Hindawi
Complexity
Volume 2018, Article ID 1240149, 8 pages
https://doi.org/10.1155/2018/1240149

Research Article

Attack Detection/Isolation via a Secure Multisensor Fusion Framework for Cyberphysical Systems

Arash Mohammadi,[1] **Chun Yang**,[2] **and Qing-wei Chen**[2]

[1]*Concordia Institute for Information System Engineering, Concordia University, Montreal, QC, Canada H3H-1M8*
[2]*College of Automation, Nanjing University of Science and Technology, Nanjing 210094, China*

Correspondence should be addressed to Arash Mohammadi; marash@ece.utoronto.ca

Received 14 September 2017; Accepted 9 January 2018; Published 11 February 2018

Academic Editor: Carlos Gershenson

Motivated by rapid growth of cyberphysical systems (CPSs) and the necessity to provide secure state estimates against potential data injection attacks in their application domains, the paper proposes a secure and innovative attack detection and isolation fusion framework. The proposed multisensor fusion framework provides secure state estimates by using ideas from interactive multiple models (IMM) combined with a novel fuzzy-based attack detection/isolation mechanism. The IMM filter is used to adjust the system's uncertainty adaptively via model probabilities by using a hybrid state model consisting of two behaviour modes, one corresponding to the ideal scenario and one associated with the attack behaviour mode. The state chi-square test is then incorporated through the proposed fuzzy-based fusion framework to detect and isolate potential data injection attacks. In other words, the validation probability of each sensor is calculated based on the value of the chi-square test. Finally, by incorporation of the validation probability of each sensor, the weights of its associated subsystem are computed. To be concrete, an integrated navigation system is simulated with three types of attacks ranging from a constant bias attack to a non-Gaussian stochastic attack to evaluate the proposed attack detection and isolation fusion framework.

1. Introduction

Cyberphysical Systems (CPSs) [1] are designed by integrating control, communication, and processing technologies with the main goal of monitoring/managing critical physical infrastructures. CPSs have attracted significant attentions recently both in academia and in industry due to their exceptional properties and as such emerged in different applications of paramount engineering importance such as medical systems [2], power/energy grids [3], aerospace [4], industrial/manufacturing process control [5], and transportation [6], where performing secure and optimal state estimation is the key concern. In recent years, sensor technologies and communication systems have gone through extensive advancements and improvements making it possible to deploy several sensors simultaneously in CPSs. Such developments have resulted in a significant increase in different CPS application domains. This increasing interest in deployment of CPSs and factoring in that safety and security is of paramount importance in such application domains, investigating security issues of CPSs from different angles has attracted great research interest recently [7–10]. A potential cyber/physical attack in CPSs could have serious ramifications from leakage of consumer information, damaging economy, loss of critical infrastructures, and even threatening humans. Consequently, it is of significant practical importance to detect, identify, and prevent zero-day attacks in real-time with high accuracy which is the focus of this paper.

In this paper, our main focus is to design an attack detection/isolation solution for multisensor state estimation problems in CPSs. The χ^2-test or as commonly called, residue-based test [11], is considered to be the conventional detection solution [12–14] typically used in CPSs. The χ^2-test utilizes a normalized version of the power of the residuals based on the steady-state innovation covariance. In such a conventional detection criterion, the system is statistically evaluated based on a predefined and assumed

model; that is, it is common to base the calculation on some functional form of the innovation sequence (e.g., using trace or determinant operators, in the case of χ^2-test, the former is used). Utilization of such functional form of the innovation sequence results in integration of diagonal and off-diagonal components of the innovation which in turn results in overlooking important statistical information.

The paper addresses this drawback. In particular, we propose a multisensor fusion framework which provides secure state estimates by assigning an interactive multiple model (IMM) filter to each sensor modality. The IMM filter adjusts the system's uncertainty adaptively via model probabilities by constructing a hybrid state model consisting of two modes: one corresponding to the ideal scenario representing clean measurements and one modeling the presence of potential attacks (referred to as the attack behaviour mode). The state χ^2-test is then incorporated through a proposed fuzzy-based fusion framework to detect and isolate potential data injection attacks. The values obtained from the χ^2-test assigned to each sensor are then used to compute the validation probability of each sensor. To overcome the difficulty in selecting an appropriate threshold, we construct the detection threshold based on the χ^2-test's values with two boundaries and an up boundary. Finally, by incorporation of the validation probability of each sensor, the weights of its associated subsystem are computed.

The rest of the paper is organized as follows: first, Section 2 formulates the attack detection/isolation problem in CPSs and presents different attack models. Section 3 develops the proposed fusion framework and attack isolation mechanism. Section 4 presents simulation results based on an integrated navigation system consisting of three observation nodes, that is, Global Navigation System (GPS), the Bei-Dou2 (BD2), and Strap-down Inertial Navigation System (SINS). The paper is finally concluded in Section 5.

2. Problem Formulation

We consider the following general linear state model to represent the underlying physical system:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \boldsymbol{\omega}_k, \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ denotes the state vector at iteration k, $\boldsymbol{\omega}_k \in \mathbb{R}^{n_x}$ is the state noise component which is considered to be distributed according to a Gaussian distribution, independent of the state vector, with zero-mean and known covariance matrix, that is, $\boldsymbol{\omega}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$. The CPS of interest is monitored using a set of N observation nodes (sensors) communicating their data to the remote processing unit referred to as the fusion centre (FC) to perform the required estimation task. The measurement model of sensor l, for $(1 \le l \le N)$, is given by

$$\mathbf{z}_k^{(l)} = \mathbf{H}_k^{(l)} \mathbf{x}_k + \boldsymbol{\xi}_k^{(l)}, \tag{2}$$

where $\mathbf{z}_k^{(l)} \in \mathbb{R}^{n_z}$ represents the observation vector collected by sensor l, for $(1 \le l \le N)$ at iteration k. The uncertainty in the observation vector is modeled by $\boldsymbol{\xi}_k^{(l)}$ which is considered to be distributed according to a Gaussian distribution with zero-mean and known covariance matrix, that is, $\boldsymbol{\xi}_k^{(l)} \sim \mathcal{N}(0, \mathbf{R}_k^{(l)})$.

In this paper, we consider attack surfaces [15–17] where an adversary compromises the underlying system by injecting a bias $\mathbf{b}_k^{(l)}$ (possibly time-varying and/or stochastic) into a subset of measurements at iteration k. Based on the original measurement model (see (2)), the measurement model under the attack, therefore, is represented as follows:

$$\mathscr{Z}_k^{(l)} = \mathbf{z}_k^{(l)} + \mathbf{b}_k^{(l)} = \mathbf{H}_k^{(l)} \mathbf{x}_k + \boldsymbol{\xi}_k^{(l)} + \mathbf{b}_k^{(l)}, \tag{3}$$

where $\mathscr{Z}_k^{(l)}$ denotes possible attacked measurement collected by the lth sensor. In particular, we consider the following three type of attack scenarios:

(i) Constant attack where the injected bias ($\mathbf{b}_k^{(l)}$) into a measurement is constant over time, that is, $\mathbf{b}_k^{(l)} = \mathbf{b}^{(l)}$

(ii) Time-varying attack where the injected bias changes over time, for instance, trigonometric functions,

$$\mathbf{b}_k^{(l)} = A * \sin(\Omega * t) \tag{4}$$

(iii) Stochastic attack where the injection randomly changes over time with some statistical properties being selected by the adversary and unknown to the detection mechanism.

Our goal in this paper is to devise a novel monitoring solution to detect such attacks in real-time with minimum latency and isolate the compromised sensors. Without loss of generality and for simplicity of the presentation, we consider the following assumption.

Assumption 1. In a sensor network with N observation nodes which is under data injection attacks, number of attacked sensors at iteration k, denoted by $\mathscr{M}$, is not equal to the overall number of available sensor nodes ($\mathscr{M} < N$).

This assumption is considered to guarantee that at each iteration at least one unattacked sensor is available for performing the state estimation task. Please note that this assumption is not restrictive as, in absence of an unattacked sensor node, the overall fusion framework continues to provide predictive state estimates while the problem is being investigated and attacked sensors are restored.

In the next section, we present our proposed attack detection/isolation framework which at each iteration isolates the attacked signal and performs the estimation task only based on the remaining clean measurements.

3. Fusion Framework with Attack Isolation

In order to design a monitoring framework capable of detecting all the three aforementioned injection attacks, first we model the two possible scenarios, that is, the attack and the ideal behaviour modes, by designing two different error covariance matrices for the state forcing terms. This design

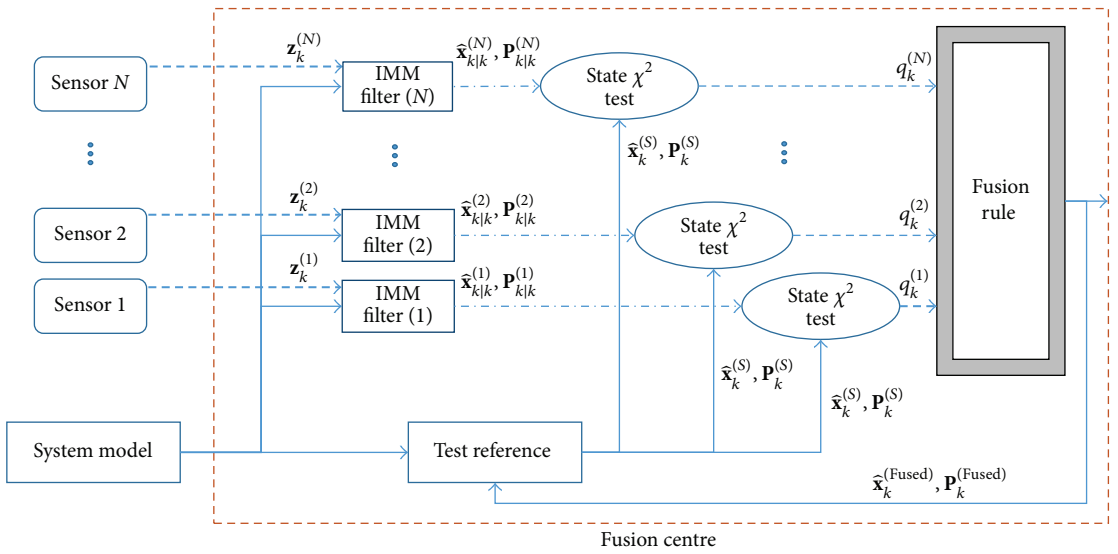

FIGURE 1: Algorithm architecture.

methodology introduces structural uncertainty into the state model for which an IMM filter is associated with each active sensor. The IMM filers are used cooperatively to provide the estimate of the underlying states.

Considered as the first protection layer, this setup will increase the accuracy of the fusion model under potential attacks. On the other hand, in order to isolate attacked measurements which are incorporated to update associated probability corresponding to each model within the pool of IMM filters, the information provided by the χ^2-test is utilized. In other words, we use the measurement which has minimum χ^2-test value for updating the associated probability of each filter in the IMM filterbank. Consequently, this proposed approach results in updating the model probabilities based on the sensor measurement which is less likely to be under attack and therefore further increases the accuracy of the fusion task.

Figure 1 illustrates the architecture of the proposed attack detection/isolation framework. In summary and at each update iteration, the proposed attack detection/isolation works as follows:

(i) Each node (subsystem) transfers its local measurements to its associated IMM filter which in turn computes an updated estimate of the state vector $\hat{\mathbf{x}}_{k|k}^{(l)}$ and its associated error covariance matrix $\mathbf{P}_{k|k}^{(l)}$ which are updated with that subsystem's measurements.

(ii) This information ($\hat{\mathbf{x}}_{k|k}^{(l)}$ and $\mathbf{P}_{k|k}^{(l)}$), for $(1 \leq l \leq N)$, is then transferred to the χ^2-square test block, associated with subsystem l to perform attack detection tasks.

(iii) The detection block computes a failure detection value $q_k^{(l)}$ and transfers it to the central node to be fused with the information from other subsystems and to perform the final attack detection/isolation.

(iv) For the purpose of selecting the best available observation to be utilized for evaluation of the IMM filters' model probabilities, the available fault detection information is used and the subsystem which has the minimum fault value is considered as the selected subsystem for updating the IMM filters' model probabilities. At the same time and to update (calibrate) the reference data (i.e., $\mathbf{P}_k^{(S)}$ and $\hat{\mathbf{x}}_k^{(S)}$), we incorporate the global fused information.

In brief, the proposed attack detection/isolation framework has total of N (number of sensor subsystems) IMM filters as illustrated in Figure 1. Filter l sequentially computes predicted values for the state vector (referred to as IMM-Predict step) and updated state estimates (referred to as IMM-Update step) in parallel to the other filters and only based on the observation it receives from the subsystem associated with that IMM filter l. Next, we first present details of the prediction step for one subsystem.

(i) Mixing Step. In this step, interaction probability $\mu_k^{i|j}$ for $(i, j \in \{I, A\})$ where $M^{(I)}$ is the model corresponding to the ideal scenario, while $M^{(A)}$ corresponds to the system under attack, is calculated as follows:

$$\bar{c}_k^j = \sum_{i=1}^{2} p_{ij} \mu_{k-1}^i,$$
$$\mu_k^{i|j} = \frac{1}{\bar{c}_k^j} p_{ij} \mu_{k-1}^i, \tag{5}$$

where p_{ij} denotes transition probability of moving from state i to j which is defined based on the following transition matrix:

$$p_{ij} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}. \tag{6}$$

Term μ_{k-1}^{1} represents the probability that model $M^{(I)}$ has dominated the previous time instant (the system was in the ideal mode at iteration $(k-1)$ and not under attack), while μ_{k-1}^{2} represents the probability that model $M^{(A)}$ was in effect at the previous iteration which represents the case where the system was under attack at time $(k-1)$. The mixed inputs for each filter are computed as follows

$$\begin{aligned}\widehat{\mathbf{x}}_{k-1|k-1}^{0j} &= \sum_{i=1}^{2}\mu_k^{i|j}\widehat{\mathbf{x}}_{k-1|k-1}^{i}\\ \mathbf{P}_{k-1|k-1}^{0j} &= \sum_{i=1}^{2}\mu_k^{i|j}\left\{\mathbf{P}_{k-1|k-1}^{i}\left[\widehat{\mathbf{x}}_{k-1|k-1}^{i}-\widehat{\mathbf{x}}_{k-1|k-1}^{0j}\right]\right.\\ &\quad\left.\times\left[\widehat{\mathbf{x}}_{k-1|k-1}^{i}-\widehat{\mathbf{x}}_{k-1|k-1}^{0j}\right]^{T}\right\}.\end{aligned}\tag{7}$$

(ii) Local Prediction Step. In this scenario, each of the two mode-matched KFs (one matched to the ideal mode and one matched to the attacked behaviour mode) performs its corresponding prediction step based on the following equations:

$$\begin{aligned}\widehat{\mathbf{x}}_{k|k-1}^{0j} &= \mathbf{F}^{j}\widehat{\mathbf{x}}_{k-1|k-1}^{0j}\\ \mathbf{P}_{k|k-1}^{0j} &= \mathbf{F}^{j}\mathbf{P}_{k-1|k-1}^{0j}\left[\mathbf{F}^{j}\right]^{T}+\mathbf{Q}_{k}^{j},\end{aligned}\tag{8}$$

which in part results in computation of the mode-matched predicted estimate of the states and its associated error covariance.

(iii) Combined Prediction Step. In this final step of the IMM-Predict module, we combine the means and covariance matrices of the attack and ideal modes to form the combined values for the predicted estimate of the states as follows:

$$\begin{aligned}\mathbf{x}_{k|k-1} &= \sum_{j=1}^{2}\mu_{k-1}^{j}\widehat{\mathbf{x}}_{k|k-1}^{0j}\\ \mathbf{P}_{k|k-1} &= \sum_{j=1}^{2}\mu_{k-1}^{j}\left\{\mathbf{P}_{k|k-1}^{0j}+\left[\mathbf{x}_{k|k-1}^{0j}-\widehat{\mathbf{x}}_{k|k-1}\right]\right.\\ &\quad\left.\times\left[\mathbf{x}_{k|k-1}^{0j}-\widehat{\mathbf{x}}_{k|k-1}\right]^{T}\right\}.\end{aligned}\tag{9}$$

This completes the prediction step of the proposed attack detection/isolation framework. Next, we present details of the update step of the proposed framework.

(iv) Mode-Matched KF Update. Local state vector associated with the KF matched to one of the two ideal or attack modes is updated as follows:

$$\mathbf{K}^{j}=\mathbf{P}_{k|k-1}^{0j}\left[\mathbf{H}^{j}\right]^{T}\left[\mathbf{R}_{k}^{j}\right]^{-1}\tag{10}$$

$$\boldsymbol{\zeta}_{k}^{j}=\mathbf{z}_{k}-\mathbf{H}^{j}\widehat{\mathbf{x}}_{k|k-1}^{0j}\tag{11}$$

$$\mathbf{S}_{k}^{j}=\mathbf{H}^{j}\mathbf{P}_{k|k-1}^{0j}\left[\mathbf{H}^{j}\right]^{T}+\mathbf{R}_{k}^{j}\tag{12}$$

$$\left[\mathbf{P}_{k|k}^{j}\right]^{-1}=\left[\mathbf{P}_{k|k-1}^{0j}\right]^{-1}+\left[\mathbf{H}^{j}\right]^{T}\left[\mathbf{R}_{k}^{j}\right]^{-1}\mathbf{H}^{j}\tag{13}$$

$$\boldsymbol{\Lambda}_{k}^{j}=\mathcal{N}\left(\boldsymbol{\zeta}_{k}^{j};\mathbf{0},\mathbf{S}_{k}^{j}\right),\tag{14}$$

where term $\boldsymbol{\Lambda}_{k}^{j}$ is the likelihood function. Note that IMM-KF l uses its specific observation ($\mathbf{z}_{k}^{(l)}$) instead of $\mathbf{z}_{k}$ in (11).

(v) Attack and Idle Model Probabilities. In this step, we need to update the probability that each of the two modes is in effect at a given iteration (k). The required probabilities are calculated as follows:

$$c_{k}=\sum_{j=1}^{2}\boldsymbol{\Lambda}_{k}^{j}\bar{c}_{k}^{j},\tag{15}$$

$$\mu_{k}^{j}=\frac{1}{c}\boldsymbol{\Lambda}_{k}^{j}\bar{c}_{k}^{j},\tag{16}$$

where term c_k in (16) is included as a normalization factor to ensure that it represents a true probability distribution.

(vi) Fusion Step. In this step, the local state estimates and covariance matrices associated with the ideal and attack modes are combined to form the fused components as follows:

$$\begin{aligned}\widehat{\mathbf{x}}_{k|k}^{(l)} &= \sum_{j=1}^{2}\mu_{k}^{j}\widehat{\mathbf{x}}_{k|k}^{j}\\ \mathbf{P}_{k|k}^{(l)} &= \sum_{j=1}^{2}\mu_{k}^{j}\left\{\mathbf{P}_{k|k}^{j}+\left[\widehat{\mathbf{x}}_{k|k}^{j}-\widehat{\mathbf{x}}_{k|k}\right]\times\left[\widehat{\mathbf{x}}_{k|k}^{j}-\widehat{\mathbf{x}}_{k|k}\right]^{T}\right\}.\end{aligned}\tag{17}$$

Once this step is completed, the update stage of the proposed framework is complete. Next, we present the attack detection and compromised measurement isolation methodologies of the proposed fusion framework.

3.1. Attack Isolation Framework. We use the state χ^2-test within the proposed framework to detect an attack. And the test value is defined as follows:

$$q_{k}^{(l)}=\left\|\widehat{\mathbf{x}}_{k|k}^{(l)}-\widehat{\mathbf{x}}_{k}^{(S)}\right\|_{(\mathbf{P}_{k}^{(S)}-\mathbf{P}_{k|k}^{(l)})^{-1}},\tag{18}$$

where $\|\cdot\|$ denotes inner product in the Euclidean space. Attacks on a measurements obtained from one sensor node is evaluated via the following detection mechanism:

$$\begin{aligned}&\text{if } q_{k}^{(l)}\geq T_{D}, \quad \text{Data injection attack mode}\\ &\text{if } q_{k}^{(l)}<T_{D}, \quad \text{Idle mode (no attack)},\end{aligned}\tag{19}$$

where the required threshold (T_D) is computed based on the available tables for χ^2-test [18].

In order to define whether sensor l, for $(1 \leq l \leq N)$, is attacked or not, a validation probability is defined corresponding to each sensor. The aforementioned validation

probability is designed to be a function of the associated χ^2-test value and is given by

$$\beta\left(q_k^{(l)}\right) = \begin{cases} 1 & q_k^{(l)} \leq T_1 \\ -\dfrac{q_k^{(l)}}{C} + \dfrac{T_2}{C} & T_1 < q_k^{(l)} \leq T_2 \\ 0 & q_k^{(l)} > T_2. \end{cases} \tag{20}$$

The above validation probability rule states that when $q_k^{(l)} \leq T_1$, the sensor is in ideal mode with high probability. On the other hand, in cases where $q_k^{(l)} > T_2$, the sensor is under attack with high probability. In the third possible scenario ($T_1 < q_k^{(l)} \leq T_2$), the sensor belongs to an intermediate state which is between the state of attack and being ideal (the sensor is softly attacked; i.e., it could be a candidate for an attacked sensor). Theoretically speaking, the quadratic term $q_k^{(l)} \in \mathbb{R}^+$ appearing in (18) has three degrees of freedom as it is distributed according to the χ^2 distribution [18]. The limit values (T_1 and T_2) defined in (20) are obtained based on this fact and using standard χ^2 tables. These values are defined to provide the required confidence level. However, utilization of a predefined threshold in practical scenarios is not feasible; therefore, an alternative solution is required. In this paper, our contribution is utilization of fuzzy logic to solve this practical issue and identified the required threshold values (T_1 and T_2). Based on 90% confidence level obtained from χ^2-test standard tables, we compute the first threshold as $T_1 = 6.25$ and, similarly based on 99% confidence level obtained from χ^2-test standard tables, we obtain $T_2 = 11.35$. Finally based on χ^2-test standard tables, the value of the only constant C defined in (20) is computed and set to $C = 5.1$.

Without loss of generality and for simplicity of the presentation, in the following discussion, we consider a two-sensor scenario where at each iteration at least one of the sensors is not under attack. The sensor's validation probability is given by

$$\begin{aligned} \lambda_k^{(1)} &= \beta\left(q_k^{(1)}\right) \\ \lambda_k^{(2)} &= \beta\left(q_k^{(2)}\right) \\ \lambda_k^{(1\cap 2)} &= \beta\left(q_k^{(1)}\right)\beta\left(q_k^{(2)}\right). \end{aligned} \tag{21}$$

Term $\lambda_k^{(1)}$ denotes the probability that Sensor 1 is in an ideal behaviour mode (not attacked). Similarly, $\lambda_k^{(2)}$ denotes the validation probability that Sensor 2 is in an ideal mode. On the other hand, $\lambda_k^{(1\cap 2)}$ relates to the case where both sensors are in ideal mode at time k. We compute an adaptive weight for each sensor based on the above-mentioned probabilities as follows:

$$\begin{aligned} \alpha_k^{(1)} &= \lambda_k^{(1)} - \lambda_k^{(1\cap 2)} \\ \alpha_k^{(2)} &= \lambda_k^{(2)} - \lambda_k^{(1\cap 2)} \\ \alpha_k^{(1\cap 2)} &= \lambda_k^{(1\cap 2)} \\ \alpha_k^{(0)} &= 1 - \lambda_k^{(1)} - \lambda_k^{(2)} + \lambda_k^{(1\cap 2)}, \end{aligned} \tag{22}$$

where $\alpha_k^{(1)}$ refers to the scenario where only Sensor 1 is not attacked (Sensor 2 is potentially under attack). Similarly, $\alpha_k^{(2)}$ denotes the scenario where only Sensor 2 is not attacked (Sensor 1 is potentially under attack). On the other hand, $\alpha_k^{(1\cap 2)}$ corresponds to the case where not one but both of the sensors are in ideal mode simultaneously. Finally, term $\alpha_k^{(0)}$ corresponds to the scenario where both sensors are under potential attacks.

The computed validation probabilities are then used to adaptively compute the estimated values of the state variables and their associated error covariance matrix. In this adaptive framework, the weights are assigned based on the validation probabilities. The fusion algorithm also incorporates the estimates for the ideal mode without presence of any attacks at iteration k and computes the updated statistics as follows:

$$\begin{aligned} \widehat{\mathbf{x}}_{k|k}^{(1\cap 2)} &= \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(1|1\cap 2)}\left(\mathbf{z}_k^{(1)} - \mathbf{H}^{(1)}\widehat{\mathbf{x}}_{k|k-1}\right) \\ &\quad + \mathbf{K}_k^{(2|1\cap 2)}\left(\mathbf{z}_k^{(2)} - \mathbf{H}^{(2)}\widehat{\mathbf{x}}_{k|k-1}\right) \end{aligned} \tag{23}$$

$$\left[\mathbf{P}_{k|k}^{(1\cap 2)}\right]^{-1} = \mathbf{P}_{k|k-1}^{-1} + \left[\mathbf{H}^{(1)}\right]^T\left[\mathbf{R}_k^{(1)}\right]^{-1}\mathbf{H}^{(1)} \tag{24}$$

$$+ \left[\mathbf{H}^{(2)}\right]^T\left[\mathbf{R}_k^{(2)}\right]^{-1}\mathbf{H}^{(2)}, \tag{25}$$

where the Kalman gains are given by

$$\mathbf{K}_k^{(1|1\cap 2)} = \mathbf{P}_{k|k}^{(1\cap 2)}\left[\mathbf{H}^{(1)}\right]^T\left[\mathbf{R}_k^{(1)}\right]^{-1} \tag{26}$$

$$\mathbf{K}_k^{(2|1\cap 2)} = \mathbf{P}_{k|k}^{(1\cap 2)}\left[\mathbf{H}^{(2)}\right]^T\left[\mathbf{R}_k^{(2)}\right]^{-1}. \tag{27}$$

Once the above set of statistics are computed, the updated values for the overall system are computed as follows:

$$\widehat{\mathbf{x}}_{k|k} = \alpha_0\widehat{\mathbf{x}}_{k|k-1} + \alpha_1\widehat{\mathbf{x}}_{k|k}^{(1)} + \alpha_2\widehat{\mathbf{x}}_{k|k}^{(2)} + \alpha_{(1\cap 2)}\widehat{\mathbf{x}}_{k|k}^{(1\cap 2)} \tag{28}$$

$$\begin{aligned} &\mathbf{P}_{k|k} \\ &= \alpha_0\mathbf{P}_{k|k-1} + \alpha_1\left[\mathbf{P}_{k|k}^{(1)} + \left(\widehat{\mathbf{x}}_{k|k} - \widehat{\mathbf{x}}_{k|k}^{(1)}\right)\left(\widehat{\mathbf{x}}_{k|k} - \widehat{\mathbf{x}}_{k|k}^{(1)}\right)^T\right] \\ &\quad + \alpha_2\left[\mathbf{P}_{k|k}^{(2)} + \left(\widehat{\mathbf{x}}_{k|k} - \widehat{\mathbf{x}}_{k|k}^{(2)}\right)\left(\widehat{\mathbf{x}}_{k|k} - \widehat{\mathbf{x}}_{k|k}^{(2)}\right)^T\right] \\ &\quad + \alpha_{(1\cap 2)}\left[\mathbf{P}_{k|k}^{(1\cap 2)} + \left(\widehat{\mathbf{x}}_{k|k} - \widehat{\mathbf{x}}_{k|k}^{(1\cap 2)}\right)\left(\widehat{\mathbf{x}}_{k|k} - \widehat{\mathbf{x}}_{k|k}^{(1\cap 2)}\right)^T\right]. \end{aligned} \tag{29}$$

The final component in the proposed framework is to compute the reference statistics, that is, $\widehat{\mathbf{x}}_k^{(S)}$ and $P_k^{(S)}$. Based on [18], state propagator is used to provide the required reference. More specifically, fused state estimate and its covariance matrix are propagated one time forward to form predicted estimates which are to be used as the reference signal. As a reference for the detection algorithm, we use $\widehat{\mathbf{x}}_k^{(S)}$ and $P_k^{(S)}$ which are transferred to local χ^2-test blocks.

To summarize, the proposed secure state estimation framework can be outlined as follows:

(S.1) In the first step, the "IMM-Predict" is implemented.

(S.2) In the second step, the "IMM-Update" will be implemented.

(S.3) Calculate the failure detection value $q_k^{(l)}$ using (18).

(S.4) In the fourth step, the probability that each sensor belongs to the attack mode is computed based on (20)-(21).

(S.5) In the fifth step, the adaptive weights associated with each sensor are computed via (22).

(S.6) In the sixth step, the second-order statistics based on each sensor is updated using (23)–(27).

(S.7) In the final step, the combined second-order statistics are computed via (28)-(29).

This completes development of the proposed framework. Next we present our simulation results to validate the effectiveness of the proposed multisensor attack detection/isolation fusion framework.

4. Experimental Results

This section presents our experimental simulations performed to evaluate the performance of the proposed framework against the aforementioned three type of data injection attacks, that is, constant attacks; time-varying attacks; stochastic attacks (possibly non-Gaussian) [19]. In this simulation experiment, we utilize sensory information from an integrated navigation system with including Global Navigation System (GPS), Strap-down Inertial Navigation System (SINS), and the Bei-Dou2 (BD2). In this integrated navigation system, the ψ-error model [20] is considered to present the evolution of the SINS state over time (state model). First-order Gauss-Markov process is utilized to model the accelerometer and gyroscope biases where time constants of τ are considered. The aforementioned model results in having state vector consisting of fifteen states (inertial states in position, velocity, attitude, accelerometer bias, and gyro bias). The monitoring sensors are the GPS and BD2. We use the position information received from the GPS and BD2 to rectify the SINS error.

In this experiment and in order to generate the trajectory of the aircraft and its associated inertial measurements, we use the "Inertial Navigation System toolbox" [21]. On the other hand and to generate GPS and BD2 positions, we use the "Satellite Navigation toolbox" [22]. Bias and power spectra of the SINS sensor are defined based on the following values: accelerometer bias: 50 μg; accelerometer white noise: 5 $\mu g/\sqrt{\text{Hz}}$; Gyro bias: 0.1 deg/hour, and; Gyro white noise: 0.001 deg/$\sqrt{\text{hour}}$. At the same time, the following measurement errors are utilized in performing the simulation experiment: GPS position error (longitude): 3.72 m; GPS position error (latitude): 3.98 m; GPS position error (vertical): 3.84 m; BD2 position error (longitude): 2.43 m; BD2 position error (latitude): 2.56 m; BD2 position error (vertical): 2.78 m. It is worth mentioning that these parameters are selected in order to simulate a real-world scenario. The transition probability matrix of the IMM filter (see (6)) is as follows: $\pi_{11} = \pi_{22} = 0.98$ and $\pi_{12} = \pi_{21} = 0.02$.

We introduce three type attacks into the GPS measurement as shown in Figure 2. The result of attack detection curve based on the proposed framework is illustrated

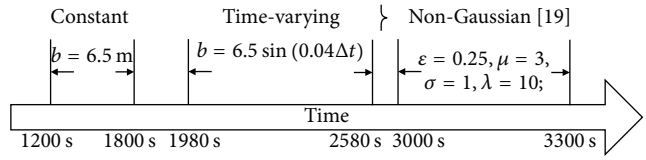

Figure 2: GPS attack timing sequence.

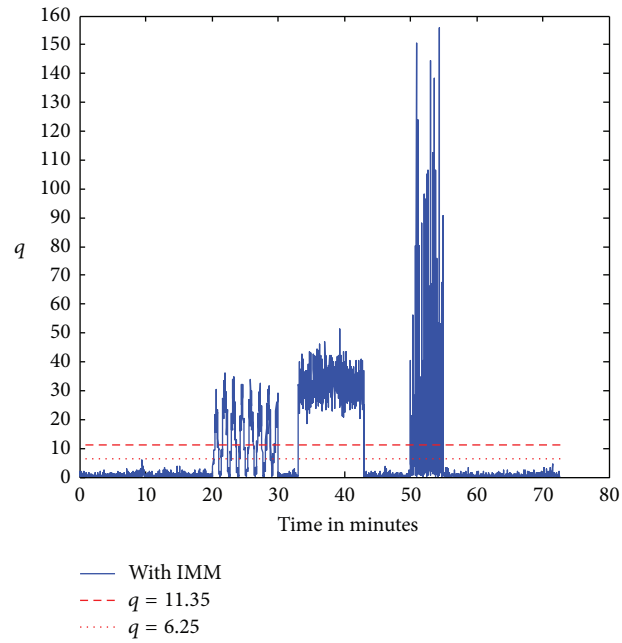

Figure 3: Attack detection curve together based on the proposed fuzzy attack detection/isolation fusion framework.

in Figure 3. Figure 3 illustrates that the proposed attack detection/isolation framework can detect constant and time-varying attacks abruptly and detect the stochastic non-Gaussian attack introduced at last reasonably well. The reason behind this behaviour, that is, the proposed framework is secure even against non-Gaussian attacks, is in its ability to adopt model probabilities to error measurement. Figure 4 illustrates the model probabilities associated with the attack and ideal behaviour modes. It is observed that the mode probabilities adopt to the attack scenario in an efficient fashion. Finally, Figure 5 illustrates the position error which shows that the proposed fusion framework keeps the error bounded and does not allow the estimation algorithm to diverge even under highly non-Gaussian attacks. This is a critical important property of the proposed framework from practical point of view.

5. Summary

In this paper, we proposed an improved and innovative secure state estimation framework which combines the IMM filter with a fuzzy-based attack isolation mechanism. In the proposed framework, we consider two separate behaviour

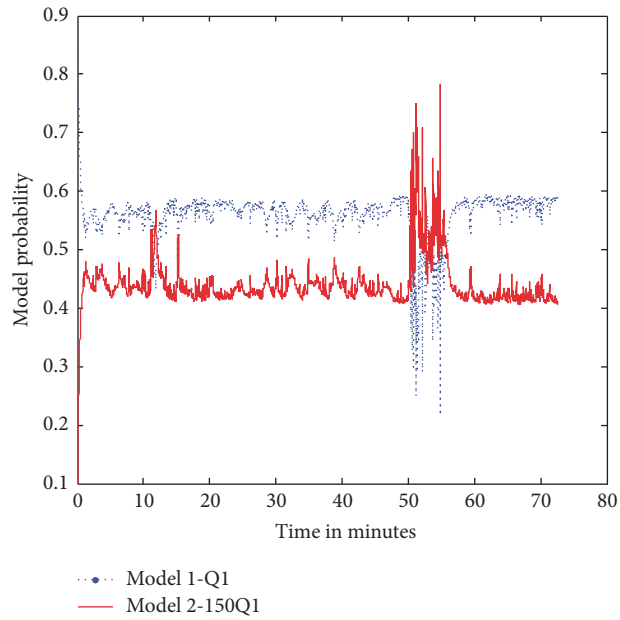

Figure 4: Model probabilities associated with the attack and ideal behaviour modes.

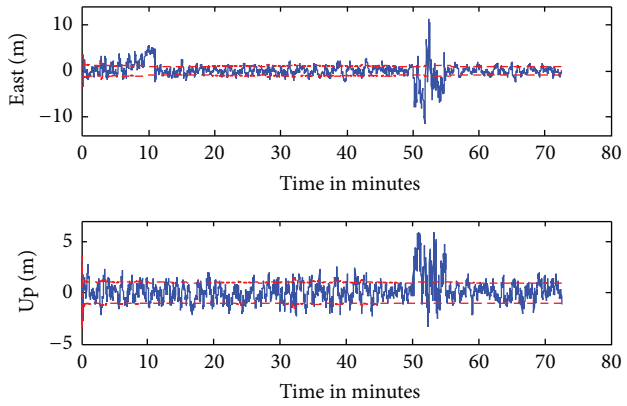

Figure 5: Position error obtained from the proposed framework.

modes, one associated with the ideal scenario and one with the attack case, where we compute adaptive weights via a modified observation update mechanism. In order to avoid utilization of attacked measurements and instead use the proper observation for updating the state estimates, local χ^2-tests are used for each modality and combined adaptively to form the global state estimates. Simulated experiments validated the effectiveness of the proposed attack detection/isolation framework.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partially supported by Natural Sciences & Engineering Research Council (NSERC) of Canada, Discovery Grant RGPIN-2016-049988.

References

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

[2] M. Pajic, R. Mangharam, O. Sokolsky, D. Arney, J. Goldman, and I. Lee, "Model-driven safety analysis of closed-loop medical systems," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 3–16, 2014.

[3] S. Xin, Q. Guo, H. Sun, B. Zhang, J. Wang, and C. Chen, "Cyber-physical modeling and cyber-contingency assessment of hierarchical control systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2375–2385, 2015.

[4] K. Sampigethaya and R. Poovendran, "Aviation cyber-physical systems: Foundations for future aircraft and air transport," *Proceedings of the IEEE*, vol. 101, no. 8, pp. 1834–1855, 2013.

[5] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, 2015.

[6] B. Martinez, X. Vilajosana, I. Vilajosana, and M. Dohler, "Lean Sensing: Exploiting Contextual Information for Most Energy-Efficient Sensing," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1156–1165, 2015.

[7] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: a data driven approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, 2015.

[8] J. Zhang, R. S. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1086–1101, 2015.

[9] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, 2015.

[10] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[11] S. Huang, K. K. Tan, and T. H. Lee, "Fault diagnosis and fault-tolerant control in linear drives using the Kalman filter," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4285–4292, 2012.

[12] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A Review of False Data Injection Attacks Against Modern Power Systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2017.

[13] J. Wei, "A data-driven cyber-physical detection and defense strategy against data integrity attacks in smart grid systems," in *Proceedings of the IEEE Global Conference on Signal and Information Processing, GlobalSIP 2015*, pp. 667–671, USA, December 2015.

[14] U. A. Khan and A. M. Stankovic, "Secure distributed estimation in cyber-physical systems," in *Proceedings of the 2013 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, pp. 5209–5213, Canada, May 2013.

[15] J. Lu and R. Niu, "Sparse attacking strategies in multi-sensor dynamic systems maximizing state estimation errors," in *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016*, pp. 3151–3155, China, March 2016.

[16] J. Lu, *Resilient dynamic state estimation in the presence of false information injection attacks*, ProQuest LLC, Ann Arbor, MI, 2016.

[17] J. Lu and R. Niu, "Malicious attacks on state estimation in multi-sensor dynamic systems," in *Proceedings of the 2014 IEEE International Workshop on Information Forensics and Security, WIFS 2014*, pp. 89–94, USA, December 2014.

[18] R. Da, "Failure detection of dynamical systems with the state chi-square test," *Journal of Guidance, Control, and Dynamics*, vol. 17, no. 2, pp. 271–277, 1994.

[19] K. N. Plataniotis, D. Androutsos, and A. N. Venetsanopoulos, "Nonlinear filtering of non-Gaussian noise," *Journal of Intelligent & Robotic Systems*, vol. 19, no. 2, pp. 207–231, 1997.

[20] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*, IET, Stevenage, UK, 2nd edition, 2004.

[21] INS TOOLBOX 3.0, "INS TOOLBOX 3.0," http://gpsoftnav.com/products/ins-toolbox-3-0/.

[22] SATNAV TOOLBOX 3.0 FOR MATLAB, "SATNAV TOOLBOX 3.0 FOR MATLAB," http://gpsoftnav.com/products/satellite-navigation-satnav-toolbox-3-0/.