Secure Particle Filtering for Cyber-Physical Systems with Binary Sensors under Multiple Attacks

Weihao Song, Zidong Wang, Jianan Wang, Fuad E. Alsaadi and Jayuan Shan

Abstract—This paper is concerned with the secure particle filtering problem for a class of discrete-time nonlinear cyber-physical systems with binary sensors in the presence of non-Gaussian noises and multiple malicious attacks. The multiple attacks launched by the adversaries, which take place in a random manner, include the denial-of-service attacks, the deception attacks and the flipping attacks. Three sequences of Bernoulli-distributed random variables with known probability distributions are employed to describe the characteristics of the random occurrence of the multiple attacks. The raw or corrupted measurements are transmitted to sensors whose outputs are binary according to engineering practice. A modified likelihood function is constructed to compensate for the influence of the randomly occurring multiple attacks by introducing the random occurrence probability information into the design process. Subsequently, a secure particle filter is proposed based on the constructed likelihood function. Finally, a moving target tracking application is elaborated to verify the viability of the proposed secure particle filtering algorithm.

Index Terms—Secure particle filtering, cyber-physical systems, binary sensors, randomly occurring attacks, target tracking.

I. INTRODUCTION

As an integrated system composed of cyber networks, physical components (e.g., sensors, controllers and monitors) and computation resources, the cyber-physical system (CPS) has become an emerging research frontier in the past few decades. Due to its significant advantages in reliability, autonomy and adaptability [7], the CPS has shown tremendous potential in practical applications of various public infrastructures such as smart grids [17] and transportation systems [37]. In [4], the CPS has been generally abstracted into the combination of a physical system and a controller, where the controller generates a control command based on the current estimate of the system state. In this sense, the proper functioning of the CPS is closely related to the performance of the chosen state estimation scheme. In fact, due to the importance of the state estimation problems, the last two decades have seen the development of a large quantity of estimation and filtering algorithms which include, but are not limited to, Kalman filtering [3], [16], extended Kalman filtering [13], [22], [35], unscented Kalman filtering [27], $H_\infty$ filtering [2], [18], [28], [34], moving-horizon estimation [26], [51], envelope-constrained filtering [29], $L_2$-$L_\infty$ filtering [32], and particle filtering [1], [20], [21] techniques. In particular, the particle filtering is one of the powerful tools in dealing with non-Gaussian noises in the filtering problems.

The CPSs are known to be vulnerable to miscellaneous security threats in both physical layers and cyber layers due primarily to their massive components and the demanding communications among different components [15]. Generally speaking, it is not an easy work to model the attacks in a unified and accurate way owing to the cunning/intelligence of the adversaries. Therefore, a great deal of research attention has been focused on the filtering/control problem of the CPSs subject to specific malicious attacks including denial-of-service attacks [24], [25], [46], deception attacks [12], [39], [47], replay attacks [41] and many more. It should be noted that the malicious attacks initiated by the adversaries cannot be always successful on account of the deployment of the security software and protection equipment. As a result, the malicious attacks in most of the existing literature are actually referred to as randomly occurred/succeed attacks. For example, in [42], the event-triggered active disturbance rejection control problem has been addressed for systems suffering from both denial-of-service attacks and physical attacks, where the randomly occurring denial-of-service attacks are characterized by the Gilbert-Elliott model. In [36], the security-guaranteed filtering scheme has been developed for delayed systems in the presence of randomly occurring sensor saturations and deception attacks, where the occurrence characteristics of the deception attacks are described by the Bernoulli process.

Apart from the security threats, the scarce resources (e.g. limited energy capacity and network bandwidth) constitute another critical issue of the CPSs due to the massive information exchange among the components [10], [14], [44]. In order to utilize the limited resources in an efficient way, considerable research effort has recently been devoted to the so-called event-triggered communication mechanism [8], [19], [38], [50], under which the data exchange is executed only when a predefined event occurs, thereby reducing the frequency of data transmissions and mitigating the network burden [11], [23]. Nevertheless, the data to be transmitted (if triggered) may still exceed the packet length restriction in some cases. An alternative approach to dealing with the data-intensive problem is to use the binary sensor whose outputs are simply binary values representing switches, contacts, and pins.
In this case, only the binary values need to be transmitted to the fusion center and the network traffic is much reduced.

Owing to their merits of low cost and simple installation, binary sensors have been welcomed in industry and have also been paid a great deal of research attention from academic communities, see e.g., [48] and the references therein. The typical binary sensors include the industrial sensors for pressure/gas/liquid monitoring, and the medical sensors with binary outcomes, to name just a few [45]. So far, in the context of filter/estimation, two kinds of particle filtering algorithms have been developed in [9] based on the data from a group of binary sensors to track a target. In [45], the fusion estimation scheme has been presented for a class of linear time-varying systems subject to bounded noises by exploiting the information at the sign switching instant of the binary signal. It should be pointed out that the binary decisions are prone to be overheard and deliberately flipped by the adversaries during the data transmission. Such kind of cyber-attacks, if not addressed well, may deteriorate the estimation performance and even paralyze the whole CPS.

Summarizing the above discussions, there appears to be a lack of systematic investigation on the secure particle filtering problem for a class of nonlinear/non-Gaussian CPSs with binary sensors subject to randomly occurring multiple attacks. As such, the primary aim of this paper is to narrow such a gap by means of designing a secure particle filtering algorithm with certain robustness to the multiple attacks in both physical layers and cyber layers. It is worth noticing that the addressed filtering problem is by no means straightforward due mainly to the technical challenges identified as follows: 1) how to establish a unified framework to take into account the simultaneous presence of denial-of-service attacks, deception attacks and flipping attacks in the measurement model? 2) how to deal with the analytical complexity induced by the random nature of the multiple attacks and the binary (hence sparse) signal from binary sensors? and 3) how to attenuate the effect from the multiple attacks on the filtering performance in the filter design?

The main contributions of this paper can be highlighted as threefold: 1) the secure filtering problem is investigated for a class of general nonlinear/non-Gaussian CPSs with binary sensors; 2) a comprehensive yet realistic measurement model is presented to simultaneously take into account the randomly occurring denial-of-service attacks, deception attacks and flipping attacks; and 3) a secure particle filtering algorithm is developed by establishing a modified likelihood function to compensate for the effect of the multiple malicious attacks.

The remainder of this paper is structured as follows. Section II formulates the secure filtering problem with binary sensors and gives some preliminaries about the particle filtering scheme. In Section III, the secure particle filtering algorithm which deals with the randomly occurring multiple attacks is developed by establishing a modified likelihood function. A two-dimensional moving target tracking problem is considered in Section IV to demonstrate the effectiveness and practicality of our proposed secure filtering algorithm. Eventually, some conclusions are presented in Section V.

Notation. Throughout this paper, the notation exploited is fairly normative. \( \mathbb{R}^n \) stands for the \( n \)-dimensional Euclidean vector space. The superscript \( T \) means the transpose operation. \( \text{diag}\{a_1, a_2, \ldots, a_n\} \) denotes a diagonal matrix with \( a_1, a_2, \ldots, a_n \) being its diagonal elements. \( p_k(\cdot) \) stands for the probability density function of a stochastic variable \( x \), i.e., \( x \sim p_k(\cdot) \), and \( cdff_x(\cdot) \) denotes the corresponding cumulative distribution function. \( \Pr\{X\} \) represents the occurrence probability of a discrete event \( X \). \( \mathbb{E}(x|z) \) denotes the mathematical expectation of \( x \) conditional on \( z \). \( \mathcal{N}(x; u, \Sigma) \) denotes the Gaussian probability density function of stochastic variable \( x \) with mean and covariance being \( u \) and \( \Sigma \), respectively. \( x_{k,l} \) is the path of \( x \) from time instant \( k \) to time instant \( l \). Other notations will be introduced when needed.

II. PROBLEM FORMULATION AND PRELIMINARIES

A. System setup

Consider a class of discrete-time nonlinear systems characterized by the following model:

\[
x_{k+1} = f(x_k) + \omega_k
\]

where \( x_k \in \mathbb{R}^n \) denotes the system state at time instant \( k \) and \( f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) represents the nonlinear state evolution function. \( \omega_k \in \mathbb{R}^n \) is the process noise satisfying \( p_{\omega_k}(\cdot) \). The measurement model of the \( s \)-th sensor is given by

\[
\bar{y}_k^s = h^s(x_k) + \nu_k^s, \quad s = 1, 2, \ldots, S
\]

where \( \bar{y}_k^s \in \mathbb{R} \) represents the measurement output of the \( s \)-th sensor at time instant \( k \) and \( h^s(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R} \) is the measurement function. \( \nu_k^s \in \mathbb{R} \) is the measurement noise on the \( s \)-th sensor satisfying \( p_{\nu_k^s}(\cdot) \).

In this paper, we assume that the measurement process is prone to attacks launched by the malicious attackers. That is to say, the actual measurements of the sensors may be falsified by the randomly occurring denial-of-service attacks or deception attacks, which are characterized by

\[
\bar{y}_k^s = (1 - \phi_k^s)(\bar{y}_k^s + \varphi_k^s)
\]

where \( \bar{y}_k^s \) is the falsified measurement of the \( s \)-th compromised sensor and \( \phi_k^s \) denotes the deception attack launched by the attacker given by

\[
\phi_k^s = -\bar{y}_k^s + \mu_k^s
\]

where \( \mu_k^s \) represents a random deception signal satisfying \( p_{\mu_k^s}(\cdot) \). The stochastic variables \( \phi_k^s \) and \( \varphi_k^s \) are assumed to be mutually independent Bernoulli-distributed white sequences, which take values on \( 0 \) and \( 1 \) with the following mathematical probabilities:

\[
\begin{align*}
\Pr\{\phi_k^s = 1\} &= \bar{\phi}^s \\
\Pr\{\phi_k^s = 0\} &= 1 - \bar{\phi}^s
\end{align*}
\]

and

\[
\begin{align*}
\Pr\{\varphi_k^s = 1\} &= \bar{\varphi}^s \\
\Pr\{\varphi_k^s = 0\} &= 1 - \bar{\varphi}^s
\end{align*}
\]

where \( \bar{\phi}^s \in [0, 1] \) and \( \bar{\varphi}^s \in [0, 1] \) are both known constants referred to as the success rates of the initiated denial-of-service attacks and deception attacks, respectively.
Next, the \( s \)th sensor processes its measurement according to:

\[
\theta^*_k = \begin{cases} 
1, & \text{if } y^*_k > T_s \\
0, & \text{otherwise}
\end{cases}
\] (5)

where \( T_s \) is a known threshold, and sends only binary value \( \theta^*_k \) to the fusion center via the wireless transmission channels where the cyber-attacker is able to flip the binary information.

By introducing another Bernoulli-distributed stochastic variable \( \alpha^*_k \), the eventually received signal contributed by the \( s \)th sensor at the fusion center is of the following form [9]:

\[
y^*_k = \tau^* \bar{\theta}^*_k + \epsilon^*_k
\] (6)

where \( \tau^* \) is the channel gain coefficient corresponding to the \( s \)th sensor, \( \epsilon^*_k \) is the channel noise satisfying \( p_{\epsilon^*_k}(\cdot) \) and

\[
\bar{\theta}^*_k = (1 - \alpha^*_k)\theta^*_k + \alpha^*_k(1 - \theta^*_k)
\] (7)

where

\[
\begin{cases} 
Pr\{\alpha^*_k = 1\} = \bar{\alpha}^* \\
Pr\{\alpha^*_k = 0\} = 1 - \bar{\alpha}^*
\end{cases}
\]

with \( \bar{\alpha}^* \in [0, 1] \) being the probability that the cyber-attacker successfully flips the binary information.

For notational brevity, all the received signals at the fusion center up to time instant \( k \) are denoted as:

\[
y^{1:S}_{1:k} = [y^1_{1:k}, y^2_{1:k}, \ldots, y^S_{1:k}]^T.
\]

**Remark 1**: For the addressed nonlinear/non-Gaussian CPSs, the measurement signals may be intentionally compromised by the adversaries in both physical layers and cyber layers. Note that the malicious attacks launched by the adversaries in both layers are less likely to work at all times due probably to the complicated network environment and the defender’s security protection. Therefore, it is reasonable to consider that the attacks occur in a random way [5], [33]. To model the randomness of the successful attacks in the physical layers, Bernoulli-distributed stochastic variables \( \phi^*_k \) and \( \varphi^*_k \) are introduced in (3). To be more specific, if \( \phi^*_k = 0 \) and \( \varphi^*_k = 0 \), the measurement process of the \( s \)th sensor is normal; if \( \phi^*_k = 0 \) and \( \varphi^*_k = 1 \), the \( s \)th sensor is successfully attacked in the form of deception attack; if \( \phi^*_k = 1 \) and \( \varphi^*_k = 1 \), the \( s \)th sensor is hijacked by the adversary and the measurement service is unavailable, i.e., only the reading “0” can be output. Similarly, the Bernoulli distributed stochastic variable \( \alpha^*_k \) is adopted in the cyber layers. It is evident from (7) that the binary signal is free from the malicious attacks during the transmission when \( \alpha^*_k = 0 \), and the binary signal is deliberately flipped by the adversary when \( \alpha^*_k = 1 \). It should be mentioned that the considered model in [9] can be regarded as a special case of our work when \( \phi^*_k = 0 \), \( \varphi^*_k = 0 \) and \( \alpha^*_k = 0 \).

**Remark 2**: It should be noted that the selection of threshold \( T_s \) is of practical importance. After the threshold is set, the output of the binary sensor will be determined accordingly, and a slight change of the threshold might cause a huge change of the measurement output. In practical applications, the selection of the threshold is closely related to the sensing principle or specific task. For example, in the sensor networks with limited sensing range, the target of interest can be detected and observed only when it moves into the sensing region of the sensors [30]. In this case, the threshold depends on the physical constraints of the sensors and the sensors can simply output the binary values to imply whether they detect the target or not. On the other hand, if the measurement output serves a specific task (e.g. camera-based surveillance within campus), the threshold is determined as a reasonable similarity of the pedestrian activity to the predefined suspicious behaviors.

Throughout this paper, we make the following two assumptions.

**Assumption 1**: The process noise \( \omega_k \), the measurement noise \( \nu^*_k \), the random deception signal \( \mu^*_k \) and the channel noise \( \varepsilon^*_k \) are mutually independent and also independent of the initial state \( x_0 \) that has the prior probability density function \( p_{x_0}(\cdot) \).

**Assumption 2**: The nonlinear functions \( f(\cdot) \) and \( h(\cdot) \) as well as the probability density functions \( p_{\omega_k}(\cdot), p_{\nu^*_k}(\cdot), p_{\mu^*_k}(\cdot) \) and \( p_{\varepsilon^*_k}(\cdot) \) are all known.

### B. Preliminaries

The key issue in sequential Bayesian filtering problem is to calculate the posterior probability density function \( p(x_k | y^{1:S}_{1:k}) \), based on which we can obtain the minimum mean-square error (MMSE) estimate for the state \( x_k \) as:

\[
\hat{x}_k^{\text{MMSE}} = \mathbb{E}\{x_k | y^{1:S}_{1:k}\} = \int x_k p(x_k | y^{1:S}_{1:k}) dx_k.
\] (8)

The posterior probability density function \( p(x_k | y^{1:S}_{1:k}) \) can be recursively derived as follows:

\[
\begin{align*}
 p(x_k | y^{1:S}_{1:k}) &= p(x_k | x_{k-1}) p(x_{k-1} | y^{1:S}_{1:k-1}) dx_{k-1}, \\
 p(x_k | y^{1:S}_{1:k-1}) &= \int p(y^{1:S}_{1:k} | x_k) p(x_k | y^{1:S}_{1:k-1}) dx_k.
\end{align*}
\] (9)

However, the closed-form expression of \( p(x_k | y^{1:S}_{1:k}) \) is generally unavailable except for some special cases, e.g., the linear and Gaussian systems. Fortunately, the sequential Monte Carlo method (i.e., particle filtering) [1] can provide an
approximation of $p(x_k|y_1:1;S)$ by a set of weighted particles \( \{x_k^m, w_k^m\}_{m=1}^M \) as
\[
p(x_k|y_1:1;S) = \sum_{m=1}^{M} w_k^m \delta(x_k - x_k^m),
\]
and then we obtain
\[
x_k^{\text{MMSE}} = \sum_{m=1}^{M} w_k^m x_k^m
\]
where \( M \) is the number of particles, \( \delta(\cdot) \) is the Dirac delta function, \( x_k^m \) is sampled from a proposal distribution \( q(x_k|x_{k-1}^m, y_1:1;S) \), and the corresponding weight \( w_k^m \) is computed by
\[
w_k^m = w_{k-1}^m \frac{p(y_1:1;S|x_k^m)p(x_k^m|x_{k-1}^m)}{q(x_k^m|x_{k-1}^m, y_1:1;S)}.
\]

The purpose of this paper is to design a secure particle filtering algorithm for a class of nonlinear/non-Gaussian CPSs with binary sensors under multiple attacks such that the MMSE estimate of the state \( x_k \) is obtained at the fusion center using the compromised measurement signals up to time instant \( k \), i.e., \( y_1:1;S_k \).

III. SECURE PARTICLE FILTERING ALGORITHM DESIGN

In this section, we investigate the secure particle filter design problem for a class of CPSs formulated in Section II-A. In fact, if we only consider the systems described by (1) and (2) in a safe environment, the estimation objective can be directly achieved by virtue of the standard particle filtering algorithm (e.g. the sampling importance resampling particle filter). However, the vulnerability of the CPSs to the malicious attacks in both physical layers and cyber layers renders the standard particle filtering scheme inapplicable, and there is an urgent need to develop a dedicated filter algorithm that can resist the cyber-attacks with satisfactory filtering accuracy.

The following theorem provides a solution to the secure particle filter design problem by giving an explicit expression of the modified likelihood function to assist in updating the importance weights.

**Theorem 1:** Consider the measurement model described by (2), the randomly occurring denial-of-service attack/deception attack model characterized by (3)-(4) and the binary transmission scheme given by (5)-(7). The modified likelihood function evaluated at \( x_k^m \), which is employed to update the corresponding importance weight at the fusion center, is given by
\[
p(y_k^{1:S}|x_k^m) = \sum_{s=1}^{S} \left\{ (1 - \tilde{\phi}^s)(1 - \tilde{\varphi}^s)p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, x_k^m) + \tilde{\varphi}^s p(y_k^s|\phi_k^s = 0, \varphi_k^s = 1, x_k^m) + \tilde{\phi}^s p(y_k^s|\phi_k^s = 1, x_k^m) \right\}
\]
where
\[
p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, x_k^m) = p_c\{\tilde{y}_k^s = 0, \tilde{\varphi}_k^s = 0, \tilde{\phi}_k^s = 1, x_k^m\} \]
and, similarly, we obtain
\[
p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, x_k^m) = p_c\{\tilde{y}_k^s = 0, \tilde{\varphi}_k^s = 0, \tilde{\phi}_k^s = 0, x_k^m\}(1 - \tilde{\phi}^s),
\]
On the other hand, it is clear from (3)-(7) that the actually received signal \( y_k^s \) from the \( s \)th sensor at the fusion center depends on the Bernoulli-distributed stochastic variables \( \phi_k^s \), \( \varphi_k^s \) and \( \alpha_k^s \), as well as the threshold parameter \( T_\delta \). To proceed with the proof, we will derive the expression of the likelihood function in the following three cases.

**Case 1:** \( \phi_k^s = 0 \) and \( \varphi_k^s = 0 \).

In this case, we have
\[
y_k^s = \tilde{y}_k^s,
\]
and it is straightforward to obtain from the law of total probability that
\[
p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, x_k^m) = p_c\{\tilde{y}_k^s = 1, \tilde{\phi}_k^s = 0, \varphi_k^s = 0, x_k^m\}
\]
and, similarly, we obtain
\[
p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, x_k^m) = p_c\{\tilde{y}_k^s = 1, \tilde{\phi}_k^s = 0, \varphi_k^s = 0, x_k^m\}\alpha^s
\]
and, similarly, we obtain
\[
p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, x_k^m) = p_c\{\tilde{y}_k^s = 0, \tilde{\phi}_k^s = 0, \varphi_k^s = 0, x_k^m\}(1 - \tilde{\phi}^s),
\]

Proof: Based on Assumption 1, we have
\[
p(y_k^{1:S}|x_k^m) = \prod_{s=1}^{S} p(y_k^s|x_k^m).
\]

In the sequel, we discuss each term on the right-hand side of the above equation. According to (7), we have
\[
\text{Pr}\{\tilde{\theta}_k^s = 1|\phi_k^s = 0, \varphi_k^s = 0, x_k^m\} = p_c\{\tilde{y}_k^s = 1, \tilde{\phi}_k^s = 0, \varphi_k^s = 0, x_k^m\}
\]
and, similarly, we obtain
\[
\text{Pr}\{\tilde{\theta}_k^s = 0|\phi_k^s = 0, \varphi_k^s = 0, x_k^m\} = p_c\{\tilde{y}_k^s = 1, \tilde{\phi}_k^s = 0, \varphi_k^s = 0, x_k^m\}\alpha^s
\]
where
\[ \Pr \{ \theta_k^e = 1 | \phi_k^e = 0, \varphi_k^e = 0, x_k^m \} = \Pr \{ \hat{y}_k^e > T_\delta | x_k^m \} = \Pr \{ h_k(x_k^m) + \nu_k > T_\delta \} = 1 - cdf_{\nu_k}(T_\delta - h_k(x_k^m)) \] (22)

and
\[ \Pr \{ \theta_k^e = 0 | \phi_k^e = 0, \varphi_k^e = 0, x_k^m \} = cdf_{\nu_k}(T_\delta - h_k(x_k^m)). \] (23)

Furthermore, it can be observed from (6) that
\[ p(y_k^e | \theta_k^e = 1, \phi_k^e = 0, \varphi_k^e = 0, x_k^m) = p(z_k^e \epsilon x_k^m - \tau^a) \]
\[ p(y_k^e | \theta_k^e = 0, \phi_k^e = 0, \varphi_k^e = 0, x_k^m) = p(z_k^e | \hat{y}_k^e). \] (24)

Then, we arrive at (14) by substituting (20)-(24) into (19).

**Case 2:** \( \phi_k^e = 0 \) and \( \varphi_k^e = 1 \).

In this case, the deception attack is successfully launched by the adversary and we know that
\[ \hat{y}_k^e = \mu_k^e. \] (25)

After some similar manipulations as those in **Case 1**, it is easy to obtain (15).

**Case 3:** \( \phi_k^e = 1 \).

In this case, the sth sensor is hijacked by the adversary and only the reading of “0” can be output. Without loss of generality, we assume that \( T_\delta \) is a positive scalar. Then, similar to the previous cases, we can have (16).

According to the law of total probability, we write the likelihood function \( p(y_k^e | x_k^m) \) associated with the sth sensor as follows:
\[ p(y_k^e | x_k^m) = p(y_k^e | \phi_k^e = 0, x_k^m) + p(y_k^e | \phi_k^e = 1 | x_k^m) \]
\[ = p(y_k^e | \phi_k^e = 0, x_k^m) \Pr \{ \phi_k^e = 0 \} + p(y_k^e | \phi_k^e = 1, x_k^m) \Pr \{ \phi_k^e = 1 \} \]
\[ = (1 - \delta^e) p(y_k^e | \hat{y}_k^e = 0, x_k^m) + \delta^e p(y_k^e | \hat{y}_k^e = 1, x_k^m) \]
\[ = (1 - \delta^e) [ (1 - \delta^e) p(y_k^e | \hat{y}_k^e = 0, \varphi_k^e = 0, x_k^m) \]
\[ + \delta^e p(y_k^e | \hat{y}_k^e = 1, x_k^m) ] \]
\[ + \delta^e p(y_k^e | \hat{y}_k^e = 0, \varphi_k^e = 1, x_k^m) \] (26)

It follows from (17) and (26) that the modified likelihood function of particle \( x_k^m \) at the fusion center can be calculated by (13), which completes the proof.

Now, we are in a position to design the secure particle filtering algorithm, whose main purpose is to get the particle-based representation of the posterior probability density function sequentially. In other words, we aim to obtain the particle-based representation of \( p(x_k | y_{1:k}^{1:S}) \) as shown in (10) given that of \( p(x_{k-1} | y_{1:k-1}^{1:S}) \).

Let a set of weighted particles \( \{ x_m^m | x_k^m \} \) to approximate \( p(x_k | y_{1:k}^{1:S}) \). If we choose the state transition probability density function \( p(x_k | x_{k-1}) \) as a proposal density \( q(x_k | x_{k-1}, y_{1:k}^{1:S}) \), then the particles at time instant \( k \) are sampled as \( x_k^m \sim p(x_k | x_{k-1}) \) [1]. As such, when we obtain the measurement signals contaminated by the randomly occurring multiple attacks, we can update the importance weight \( w_k^m \) associated with particle \( x_k^m \) according to
\[ w_k^m = w_{k-1}^m p(y_{1:k}^e | x_k^m). \] Meanwhile, in order to mitigate the phenomenon of particle degeneracy during the iterative update of particles, the resampling strategy is added at each iteration by removing the particles with negligible weights and duplicating the particles with significant weights [1]. It should be noted that, even though we design the secure particle filtering algorithm in the framework of the sampling importance resampling particle filter, extensions to other types of particle filters (e.g. auxiliary particle filter [31]) are fairly straightforward.

In summary, the pseudo-code of the secure particle filtering algorithm for the CPSs with binary sensors subject to multiple attacks is provided in Algorithm 1.

**Remark 3:** So far, we have addressed the secure filtering problem for a class of nonlinear/non-Gaussian CPSs with binary sensors in the framework of sequential Bayesian estimation. The available information \( y_{1:k}^{1:S} \) at the fusion center has been employed in the proposed filter. To compensate for the effect of the malicious attacks on the filtering performance, the probability information of the randomly occurring attacks has been taken into account in the process of filter design. A modified likelihood function has been explicitly constructed in (13) to update the importance weights. In this sense, the developed particle filtering algorithm has certain robustness against the randomly occurring denial-of-service attacks, deception attacks and flipping attacks. Note that, if the probability information of the randomly occurring attacks is not available, one could employ an online detector to detect the random attacks at each time instant, which, however, might be time- and cost-consuming. In fact, it is an interesting yet challenging task to design an efficient secure filtering scheme under the random occurring multiple attacks without prior statistics, which would be one of the promising research topics.

**Remark 4:** The filtering problem for CPSs under cyber-attacks has been extensively studied in the literature. Our main results distinguish from existing ones in the following three aspects: 1) the secure filtering problem addressed is new in the sense that the CPS is nonlinear, the underlying noises are allowed to be non-Gaussian and the sensor outputs are binary; 2) the model for malicious attacks is new as it takes three kinds of random occurring attacks (denial-of-service attacks, deception attacks and flipping attacks) into simultaneous consideration; and 3) the developed secure particle filtering algorithm with a modified likelihood function is able to compensate for the effect of the multiple malicious attacks.

## IV. Simulation Results

In this section, a practical application to the moving target tracking is presented to demonstrate the usefulness of our proposed secure particle filtering algorithm.

### A. Moving target tracking scenario

Consider the moving target tracking problem in a two-dimensional (2-D) Cartesian coordinate system. The mathe-
Algorithm 1 Secure particle filtering algorithm for the CPSs with binary sensors subject to multiple attacks

1: Initialization: Draw $M$ particles from the prior density, i.e., $x_0^m \sim p_{x_0} (\cdot), m = 1, 2, \ldots, M$ and set the corresponding importance weights $w_0^m$ as $\frac{1}{M}$. The maximum recursive time instant is chosen as $K$.

2: for $k = 1, 2, \ldots, K$ do
3:  for $m = 1, 2, \ldots, M$ do
4:      Step 1: Importance sampling
5:      Sample particle $\tilde{x}_k^m$ from the transition probability density function $p(x_k|x_{k-1})$.
6:      Step 2: Measurement update
7:      Collect all the compromised sensor signals $y_k^{1:S}$ at the fusion center.
8:      Step 3: Importance weight calculation
9:      Calculate the unnormalized importance weights $\{\bar{w}_k^m\}_{m=1}^M$ according to
10:     $\bar{w}_k^m = w_{k-1}^m \sum_{s=1}^S (1 - \bar{\phi}^s)(1 - \bar{\varphi}^s)$
11:        $\times p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, \bar{x}_k^m)$
12:        $+ \bar{\phi}^s p(y_k^s|\phi_k^s = 0, \varphi_k^s = 1, \bar{x}_k^m)$
13:        $+ \bar{\varphi}^s p(y_k^s|\phi_k^s = 1, \bar{x}_k^m)$,
14: where $p(y_k^s|\phi_k^s = 0, \varphi_k^s = 0, \bar{x}_k^m)$, $p(y_k^s|\phi_k^s = 0, \varphi_k^s = 1, \bar{x}_k^m)$ and $p(y_k^s|\phi_k^s = 1, \bar{x}_k^m)$ are defined in (14)-(16), respectively.
15: end for
16: for $m = 1, 2, \ldots, M$ do
17:      Step 4: Weight normalization
18:      Normalize the importance weights according to
19:      $\bar{w}_k^m = \frac{w_k^m}{\sum_{m=1}^M w_k^m}$.
20:      Step 5: State estimate extraction
21:      Update the MMSE estimate of state $x_k$ and the corresponding estimation error covariance as
22:      $\hat{x}_k = \sum_{m=1}^M \bar{w}_k^m \tilde{x}_k^m$,
23:      $\hat{P}_k = \sum_{m=1}^M \bar{w}_k^m (\tilde{x}_k^m - \hat{x}_k)(\tilde{x}_k^m - \hat{x}_k)^T$.
24:      Step 6: Resampling
25:      Resample new particle $x_k^m$ from the distribution $\sum_{m=1}^M w_k^m \delta(x_k - \tilde{x}_k^m)$.
26: end for
27: end for

matematical model for the target movement, adopted from [6], is expressed as:

$$x_{k+1} = \begin{bmatrix} 1 & t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + \omega_k$$  \hspace{1cm} (27)$$

where $x_k$ is specified by

$$\begin{bmatrix} \phi_{tar}^x \phi_{tar}^y \phi_{tar}^x \phi_{tar}^y \end{bmatrix}^T$$

is the state vector of the moving target at time instant $k$, which determines the target position $(\phi_{tar}^x, \phi_{tar}^y)$ and velocity $(\phi_{tar}^x, \phi_{tar}^y)$ in the 2-D plane. $t$ stands for the sampling period and $\omega$ is the zero-mean Gaussian white noise with covariance matrix $\text{Cov}_w$ defined as follows:

$$\text{Cov}_w = \begin{bmatrix} \frac{\sigma^2}{2} & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \frac{\sigma^2}{2} \end{bmatrix}$$

where $\sigma$ denotes the acceleration variance.

For the purpose of target tracking, $S$ binary sensors are deployed in the surveillance areas to detect and receive the energy produced by the moving target of interest. At time instant $k$, the measurement at the $s$th sensor is described by (2), where the measurement function is written as [9]

$$h^s(x_k) = \Upsilon \begin{bmatrix} d_0 \\ \|I_{x,s}^k - I_{\phi_{tar}^x, \phi_{tar}^y}^k\| \end{bmatrix} \lambda^s$$

where $\Upsilon$ denotes the produced energy by the target at a reference distance $d_0$, $(\phi_{tar}^x, \phi_{tar}^y)$ represents the location of the $s$th sensor (we assume that the information of the sensor locations is available to the fusion center) and $\lambda^s$ is a known environment-dependent propagation loss parameter of the $s$th sensor.

The measurement noise $\nu^s_k$ on the $s$th sensor is represented by a two-component Gaussian mixture model, i.e.,

$$p(\nu^s_k) = (1 - \beta^s)\mathcal{N}(\nu^s_k; \mu^s_1, \Sigma^1) + \beta^s\mathcal{N}(\nu^s_k; \mu^s_2, \Sigma^2)$$

where $\beta^s$ is the glint probability. In addition, the channel noise $\varepsilon^s_k$ associated with the $s$th sensor is assumed to be zero-mean Gaussian white noise with variance $(\sigma^2)^2$ and the deception signal $\mu^s_k$ satisfies a uniform distribution over the interval $[a, b]$.

Once the measurement process is completed, each sensor compares the obtained measurements with the predefined threshold parameter and only a single binary digit is transmitted to the fusion center. The above-mentioned processes are, of course, prone to be attacked by the adversaries and the corresponding parameters are given in Section IV-C.

B. Performance metric

The root mean square error (RMSE) on the position and velocity estimates averaged over $N$ Monte Carlo trials are
selected as the performance metrics in our work to assess the tracking performance, which are respectively defined by

\[
\text{RMSE}_{p,k} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( (\hat{x}_{i,k} - \bar{x}_{i,k})^2 + (\hat{y}_{i,k} - \bar{y}_{i,k})^2 \right)},
\]

\[
\text{RMSE}_{v,k} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( (\hat{v}_{x,i,k} - \bar{v}_{x,i,k})^2 + (\hat{v}_{y,i,k} - \bar{v}_{y,i,k})^2 \right)},
\]

where the subscripts \( p, k \) and \( v, k \) indicate, respectively, the position and velocity, \((\hat{x}_{i,k}, \hat{y}_{i,k})\) and \((\hat{v}_{x,i,k}, \hat{v}_{y,i,k})\) respectively represent the realization of \((x_{i,k}, y_{i,k})\) and \((v_{x,k}, v_{y,k})\) in the \( i \)th Monte Carlo trial, and their estimates are respectively given by \((\bar{x}_{i,k}, \bar{y}_{i,k})\) and \((\bar{v}_{x,i,k}, \bar{v}_{y,i,k})\).

C. Common simulation parameters

In the simulation, the moving target is observed by \( S = 16 \) binary sensors, whose positions are depicted in Fig. 2. The target trajectories are simulated by setting initial state \( x_0 = [15, 0.4, 20, 0.3]^T \), sampling period \( t = 1 \), and acceleration variance \( \Xi = 0.045^2 \). To sample the particles in the initialization step, a procedure adopted from [6] is employed. To be more specific, the position components are directly sampled from a Gaussian prior distribution with mean \([15, 20]^T\) and covariance matrix \(\text{diag}\{100, 100\}\), and the velocity components are indirectly sampled from a Gaussian prior distribution with mean \([0.5, \arctan(3/4)]^T\) and covariance matrix \(\text{diag}\{0.25^2, (\pi/6)^2\}\) by noting that the prior knowledge of the resultant velocity and the azimuth is more common in practice.

The number of particles is \( M = 500 \) and \( N = 50 \) different realizations are conducted for the Monte Carlo simulations. Other parameter setups related to the binary sensors and the randomly occurring attacks are presented in TABLE I.

![Fig. 2: One realization of the target trajectory and its estimate obtained from our proposed Sec-PF. The blue diamonds denote the positions of the binary sensors.](image-url)

![Fig. 3: The measurements of the sensor locating at (10, 10) before transmitted and the corresponding binary values.](image-url)

D. Simulation results and discussions

One realization of the target trajectory and the estimated trajectory obtained from the proposed secure particle filtering algorithm (abbreviated as Sec-PF) are presented in Fig. 2, from which we can see that the trajectory estimated by the Sec-PF is close to the true trajectory of the moving target. For the binary sensor locating at (10, 10), Fig. 3 displays its measurements corrupted by the randomly occurring denial-of-service attacks/deception attacks and the corresponding binary values subject to the randomly occurring flipping attacks.

In the next simulations, we aim to compare the tracking performance under the following three scenarios: (i) tracking with Sec-PF; (ii) tracking with the standard particle filtering algorithm but neglecting the effect of the randomly occurring attacks (abbreviated as Sta-PF-Neg); and (iii) tracking with the standard particle filtering algorithm using the uncorrupted measurement signals (abbreviated as Sta-PF and used as a benchmark). The behaviors of the RMSEs on position and velocity estimates obtained from the above-mentioned three algorithms are compared in Figs. 4-5, respectively. We observe that the Sec-PF is able to provide the estimates that are close to the Sta-PF, while the Sta-PF-Neg performs the worst with the highest estimation errors. As expected, our proposed Sec-PF possesses certain robustness against the randomly occurring multiple attacks.

In order to investigate the impact of the randomly occurring multiple attacks on the tracking performance, three groups of simulations are further conducted with different occurrence
probabilities of attacks. In each group, only one parameter varies and the others remain unchanged. The corresponding simulation results are plotted in Figs. 6-11, which indicates that the occurrence probabilities of attacks (i.e., $\bar{\phi}^s$, $\bar{\phi}^a$ and $\bar{\alpha}^s$) do have a significant effect on the tracking performance. We figure out that, as the occurrence probabilities of attacks increase, the tracking performance will gradually degrade.

In addition, we conduct further simulations to compare the average running time for Steps 1-6 at each time instant, average RMSEs on position estimates, and average RMSEs on velocity estimates with different numbers of particles. The corresponding simulation results (obtained on a PC with 2.50 GHz CPU) are summarized in TABLE II. It is clear that, the increase of the number of particles will usually improve the filtering performance at the cost of higher average running time. As such, the designers/operators should consider the real-world engineering specifications (e.g. the sampling period) and choose a proper number of particles to attain a balance between the computational burden and the filtering performance.
TABLE II: Performance comparisons with different numbers of particles.

<table>
<thead>
<tr>
<th>M</th>
<th>200</th>
<th>400</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average running time (s)</td>
<td>0.0196</td>
<td>0.0387</td>
<td>0.0584</td>
</tr>
<tr>
<td>Average RMSEs on position estimates</td>
<td>3.7008</td>
<td>3.5354</td>
<td>3.3978</td>
</tr>
<tr>
<td>Average RMSEs on velocity estimates</td>
<td>0.1741</td>
<td>0.1699</td>
<td>0.1643</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

In this paper, we have addressed the secure particle filter design problem for a class of nonlinear/non-Gaussian CPSs with binary sensors subject to the randomly occurring multiple attacks. Three Bernoulli-distributed random variables with known probabilities have been introduced to describe the randomly occurring denial-of-service attacks, deception attacks and flipping attacks, respectively. In order to mitigate the impact of the malicious attacks launched by adversaries on the filtering performance, we have made an effort to establish a modified likelihood function in which the occurrence probabilities of the multiple attacks have been fully exploited. Based on the theoretical analysis, a secure particle filtering algorithm has been developed and applied for the moving target tracking. The Monte Carlo simulation results have been presented to elucidate the usefulness of the developed secure particle filtering algorithm. In the future, our research topics would focus on the secure filtering problem for more complicated scenarios, such as the distributed denial-of-service attacks [40], redundant channels [49], and the conic-type nonlinear Markov jump systems [43].

REFERENCES


D. Zhao, Z. Wang, G. Wei and Q.-L. Han, A dynamic event-triggered approach to observer-based PID security control subject to deception attacks, *Automatica*, vol. 120, art. no. 109128, 2020.


Weihao Song received the B.S. degree in flight vehicle design and engineering from Beijing Institute of Technology, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree in aeronautical and astronautical science and technology with Beijing Institute of Technology, Beijing, China. From May 2019 to May 2020, he was a Visiting Scholar with the Department of Computer Science, Brunel University London, London, U.K. His research interests include Bayesian state estimation, distributed state estimation, nonlinear filtering, and networked control systems.

Zidong Wang (SM’03–F’14) was born in Jiangsu, China, in 1966. He received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, and the M.Sc. degree in applied mathematics in 1990 and the Ph.D. degree in electrical engineering in 1994, both from Nanjing University of Science and Technology, Nanjing, China. He is currently a Professor of Dynamical Systems and Computing in the Department of Computer Science, Brunel University London, U.K. From 1990 to 2002, he held teaching and research appointments in universities in China, Germany and the UK. Prof. Wang’s research interests include dynamical systems, signal processing, bioinformatics, control theory and applications. He has published around 600 papers in refereed international journals. He is a holder of the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, William Mong Visiting Research Fellowship of Hong Kong.

Prof. Wang serves (or has served) as the Editor-in-Chief for International Journal of Systems Science, the Editor-in-Chief for Neurocomputing, and an Associate Editor for 12 international journals including IEEE Transactions on Automatic Control, IEEE Transactions on Control Systems Technology, IEEE Transactions on Neural Networks, IEEE Transactions on Signal Processing, and IEEE Transactions on Systems, Man, and Cybernetics-Part C. He is a Member of the Academia Europaea, a Fellow of the IEEE, a Fellow of the Royal Statistical Society and a member of program committee for many international conferences.

Jiayuan Shan received the B.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 1988, and the M.S. and Ph.D. degrees from Beijing Institute of Technology, Beijing, China, in 1991 and 1999, respectively. He is currently a Professor with the School of Aerospace Engineering, Beijing Institute of Technology. His research interests include guidance, navigation and control of the aircraft and hardware-in-the-loop simulation. He is the Principal Professor in the direction of Flight Dynamics and Control.

Jianan Wang (Senior Member, IEEE) received the B.S. and M.S. degrees in control theory and engineering from Beijing Jiaotong University and Beijing Institute of Technology, Beijing, China, in 2004 and 2007, respectively. He received the Ph.D. degree in aerospace engineering from Mississippi State University, Starkville, MS, USA, in 2011. He is currently an Associate Professor with the School of Aerospace Engineering, Beijing Institute of Technology. His research interests include cooperative control of multiple dynamic systems, UAV formation control, cooperative guidance, and estimation of sensor networks. He is a senior member of IEEE and AIAA.

Fuad E. Alsaadi received the B.S. and M.Sc. degrees in electronic and communication from King AbdulAziz University, Jeddah, Saudi Arabia, in 1996 and 2002. He then received the Ph.D. degree in Optical Wireless Communication Systems from the University of Leeds, Leeds, UK, in 2011. Between 1996 and 2005, he worked in Jeddah as a communication instructor in the College of Electronics & Communication. He is currently an associate professor of the Electrical and Computer Engineering Department within the Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. He published widely in the top IEEE communications conferences and journals and has received the Carter award, University of Leeds for the best PhD. He has research interests in optical systems and networks, signal processing, synchronization and systems design.