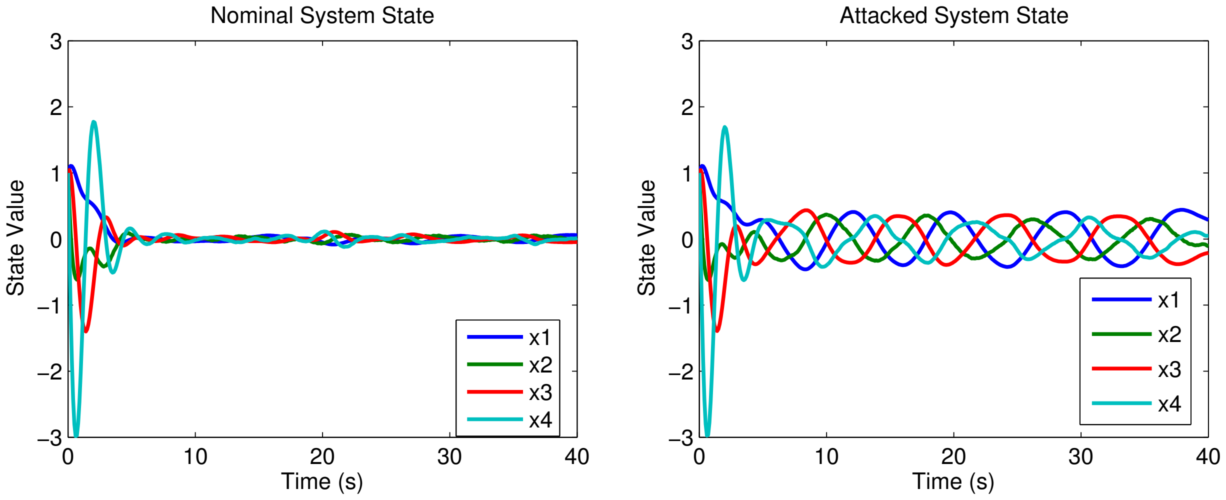


Figure 2: Optimal attack input for example system

The attack input for this system was determined by numerically solving the optimization problem formed in Section D using the MATLAB fmincon function and the sequential quadratic programming algorithm. This input is shown in Figure 2. A comparison of the nominal and attacked state trajectories for the system given this attack input are shown in Figure 3. As can be seen, the performance of the system is considerably degraded by the attack, with large steady-state oscillations being introduced.



(a) Nominal system state

(b) Attacked system state

Figure 3: State trajectory comparison under false data injection attack

The value of the residual test signal throughout this simulation is shown in Figure 4. The fault detection threshold for this system, was chosen to be 50. As can be seen in the plot, the actual values of this signal over the attack window are centered about this designed value. By choosing this design threshold to be sufficiently smaller than the actual detection threshold, the attack can go undetected.

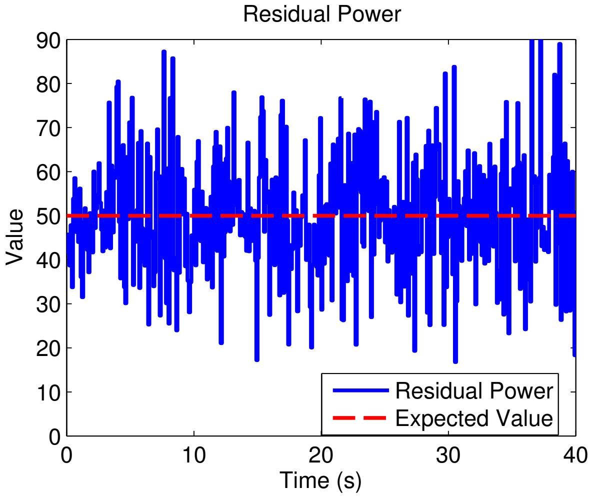


Figure 4: Optimal attack input for example system

VI. Conclusion

A model of false data injection attacks on linear systems that use Kalman filters to implement feedback has been presented. This model was used to formulate an optimization problem that identifies the optimal false data injection attack on a subset of available sensors that can be input into the system to maximize the weighted state trajectory error relative to the nominal state trajectory. Constraints were placed on this attack to limit the expected value of a residual test signal to be less than a designed threshold, enabling the attack to go undetected by vehicle monitoring systems. This provides an easy method to identify the worst case false data injection attacks and evaluate system performance in that case. The system can then be secured as needed to ensure acceptable performance in the event of such stealthy attacks.

References

[1] Anderson, R., *Security Engineering: A guide to building dependable distributed systems*, Wiley, 2010.

[2] Liu, W., Kwon, C., Aljanabi, I., and Hwang, I., "Cyber Security Analysis for State Estimators in Air Traffic Control Systems," *Guidance, Navigation, and Control and Co-located Conferences*, American Institute of Aeronautics and Astronautics, Aug. 2012.

[3] Kim, A., Wampler, B., Goppert, J., Hwang, I., and Aldridge, H., "Cyber Attack Vulnerabilities Analysis for Unmanned Aerial Vehicles," *Infotech@Aerospace Conferences*, American Institute of Aeronautics and Astronautics, June 2012.

[4] Pasqualetti, F., Bicchi, A., and Bullo, F., "Distributed intrusion detection for secure consensus computations," *Decision and Control, 2007 46th IEEE Conference on*, dec. 2007, pp. 5594 –5599.

[5] Pasqualetti, F., Bicchi, A., and Bullo, F., "On the security of linear consensus networks," *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, dec. 2009, pp. 4894 –4901.

[6] Pasqualetti, F., Bicchi, A., and Bullo, F., "Consensus Computation in Unreliable Networks: A System Theoretic Approach," *Automatic Control, IEEE Transactions on*, Vol. 57, No. 1, jan. 2012, pp. 90 –104.

[7] Pasqualetti, F., Carli, R., Bicchi, A., and Bullo, F., "Identifying cyber attacks via local model information," *Decision and Control (CDC), 2010 49th IEEE Conference on*, dec. 2010, pp. 5961 –5966.

[8] Sundaram, S. and Hadjicostis, C., "Distributed Function Calculation via Linear Iterative Strategies in the Presence of Malicious Agents," *Automatic Control, IEEE Transactions on*, Vol. 56, No. 7, july 2011, pp. 1495 –1508.

[9] Krozel, J. and Andrisani, D., "Independent ADS-B Verification and Validation," *Aviation Technology, Integration, and Operations (ATIO) Conferences*, American Institute of Aeronautics and Astronautics, Sept. 2005, pp. –.

[10] Oshman, Y. and Koifman, M., "Robust Navigation Using the Global Positioning System in the Presence of Spoofing," *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 1, Jan. 2006, pp. 95–104.

[11] Liu, Y., Ning, P., and Reiter, M. K., "False data injection attacks against state estimation in electric power grids," *Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, ACM, New York, NY, USA, 2009, pp. 21–32.

[12] Liu, Y., Ning, P., and Reiter, M., "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, Vol. 14, No. 1, 2011, pp. 13.

[13] Teixeira, A., Amin, S., Sandberg, H., Johansson, K., and Sastry, S., "Cyber security analysis of state estimators in electric power systems," *Decision and Control (CDC), 2010 49th IEEE Conference on*, IEEE, 2010, pp. 5991–5998.

[14] Kosut, O., Jia, L., Thomas, R., and Tong, L., "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, IEEE, 2010, pp. 220–225.

[15] Yang, Q., Yang, J., Yu, W., Zhang, N., and Zhao, W., "On a Hierarchical False Data Injection Attack on Power System State Estimation," *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, dec. 2011, pp. 1 –5.

[16] Rahman, M., *False Data Injection Attacks with Incomplete Information*, Master's thesis, Texas Tech University, 2012.

[17] Dan, G. and Sandberg, H., "Stealth Attacks and Protection Schemes for State Estimators in Power Systems," *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, oct. 2010, pp. 214 –219.

[18] Xie, L., Mo, Y., and Sinopoli, B., "False data injection attacks in electricity markets," *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, IEEE, 2010, pp. 226–231.

[19] Bishop, A. and Savkin, A., "On false-data attacks in robust multi-sensor-based estimation," *Control and Automation (ICCA), 2011 9th IEEE International Conference on*, IEEE, 2011, pp. 10–17.

[20] Mo, Y., Garone, E., Casavola, A., and Sinopoli, B., "False data injection attacks against state estimation in wireless sensor networks," *Decision and Control (CDC), 2010 49th IEEE Conference on*, IEEE, 2010, pp. 5967–5972.

[21] Mo, Y. and Sinopoli, B., "False data injection attacks in control systems," *Preprints of the 1st Workshop on Secure Control Systems*, 2010.

[22] Lamport, L., Shostak, R., and Pease, M., "The Byzantine Generals Problem," *ACM Trans. Program. Lang. Syst.*, Vol. 4, No. 3, July 1982, pp. 382–401.

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

[24] Ding, S., *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools*, Springer, 2008.

[25] Willsky, A. and Jones, H., "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *Automatic Control, IEEE Transactions on*, Vol. 21, No. 1, feb 1976, pp. 108 – 112.

[26] Boyd, S. and Vandenberghe, L., *Convex optimization*, Cambridge university press, 2004.

[27] Dhara, A. and Dutta, J., *Optimality Conditions in Convex Optimization: A Finite-Dimensional View*, CRC Press, 2012.