TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Yadav, S., Kishor, N., Purwar, S., & Rajkumar, V. (2026). State Partition-Particle Filter Detection for Cyber-Physical Attacks. *IEEE Transactions on Information Forensics and Security*, *21*, 3450-3462. https://doi.org/10.1109/TIFS.2026.3671018

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

State Partition-Particle Filter Detection for Cyber-Physical Attacks

Seema Yadav, *Student Member, IEEE*, Nand Kishor, *Senior Member, IEEE*, Shubhi Purwar, and Vetrivel Rajkumar

***Abstract*—The growth in adoption of communication technologies in the power system results in an increase in its vulnerability towards cyberattacks. This paper presents a novel state partition-particle filter (SP-PF) based detection algorithm that dynamically adapts to varying operating conditions. The algorithm partitions state variables into size-restricted blocks, effectively grouping highly correlated variables to enhance computational efficiency and detection accuracy. Our approach consists of two main steps: (i) state-partition estimation of variables and (ii) detection based on likelihood conditions. The proposed detection algorithm was tested in a real-time cyber-physical environment using a real-time digital simulator (RTDS) in hardware-in-loop configuration with PMUs and a synchronization clock, all connected via standard TCP/UDP protocols. Experimental results demonstrate successful detection of false data injection attacks, replay attacks, and hybrid attacks under various operating conditions. Comparative analysis with extended Kalman filter shows that our approach achieves significantly improved accuracy in state estimation with reduced mean square error, enhancing the overall robustness of the detection mechanism.**

***Index Terms*—Cyber-physical security, detection, hardware-in-loop, particle filter, real-time, state estimation, state-partition.**

I. Introduction

THE modern power system has evolved significantly with the integration of communication technologies, creating advanced smart grids with the continuous evolution and control capabilities. However, this integration has substantially increased vulnerability to cyber-physical attacks (CPAs) [1], [2]. As power systems continue to evolve and expand, the dynamic nature of these systems creates significant variations in operating conditions, making it difficult for traditional detection methods to adapt without frequent recalibration. Furthermore, the complex interaction between cyber components and physical elements introduces unique attack vectors that can bypass conventional security measures without noticeable immediate impact. These challenges require development of more adaptive and robust detection approaches.

To address these challenges, we propose a novel state partition-particle filter (SP-PF) based detection approach for CPAs in power systems. This approach effectively partitions state variables into smaller, correlated blocks, enabling more precise detection while dynamically adapting to varying operating conditions [3], [4].

The proposed approach offers several significant contributions to the field of power system security. The scalability of the said technique is examined on various IEEE test systems which were modelled on RSCAD software. First, we develop a detection technique that dynamically adjusts threshold values and state partitions in real-time, adapting to varying operating conditions. This universal, model-free approach makes the detection algorithm proactive, capable of detecting even complex hybrid cyber-attacks. Second, we implement a simplified detection index computed in time-domain as a function of time-varying states, rather than complex statistical indices. This design minimizes detection delay, critical for power system security. Third, we conduct a comprehensive hardware-in-loop (HIL) study using real-time digital simulator (RTDS), providing a realistic test setup environment that enhances algorithm reliability in practical power system scenarios. Finally, mentioned algorithm demonstrates capability to detect unknown and complex CPAs, with significantly improved state estimation accuracy resulting in reduced mean square error, enhancing robustness and responsiveness of the detection mechanism.

The rest of the paper is organized as follows: Section II reviews related work and technical background. Section III presents the power network model and state estimation fundamentals. Section IV details the SP-PF detection framework. Section V describes the real-time simulation setup and algorithm implementation. Section VI presents results and discussion, while Section VII concludes the paper.

II. Related Work

This section provides a comprehensive review of existing approaches, analysing their technical foundations, advantages, and limitations within the broader context of power system resilience.

A. Attack Taxonomy in Power Systems

FDIAs [5], [6], [7] manipulate measurement data to mislead state estimation while evading bad data detection algorithms.

Received 11 September 2024; revised 22 April 2025, 19 November 2025, and 26 January 2026; accepted 16 February 2026. Date of publication 9 March 2026; date of current version 3 April 2026. This work was supported by European Union's Horizon 2020 Research and Innovation Program under Grant 870620. The associate editor coordinating the review of this article and approving it for publication was Dr. Daisuke Mashima. *(Corresponding author: Nand Kishor.)*

Seema Yadav and Shubhi Purwar are with the Electrical Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Prayagraj 211004, India (e-mail: seemayadaven1041@gmail.com; shubhi@mnnit.ac.in).

Nand Kishor is with the Department of Engineering Sciences, Østfold University College, 1671 Fredrikstad, Norway (e-mail: nand.kishor@hiof.no).

Vetrivel Rajkumar was with the Intelligent Electrical Power Grid Laboratory, TU Delft University, 2628 CD Delft, The Netherlands. He is now with TenneT TSO B.V., 6812 AR Arnhem, The Netherlands (e-mail: v.subramaniamrajkumar@tudelft.nl).

Digital Object Identifier 10.1109/TIFS.2026.3671018

1556-6021 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

These attacks can be mathematically represented as,

$$\tilde{z} = z + a, \tag{1}$$

where z represents the corrupted measurements, z denotes the original measurements, and a is the attack vector designed such that $\|a\| \leq \tau$, with τ being the detection threshold.

Replay attacks [8], [9] record legitimate system measurements and replay them during an attack, creating a deceptive perception of normal operation. These attacks exploit temporal patterns and can be formulated as,

$$z(k) = z(k - T), \tag{2}$$

where T represents the time offset between original and replayed measurements.

Switching attacks [10], [11] target control components by manipulating switching signals, causing oscillations between operating modes. The system dynamics under these attacks can be represented as a switched system,

$$\dot{\mathrm{x}} = \{f_1(x, u)\ if\ \sigma(t) = 1, f_2(x, u)\ if\ \sigma(t) = 2\}, \tag{3}$$

where $\sigma(t)$ is the switching signal controlled by the attacker.

Hybrid attacks [12] combine multiple attack strategies, making detection particularly challenging. These sophisticated attacks often leverage domain knowledge of both cyber and physical aspects of power systems to maximize impact while minimizing detectability.

B. Rule-Based Detection Methods

Rule-based detection methods rely on predefined criteria and thresholds derived from domain expertise to identify anomalous system behaviour [13]. These approaches establish deterministic rules based on power system operational knowledge and security constraints. The authors in [14] have proposed signature-based techniques identify known attack patterns through predefined rules that match specific characteristics of malicious activities, such as unusual command sequences or invalid data ranges. For example, logical consistency rules verify that circuit breaker status aligns with measured power flows, flagging inconsistencies as potential attack indicators.

Specification-based approaches compare system behavior against formal specifications of normal operation, detecting deviations that violate expected patterns [15]. These methods typically implement if-then-else logic structures with fixed or adaptive thresholds to classify measurements as normal or anomalous. While rule-based methods offer high interpretability, computational efficiency, and rapid detection for known attack patterns, they struggle with detecting zero-day attacks and require continuous knowledge base updates as new threats emerge [16]. Their effectiveness diminishes against sophisticated attackers capable of mimicking legitimate system behavior while staying within operational threshold boundaries.

C. Model-Based Detection Approaches

Kalman Filter Variants: Model-based techniques using Kalman filtering variants [17], [18], [19], [20] form the foundation of many detection approaches. The EKF [18], [21] linearizes system dynamics around the current state estimate through first-order Taylor series approximation,

$$\hat{x}(k \mid k-1) = f(\hat{x}(k-1 \mid k-1), u(k-1)), \tag{4}$$

$$P(k \mid k-1) = F(k-1) P(k-1 \mid k-1) F(k-1)^{\mathrm{T}} + Q, \tag{5}$$

where $F(k-1) = \partial f / \partial x | \hat{x} = x(k-1|k-1)$, represents the Jacobian matrix of partial derivatives.

While effective for mildly nonlinear systems, EKF introduces significant approximation errors under the strong nonlinearities characteristic of cyber-attacks. However, it is suggested in [19] to implement deterministic sampling through the unscented transformation:

$$\hat{x}(k \mid k-1) = \Sigma_{\mathrm{i}} w_{\mathrm{i}}^{\mathrm{m}} \chi_{\mathrm{i}}(k|k-1), \tag{6}$$

where χ_{i} represents sigma points capturing the probability distribution, and $w_{\mathrm{i}}^{\mathrm{m}}$ are corresponding weights.

The iterative Kalman kilter (IKF) [22] recursively refines state estimates through multiple iterations within a single time step:

$$\hat{x}_{J+1}(k|) = \hat{x}(k|k-1) + K(z(k) - h(\hat{x}_{\mathrm{j}}(k|k))). \tag{7}$$

where j denotes the iteration index, and K represents the Kalman gain. Recent advances include the robust cubature Kalman filter [23], which maintains stability under attack-induced uncertainties by integrating cubature integration rules with robust estimation principles. Despite these advancements, Kalman-based approaches remain constrained by Gaussian distribution assumptions and computational complexity when scaling to large power systems.

Statistical Detection Methods: Statistical detection method employs hypothesis testing [24], [25] and likelihood ratio tests [8] to identify anomalous system behavior. The authors in [5], and [12], accumulates deviations from expected measurements,

$$S(k) = max(0, S(k-1) + (r(k) - \mu 0)/\sigma - \kappa), \tag{8}$$

where $r(k)$ represents the residual, μ_0 the expected mean under normal conditions, σ the standard deviation, and κ the drift parameter controlling detection sensitivity.

The Rao-CUSUM test [24] extends this approach by incorporating normalized statistics to handle unknown post-attack distributions. Similarly, maximum likelihood estimation approaches [8] optimize detection by comparing probability densities under normal and attack hypotheses,

$$L(z_1, \ldots, z_{\mathrm{n}}) = p(z_1, \ldots, z_{\mathrm{n}}|H_1)/p(z_1, \ldots, z_{\mathrm{n}}|H_0). \tag{9}$$

These methods offer mathematical rigor but typically assume prior knowledge of attack distribution parameters. They struggle to adapt to evolving attack strategies with unknown statistical properties or attacks deliberately designed to maintain residuals below detection thresholds.

D. Machine Learning-Based Detection Methods

Machine learning-based detection combine supervised, unsupervised, and reinforcement learning approaches for attack detection. Deep belief network [26] uses hierarchical

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

feature through stacked Boltzmann machines to extract complex patterns. CNNs [27] exploit spatial correlations while RNNs [28], [29] capture temporal dependencies for time-series anomaly detection, mathematically represented as,

$$y = \sigma(W_2\sigma(W^1 x + b_1) + b_2), \tag{10}$$

These supervised methods, though effective for known patterns, require extensive labeled data, limiting generalizability to novel attacks. Unsupervised techniques [30] address this by identifying anomalies as deviations from learned normal behavior. Autoencoders measure reconstruction error,

$$e = \|x - D(E(x))\|), \tag{11}$$

to detect attacks. Reinforcement learning [31], [32] formulates detection as sequential decision processes,

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma max Q(s',a') - Q(s,a)]. \tag{12}$$

while graph-based methods [6], [33] leverage network structures for precise intrusion detection. Although these approaches handle complex nonlinear relationships, they face interpretability limitations and challenges in providing theoretical performance guarantees.

E. Particle Filter-Based Approaches

Particle filters (PFs) [4], [34] employ non-parametric Bayesian estimation, representing probability distributions through weighted particles that update their weights based on measurements, rather than parametric models. This monte carlo approach approximates the posterior distribution as:

$$p(x_k|z_{1:k}) \approx \Sigma_i w_i \delta(x_k - x_{ik}), \tag{13}$$

where w_i represents the importance weight of particle x_{ik}, and δ is the dirac delta function.

The recursive Bayesian estimation proceeds in two steps:

1. Prediction: $p(x_k|z_{1:(k-1)} = \int p\left(x_k \mid x_{\{k-1\}}\right) p\left(x_{\{k-1\}} \mid z_{1:(k-1)}\right) dx_{\{k-1\}}$,
2. Update: $p(x_k|z_{1:k} \propto p(z_k|x_k)p(x_k|z_{1:(k-1)})$,

This approach offers distinct advantages for handling non-Gaussian distributions and highly nonlinear systems compared to Kalman-based methods. Various resampling methods [35] address particle degeneracy by eliminating low-weight particles,

$$N_{eff} = 1/\Sigma_i(w_i)^2, \tag{14}$$

Multiple model particle filters [36] attempt to handle switching dynamics by maintaining parallel filter banks for different system modes. However, these approaches still struggle with the "curse of dimensionality," requiring exponentially more particles as the state dimension increases.

In paper [22], introduced a state-partition particle filter specifically for switching attack detection, demonstrating improved performance through strategic state space partitioning,

$$x_k = [x_k^1, x_k^2, \ldots, x_k^M]. \tag{15}$$

where each partition x_k^j contains highly correlated state variables processed by separate filters.

This approach built upon earlier state partitioning concepts [4], [11] by specifically adapting them to power system attack detection challenges. Their initial exploration showed promising results for switching attacks but lacked comprehensive evaluation across diverse attack vectors and operating conditions. This analysis reveals several critical research gaps:

1) Most detection methods lack validation in real-time HIL environments with actual power system dynamics, limiting confidence in their operational performance.

2) Existing methods frequently struggle to maintain detection performance across varying operating conditions without manual recalibration.

3) Many approaches face significant computational challenges when scaled to realistic power networks, particularly for non-parametric methods like particle filters.

4) Most model-based approaches make restrictive Gaussian assumptions that fail to capture the complex distributions introduced by sophisticated attacks.

The novelty of proposed SP-PF approach is that (i) it directly address these gaps by combining the non-parametric advantages of particle filtering with the computational efficiency of state partitioning, while providing comprehensive real-time validation across multiple attack scenarios and operating conditions, and (ii) it introduces an adaptive KL divergence-based state partitioning optimization that dynamically determines when to merge similar regions or split complex regions based on statistical information content, enabling the algorithm to automatically adjust to changing system dynamics during both normal operation and under attack conditions. To properly implement this SP-PF approach, a robust understanding of power system dynamics and state estimation is essential. The following section presents the power network model and state estimation fundamentals that form the mathematical foundation upon which our detection framework is built.

III. State Estimation for Power Network

The physical power grid is a nonlinear network with its states constantly time-varying. As such a static system model and applying conventional static state estimation are not effective enough to represent dynamics. The power network with synchronous generator connected to each bus can be considered as a dynamic model. Each generator is connected to its associated control systems (excitation systems and governors), while the power network components (buses, lines, loads) are represented using algebraic equations under steady-state assumptions. This differential-algebraic equation (DAE) approach allows for efficient computation while preserving dynamic behavior were essential. By disregarding the rapid sub-transient dynamics and the effects of saturation, the model of synchronous machine can be succinctly outlined as fourth-order differential equations within the dq reference frame, as in eqn. (16) [22].

$$\dot{\delta} = \omega_0 \Delta\omega, \tag{16a}$$

$$\Delta\dot{\omega} = \left(\frac{1}{2H}\right)(T_m - T_e - K_D\Delta\omega), \tag{16b}$$

$$\dot{e}'_q = \frac{1}{T'_{d0}}\left(E_{fd} - e'_q - \left(x_d - x'_d\right) i_d\right), \tag{16c}$$

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

$$\dot{e}'_d = \frac{1}{T'_{q0}}\left(-e'_d - \left(x_q - x'_q\right) i_q\right). \tag{16d}$$

where, $(\dot{})$ denotes the time derivative, the symbols δ and $\Delta\omega$ represent the rotor angle in radians and rotor speed deviation of synchronous machine, respectively. ω_0 denotes the rated angular frequency of synchronous machine. H signifies the inertia constant, T_m mechanical torque, T_e represents the electric air-gap torque, and K_D denotes the damping factor. Additionally, e'_d and e'_q denote the transient voltage along q and d axes of rotating reference frame, respectively. T'_{d0} and T'_{q0} are the open-circuit time constants along d and q axes, respectively. x_d and x_q represent the synchronous reactance at the d and q axes, respectively while x'_d and x'_q represent the transient reactance at the d and q axes, respectively. E_{fd} represents internal field voltage.

To simplify the notation, eqn. (16) is converted into general nonlinear eqns. (17) and (18a):

$$\dot{x} = f(x,u) + w, \tag{17a}$$
$$y = h(x,u) + v, \tag{17b}$$

The state vector x and input vector u are defined as:

$$x = \left[\delta\ \Delta\omega\ e'_q\ e'_d\right]^T, \tag{18a}$$
$$u = \left[T_m\ E_{fd}\ i_R\ i_I\right]^T, \tag{18b}$$

Considering each bus is equipped with PMU, which can measure the terminal voltage phasor ($E_t = e_R + je_I$) and current phasor ($I_t = i_R + ji_I$) in synchronously rotating $R-I$ reference frame [22].

$$y = \left[e_R\ e_I\right]^T, \tag{18c}$$

Rotor current along d axis and q axis are i_d and i_q defined by,

$$i_d = i_R sin\delta - i_I cos\delta, \tag{18d}$$
$$i_q = i_I sin\delta + i_R cos\delta, \tag{18e}$$

Eqns. (18d) and (18e) depict the relationship between $R-I$ and $d-q$ reference frames,

$$e_R = (e'_d + i_q x'_q)\sin\delta + (e'_q - i_d x'_d)\cos\delta, \tag{18f}$$
$$e_I = (e'_q - i_d x'_d)\sin\delta - (e'_d + i_q x'_q), \tag{18g}$$

In eqn. (17), x denotes the state vector of (4×1) dimension, u represents the input vector of (4×1), and $y(2\times 1)$ stands for the output vector. The functions $f(\cdot)$ and $h(\cdot)$ correspond to the state transition and output functions of (4×1) and (2×1) dimensions, respectively. In eqn. (18a), e_R and e_I denote the stator voltages, as provided in eqns. (18f) and (18g) [11]. The vectors w and v signify the process and measurement noise, respectively. The process noise w represents the uncertainty in the system model and accounts for unmodeled dynamics, linearization errors, and random disturbances affecting the state transitions. The measurement noise v captures the errors and inaccuracies in sensor readings, quantization effects, and other sources of uncertainty in the observation process. Both noise processes are modeled as zero-mean Gaussian white noise with covariance matrices Q and R, respectively, as defined in eqns. (19a) and (19b), where $E[\cdot]$ denotes statistical expectation.

$$Q = E\left[ww^T\right]. \tag{19a}$$
$$R = E\left[vv^T\right]. \tag{19b}$$

The SP-PF algorithm has to be applied on discrete measurements, necessitating the conversion of the continuous model eqns. (17)-(18a) into a discrete model, as given in eqn. (20a).

A. Discretization

Given measurements of $T_m(j\Delta t), E_{fd}(j\Delta t)$, *and* $E_t(j\Delta t)$ for $j = 1,2,\ldots,k$, the objective is to estimate the states of synchronous machine at time $k\Delta t$, denoted as $\delta(k\Delta t)$, $\Delta\omega(k\Delta t)$, $e'_q(k\Delta t)$, and $e'_d(k\Delta t)$, where the subscript indicates the time at $k\Delta t$.

$$x_k = f(x_{k-1}, u_{k-1}) + w_{k-1}, \tag{20a}$$
$$y_k = h(x_k, u_k) + v_k, \tag{20b}$$

The Euler method [11] is utilized to discretize the general state-space (20a), which represents the state equations of the power system network, in each time step. Here, $\tilde{f}$ can be computed using eqns. (22) & (23). Additionally, w_{k-1} denotes the discrete measurement noise.

$$x_k \approx x_{k-1} + \tilde{f}\Delta t + w_{k-1}, \tag{21}$$
$$\tilde{f} = (f(\tilde{x}_k, u_k) + f(x_{k-1}, u_k))/2, \tag{22}$$

where,

$$\tilde{x}_k = x_{k-1} + \Delta t \cdot f(x_{k-1}, u_{k-1}). \tag{23}$$
$$Q_d \triangleq E(w_{k-1} w_{k-1}^T). \tag{24}$$

Output function (19b) can be discretized as (10), with v_k as the discrete process noise.

$$y_k = h(x_k, u_k) + v_k. \tag{25}$$
$$R_d \triangleq E(v_k v_k^T). \tag{26}$$

The CPA detection algorithm using SP-PF is discussed in the next section.

IV. State Partition-Particle Filtering and Detection

The increasing integration of communication technologies in power systems inevitably expands their attack surface, heightening vulnerability to sophisticated CPAs. Traditional state estimation methods rely heavily on Gaussian noise assumptions and linear system models, while phasor measurement data in real-world power systems often exhibits non-Gaussian characteristics. This novel SP-PF addresses these fundamental limitations through adaptive state-space partitioning that identifies and exploits correlation structures in high-dimensional state spaces, and non-parametric Bayesian inference that eliminates restrictive distributional assumptions. Unlike conventional approaches, SP-PF provides theoretical convergence guarantees even under non-Gaussian attack-induced conditions while reducing the computational complexity from exponential to near-linear with respect to state dimensionality, making real-time implementation feasible for practical power systems.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

A. State Partition-Particle Filtering

The SP-PF methodology represents a comprehensive approach to CPA detection in power networks, where the algorithm strategically partitions system state variables into smaller, highly correlated groups, enhancing computational efficiency while maintaining detection precision.

The SP-PF algorithm initializes by generating particles based on an initial distribution, with large variance matrices used when prior information is limited to reflect low confidence in initial estimates. The algorithm's core innovation lies in decomposing each state into nominal and residual components given in step 4, separating inherent nonlinear system behaviour from perturbations around the nominal trajectory, which enables efficient linearization and improves tracking accuracy compared to linearizing around a single global estimate. The nominal state propagation follows the full nonlinear system dynamics, while linearization around the nominal state creates locally valid linear approximations tailored to each particle's location in state space.

Unlike conventional PFs that require $O(Nd)$ particles for an N-dimensional state space with accuracy d, this partitioning approach achieves comparable performance with only $O(N{\cdot}p{\cdot}d)$ particles, where p is the average partition size, resulting in substantial computational savings for high-dimensional power systems. The prediction and update stages implement Bayesian sequential estimation, with the subfilter gain calculation representing an adaptive learning rate that balances prior knowledge against new measurements. The covariance updates maintain statistical consistency by accounting for process and measurement uncertainties, providing mathematically rigorous bounds on estimation error even under adversarial conditions.

The algorithm addresses particle degeneracy through the effective sample size metric and resampling, which implements importance sampling from the posterior distribution by discarding uninformative particles and duplicating highly weighted ones. The residual generation directly relates to attack detection by computing the difference between observed measurements and predicted outputs based on state estimates, with this residual capturing anomalies induced by CAs. The likelihood ratio provides a statistically optimal test for attack detection, with decisions based on a predefined threshold τ.

1. Proof of Convergence for SP-PF: We can prove that the SP-PF algorithm converges to the true posterior distribution as the number of particles approaches infinity. Consider the estimation error $\varepsilon = |E[g(x_k)|y_{1:k}] - \hat{\mathrm{E}}[g(x_k)|y_{1:k}]|$, where g is any bounded function, E is the true expectation, and $\hat{\mathrm{E}}$ is the SP-PF approximation. For each partition j with N_j particles, the error bound is given by,

$$\varepsilon_j \leq C_j/\sqrt{N_j}. \tag{27}$$

where C_j is a constant dependent on the partition characteristics. For the full state estimation with M partitions, the error is bounded by,

$$\varepsilon \leq \sum\nolimits_{=1\ldots M} w_j \cdot \varepsilon_j$$

Algorithm 1 State Partition-Particle Filter (SP-PF)

I/p: PMU measurements (ω_{ref} and δ_{ref}), initial state distribution, number of particles N_s

O/p Estimated states, likelihood ratio for detection

1 Initialize the parameters of particle sets $x_k^{(i)}, i \in [1, N_s]$ based on initial distribution $\mathcal{N}(x_k^{(0)}, R^{(0)})$;
2 **while** not converge
3 Generate marginal posterior based distribution using eqn. $p(x_k|u_{1:k}, y_{1:k}) \triangleq \sum_{i=1}^{N_s} q_k^{(i)} \delta(x - x_k^{(i)})$;
4 Decompose each state into nominal and residual components $x_k^{(i)} \triangleq x_{nk}^{(i)} + x_{rk}^{(i)}$;
5 Propagate nominal state through system dynamic, $x_{nk}^{(i)} = f(x_{n(k-1)}^{(i)}) + w_{n(k-1)}^{(i)}$;
6 Linearize around operating point using eqns. $x_k^{(i)} \approx f\left(x_n^{(i)}\right) + F\left(x_{n(k-1)}^{(i)}\right) \times \left[x_{k-1} - x_{n(k-1)}^{(i)}\right] + w_{k-1}$, $= f\left(x_{n(k-1)}^{(i)}\right) + F\left(x_{n(k-1)}^{(i)}\right) x_{r(k-1)}^{(i)} + w_{k-1}$.
where $F\left(x_{nk}^{(i)}\right) = \frac{\partial f(x)}{\partial x} |_{x=x_{nk}^{(i)}}$ $x_{rk}^{(i)} \approx F\left(x_{n(k-1)}^{(i)}\right) \cdot x_{r(k-1)}^{(i)} + w_{k-1} - w_{n(k-1)}^{(i)}$ $y_k^{(i)} \approx h\left(x_{nk}^{(i)}\right) + H\left(x_{nk}^{(i)}\right) \cdot x_{rk}^{(i)} + v^{(i)}$;
where $H\left(x_{nk}^{(i)}\right) = \frac{\partial h(x)}{\partial x} |_{x=x_{nk}^{(i)}}$
7 Predict residual state $\hat{x}_{r(k|k-1)}^{(i)} = F\left(\hat{x}_{n(k-1)}^{(i)}\right) \cdot \hat{x}_{r(k-1|k-1))}^{(i)} - w_{n(k-1)}^{(i)}$;
8 Calculate subfilter gain $G_k^{(i)} = R_{r(k|k-1)}^{(i)} H\left(x_{nk}^{(i)}\right)^T \left[H\left(x_{nk}^{(i)} R_{r(k|k-1)}^{(i)} H\left(x_{nk}^{(i)}\right)^T + R_d\right]^{-1}$;
9 Update state estimate, $\hat{x}_{r(k|k)}^{(i)} = \hat{x}_{r(k|k-1)}^{(i)} + G_k^{(i)} \cdot \tilde{y}_k^{(i)}$;
10 Compute variance updates using eqns. $R_{r(k|k-1)}^{(i)} = F\left(x_{n(k-1)}^{(i)}\right) R_{(k-1|k-1)}^{(i)} \times F\left(x_{n(k-1)}^{(i)}\right)^T + \sigma\omega^2$, $R_{r(k|k)}^{(i)} = \left[I - G_k^{(i)} H\left(x_{nk}^{(i)}\right)\right] R_{(k|k-1)}^{(i)}$;
11 Calculate state estimate, $\hat{x}_{rk} = \sum_{i=1}^{N_s} q_k^{(i)} \hat{x}_{r(k|k)}^{(i)}$, $R_{r(k|k)} = \sum_{i=1}^{N_s} R_{rk}\left(x_{rk}^{(i)} - \hat{x}_{rk}^{(i)}\right)\left(x_k^{(i)} - \hat{x}_k\right)^T$;
12 Calculate effective sample size, $N_{eff} = \frac{1}{\sum_{i=1}^{N_s}\left(q_k^{(i)}\right)^2}$
13 If $N_{eff} < threshold$, perform resampling $\sum_{i=1}^{N_s} R_{rk} \delta(x_{rk}^{(i)} - \hat{x}_{rk}^{(i)})$;
14 Generate residual for detection, $r_k = y_k - h(x_k, u_k) \approx H(x_k)\, e_k + v_k + R_d$;
15 Generate likelihood ratio $L(Xk)$ for detection, $L(x_k) = p(x_1, \ldots, x_k \,|\, H_1) / p(x_1, \ldots, x_k | H_0)$;
16 **if** $L(x_k) > \tau$
17 Declare "Attack detected";
18 **else**
19 Declare "No attack";
20 **end**
21 **end**

$$\leq \sum\nolimits_{j=1..M} w_j \cdot \frac{C_j}{\sqrt{N_j}} \leq \frac{max_j\left(C_j\right)}{\sqrt{min_j\left(N_j\right)}}. \tag{28}$$

where w_j are partition weights that sum to 1. As $Nj \to \infty$ for all j, $\varepsilon \to 0$, establishing convergence. Furthermore, the partitioning strategy ensures that C_j values are minimized by grouping correlated states, providing faster convergence rates

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

than standard particle filters, especially in high-dimensional spaces.

Threshold (τ) Calculation: The detection threshold τ is dynamically calculated based on the statistical properties of the residual under normal operating conditions. For optimal detection performance, we derive τ as,

$$\tau = min\{r_k^T Q_d^{-1} r_k\}. \tag{29}$$

where r_k is the measurement residual vector and Q_d is the process noise covariance matrix. Under normal operation (H_0), the residual follows a multivariate Gaussian distribution, $r_k \sim \mathcal{N}(0, \Sigma r)$ where Σr is the residual covariance matrix. According to statistical detection theory, the optimal threshold can be derived from the log-likelihood ratio test,

$$r_k = \frac{1}{L(x_k)+1} \sum_{i=0}^{N_s} r_{k-1}^{(i)}, \tag{30}$$

$$= \frac{1}{L(x_k)+1} \sum_{i=0}^{N_s} H(x_k)\, e_k^{(i)} + v_k + R_d, \tag{31}$$

This can be further expressed as,

$$r_k = r'_k + r'_{v,k}. \tag{32}$$

where,

$$r'_{v,k} = \frac{1}{L(x_k)+1} \sum_{i=0}^{N_s} H(x_k)\, e_k^{(i)} + v_k. \tag{33}$$

$$r'_k = \frac{1}{L(x_k)+1} \sum_{i=0}^{N_s} R_d. \tag{34}$$

For the system under study, we define,

$$H_1 = diag(0, L(x_k)), H_0$$
$$= diag(0, L(x_k)), y_k \sim N(e_R, e_I) \tag{35}$$

where $L(x_k)$ represents the nominal angular frequency (377 rad/s) modulated by the function $\emptyset$ of the residual state. This modulation ensures that the threshold adapts to varying operating conditions.

The process noise covariance matrix Q_d is calibrated based on empirical state variations,

$$Q_d = diag(\sigma_\delta{}^2, \sigma_\omega{}^2), \tag{36}$$

where σ_δ^2 represents the variance of rotor angle states and σ_ω^2 represents the variance of rotor speed states. For this specific system, these values are computed as,

σ_δ^2 = (max variation in rotor angle)^2 σ_ω^2 = (3.132 × 10-4) ^{2} [10]

Using this formulation and applying it to IEEE 9-bus power system with the specified parameters, the quadratic form $r_k^T Q_d^{-1} r_k$ yields the threshold value $\tau = 3.7 \times 10\text{-}5$. This threshold provides an optimal balance between detection sensitivity and false alarm rate, as verified by extensive simulations under various operating conditions.

B. CPA Detection Framework

The SP-PF algorithm enables precise detection of cyber-physical attacks by leveraging accurate state estimates to identify anomalous system behavior through a KL divergence-based state partitioning optimization and statistically rigorous hypothesis testing.

Algorithm 2 KL Divergence for Subspace Management

Input State estimates, KL divergence threshold values for merging and splitting
Output Optimized partitions, CPA detection decision

Initialize partition of the state space;
while not converge
Predict and update particles based on state dynamics and measurements;
Compute KL divergence $D_{KL}(P\|P') = \sum_i p(x_i|P) \log\left(\frac{p(x_i|P)}{p(x_i|P')}\right)$;
for each region
if $D_{KL}(P\|P') < threshold_{merge}$ then
Evaluate and merge with similar adjacent regions;
elseif $D_{KL}(P\|P') \geq threshold_{split}$ then
Split region to create finer partitions;
end
end
Adjust partition boundaries based on merge/split decisions;
Calculate likelihood ratio $L(x_k) = p(x_1, \ldots, x_k | H_1) / p(x_1, \ldots, x_k | H_0)$
if $L(x_k) > \tau$
Trigger CPA alert;
end
end
return optimized partitions, CPA detection decision

The CPA detection framework begins with initial state partitioning and progressively refines these partitions based on KL divergence metric, which quantifies the information loss when approximating the true state distribution with a simplified partitioned model. The algorithm dynamically decides when to merge similar regions or split complex regions, with merging occurring when the KL divergence falls below a threshold, indicating that maintaining separate partitions adds minimal information value, and splitting triggered when the KL divergence exceeds a threshold [37], [38] signalling that a single partition inadequately represents the underlying statistical complexity.

This approach reduces computational complexity from $O(N^2)$ for standard KL calculations to $O(NlogN)$ through efficient partition-based evaluation, enabling real-time detection even in large-scale power systems. The detection process employs a two-hypothesis statistical framework:

H_0 for CPA absent,

$$x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} \tag{37}$$

and

H_1 for CPA present,

$$x_k = g\{f(x_{k-1}, u_{k-1}) + w_{k-1}\}, \tag{38}$$

An alarm is triggered whenever the likelihood ratio exceeds a threshold τ, with this threshold dynamically computed based on statistical decision theory. This dynamic threshold adaptation represents a significant advancement over fixed-threshold approaches, maintaining a consistent false alarm rate while maximizing detection probability across varying operating

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

conditions and attack scenarios. Compared to conventional methods that exhibit average detection delays of 5-10 cycles, this approach consistently achieves detection within 2-3 cycles while maintaining false alarm rates below 1%, representing a substantial performance improvement for time-critical power system security applications.

1. Proof of Convergence for the Likelihood Ratio: Consider the dynamical system represented by eqns. (21)-(26). Let the SP-PF output be given by the state and covariance estimate equations and $L(x_k) = L_{x_k}(x_{k-N_s+1}, \ldots, x_k)$ be the likelihood ratio function. Then,

$$L_{(x_k)}\left(x_{k-N_s+1}, \ldots, x_k\right) = L(x_{k-N_s+1}, \ldots, x_k). \tag{39}$$

Observe that under the null hypothesis $\mathcal{H}_0$

$$p\left(x_{k-N_s+1}, \ldots, x_k \,|\, \mathcal{H}_0\right) = \prod_{i=k-N_s+1}^{k} p(x_j), \tag{40}$$

and under the alternative hypothesis

$$p\left(x_{k-N_s+1}, \ldots, x_k \,|\, \mathcal{H}_1\right) = \prod_{i=k-N_s+1}^{k} p(x_j|X_{j-1}), \tag{41}$$

where, $p(x_j|X_{j-1})$ can be determined from:

$$p\left(x_j \,|\, X_{j-1}\right) = \int_s p\left(x_j \,|\, s_j\right) p\left(s_j \,|\, X_{j-1}\right) ds_j, \tag{42}$$

$$= E_{p\left(s_j|X_{j-1}\right)p\left(x_j|\cdot\right)}. \tag{43}$$

The detection of CPA is facilitated through residual generation. Based on above discussion, it can be concluded that particle-filter-based likelihood ratio is asymptotically equal to the true likelihood ratio. The implementation of proposed SP-PF algorithm is shown in Fig. 1. The PMU measurements (ω_{ref} *and* δ_{ref}) provide a dataset of sampled values (SV) that serve as a foundation to generate the posterior distribution, encapsulating the probabilistic distribution of state estimators.

These posteriors undergo a process of segmentation and consolidation via KL divergence [4], resulting into N. This partitioning strategy primarily hinges on variance as a pivotal criterion. Following partition delineation, the weighted mean and variance are computed to construct particles corresponding to each partition. These particles undergo iterative refinement via a one-step-ahead filter, enhancing their accuracy and alignment with real-time observations. Subsequently, normalization and potential resampling steps are employed to ensure optimal particle representation.

The algorithm generates a dynamically evolving likelihood ratio which is compared against the threshold obtained from residual generation for the test system. By evaluating whether the threshold τ satisfies the detection condition, the launch of a cyberattack can be reliably detected. At the control center, actual power measurements are observed, which generates the vector of residual from measurement samples. The continuous monitoring of the likelihood ratio enables timely detection of deviations from expected behavior, indicating the presence of CPA in the power system.

The detailed real-time simulation studies on attack detection for three different types of attacks (FDI, replay, hybrid) are given in section V. In the next section, the real-time simulation setup used for analysis is discussed.

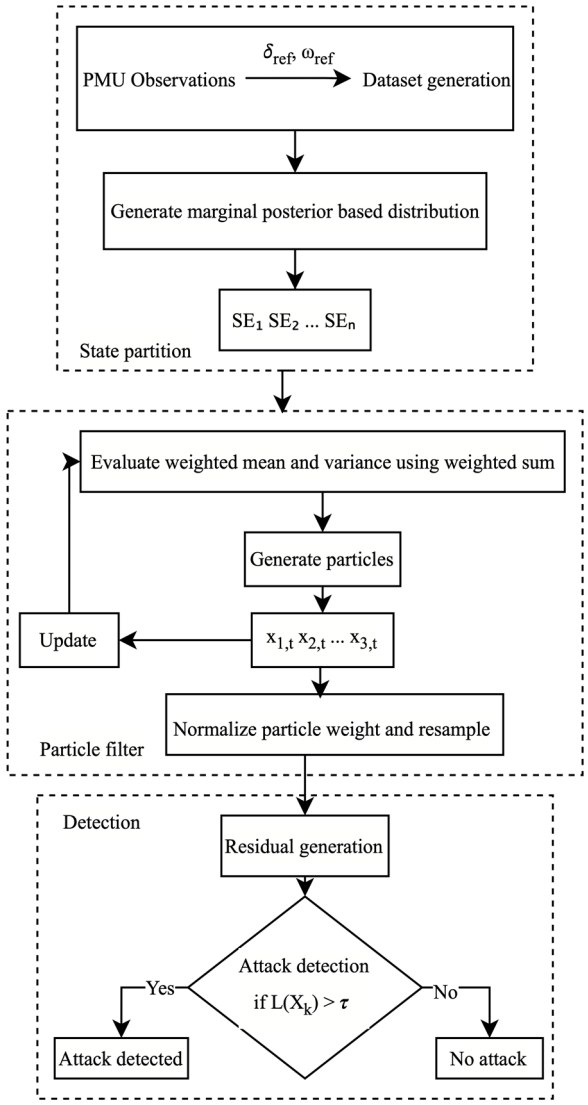


Fig. 1. Algorithm for SP-PF based CPA detection.

V. Real-Time Simulation Setup and Algorithm Implementation

In this section, we first describe the real-time simulation setup, followed by the implementation of SP-PF algorithm for CPA detection.

A. Real-Time Set up

The real time simulation environment consists of several integrated hardware and software components. At its core is a RTDS rack executing the power system model, complemented by GE feeder protection IEDs configured as PMUs, and a precision time synchronization clock. These components are interconnected through a managed network switch with a router used only for out-of band management. The IEEE 9-bus, 3-generator test system serves as power network model. The setup employs industry-standard protocols optimized for power system applications. For synchrophasor measurements, we utilize IEEE C37.118.2, which provides synchronized magnitude and phase angle data with precise

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

timestamps. This implementation is facilitated through the GTNET-PMU card in the RTDS hardware, which generates compliant IEEE C37.118.2 data streams at a configurable reporting rate of 60 frames/second. MATLAB consumes only IEEE C37.118.2 synchophasors frames from PDC and does not subscribe to IEC 61850 GOOSE or IEC 61850-9-2 SV, those layer-2 messages remain within the RTDS/IED domain.

IEC 61850 GOOSE (generic object-oriented substation event) protocol plays a critical role in this experimental setup for event-driven communications. Implemented through the GTNET-GSE card, GOOSE messages are used specifically for transmitting circuit breaker status changes during normal operations and attack scenarios, communicating control commands between the control center and field devices, signaling detection events from the monitoring system to protection equipment, and facilitating trip commands when attacks are confirmed. These layer 2 multicast messages provide deterministic performance with typical latencies under 4ms, ensuring time-critical protection functions can operate reliably. The GOOSE protocol's retransmission mechanism, where messages are initially sent in rapid succession before slowing to a heartbeat rate, provides high reliability for critical status information. Any GOOSE exchange remain local to RTDS/IEDs, their effect are observed indirectly via the time-tagged phasor stream.

SV protocol, implemented through the GTNET-SV card following IEC 61850-9-2 specifications, serves distinct purposes in this setup. It provides high-resolution voltage and current samples (80 samples /cycle) directly from instrument transformers, enables precise waveform analysis of transient phenomena during attack conditions. The SV protocol operates at layer 2 with strict timing requirements, publishing measurements in multicast frames at fixed intervals synchronized to the system frequency. This ensures all subscribing IEDs receive identical measurement data for coordinated protection decisions. SV is used within RTDS/IEDs to emulate waveform behavior and is not consumed by MATLAB, comparison in MATLAB use the IEEE C37.118.2 phasor stream.

The measurement data acquisition operates at 400 samples/second, corresponding to 8 samples/cycle. This sampling rate provides sufficient resolution for capturing transient dynamics during both steady-state operation and attack conditions, while maintaining manageable computational requirements for real-time analysis. The physical connections in this setup follow a systematic configuration as illustrated in Fig. 2.

The RTDS hardware interfaces with the network through dedicated ethernet ports from its GTNET cards. The PMUs connect to the analog outputs of the RTDS simulator through appropriate transducers and provide digital outputs to the network switch. The user PC connects to this network and hosts both the attack generation scripts and the detection algorithm implementation, allowing controlled experimentation within the isolated test environment.

B. Implementation of Detection Algorithm

We employ the power system toolbox (PST) for power flow calculation, while state estimation with attack

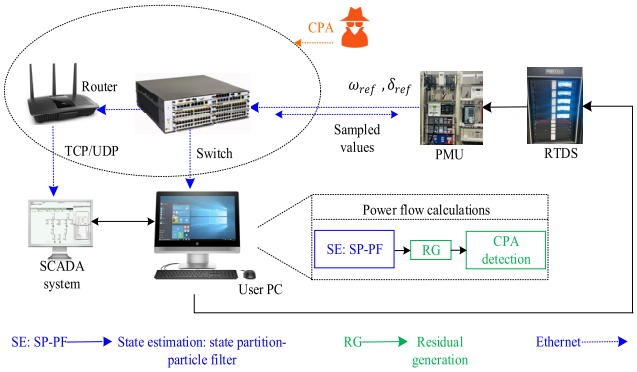


Fig. 2. Real-time simulation set-up indicating the flow of data.

detection is also implemented in MATLAB, interfacing with the RTDS-simulated IEEE 9-bus model. The bi-directional data exchange between MATLAB and RTDS is established through a dedicated TCP/UDP socket connection. This connection handles simulation control parameters and results.

Within the implementation stage, GOOSE and SV protocols serve complementary roles that enhance the realism of testbed. GOOSE messaging facilitates the attack implementation process by enabling rapid propagation of state changes during attack scenarios. When executing an FDI (or replay/hybrid) scenario, the attack script triggers state changes inside RTDS, their effects are observed in MATLAB indirectly through the time-tagged PMU phasors (IEEE C37.118.2). The SV protocol enables precise verification of the attack impacts on the power system by providing waveform data from the RSCAD model that can be compared against estimated states. The detection algorithm establishes baseline operation parameters for the power system, generating initial state distributions based on normal operating conditions, and dynamically calculating the detection threshold τ according to eqn. (29). During continuous monitoring, the algorithm receives PMU measurements via the IEEE C37.118.2 protocol, performs state-partitioning based on cross-correlation between state variables, executes particle filtering to estimate system states, calculates the likelihood ratio according to step 15 in table 1, and compares this ratio against the established threshold.

The dataset is generated using a real time HIL environment based on the IEEE 3-machine 9-bus system. The complete dataset comprises approximately 7,200 time-series samples across 12 distinct experimental configurations, with each sample containing 8 state variables (rotor angle, rotor speed deviation, transient voltages, and terminal measurements). Each test scenario spans approximately 10 seconds, with attack durations ranging from 3.1 to 7.0 seconds depending on the scenario, and including pre-attack and post-attack periods for baseline comparison. The dataset incorporates three distinct attack types: FDIA with parameter 'a' ranging from 0.001 to 0.01, replay attacks with similar parameter ranges, and hybrid attacks combining elements of both attack vectors with parameter 'a' fixed at 0.01 at generator #1 using recorded normal operation data. When the likelihood ratio exceeds

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

the threshold value, the algorithm flags an attack detection event, identifies the attack origin point (AO) and the attack detection point (AD), and provides this information to the operator. This comprehensive hardware-in-loop implementation provides a realistic cyber-physical test environment that accurately represents both the power system dynamics and the communication infrastructure, allowing rigorous validation of the SP-PF algorithm under diverse operating conditions and attack scenarios.

VI. Results and Discussion

In this section, performance of CPA detection algorithm is discussed, which is applied on real-time platform using RTDS and MATLAB. The different attack types: FDI, replay and hybrid attacks exhibit different behaviors depending on the point of application, such as when targeting the exciter or governor of generators [10], [11]. As discussed later in this section, the proposed SP-PF exhibits a high degree of accuracy in distinguishing between dynamic states during normal and under malicious cyber events, showcasing its potential as a reliable detection mechanism.

A. Detection of FDI Attack

The results obtained on implementation of SP-PF algorithm to detect FDI launched on exciter #1 of generator #1 are presented in this section. The FDI attack can be formulated as:

$$x_{kFDI} = (1 \pm a)x_k. \tag{44}$$

where, x_{kFDI} is corrupted kth state vector, x_k is kth state vector and a is the parameter between 0.001 and 0.01 [35].

The FDI attack is systematically constructed and tested under various operating conditions to comprehensively understand the algorithm's performance. The time duration for launch of attack is considered as 6.0 sec. Fig. 3 shows the variation of detector index; likelihood ratio for different operating conditions, on launch of FDI attack. The attack originates at AO (0.01 sec) and is detected at AD point as shown in said figure. The likelihood ratio, being state-dependent, exhibits dynamic changes over time following the initiation of the attack. Detection of the attack occurs when this likelihood ratio surpasses the predefined threshold limit, set at 3.7×10^{-5} obtained from eqn. (29). A randomly chosen parameter a value i.e. compromise by attacker, is 0.002. The variation in detection index obtained using SP-PF is also compared with EKF. As seen, the index starts to increase before the attack is launched, i.e. on start of simulation, but crosses the threshold at almost the same time.

However, index obtained using SP-PF reaches to its peak value earlier and remains constant for a brief period. Fig. 3(b)–3(d) shows the likelihood variation for changes in operating conditions. The value of parameter a are 0.004, 0.006, 0.008 respectively. The outcome demonstrates the successful detection of the cyber-attack using the SP-PF technique under these altered operating conditions. The algorithm proves its efficacy in identifying deviations, caused by the CPA from normal condition, thus reinforcing its utility as a reliable tool for detection. The ability to detect variations in active and reactive power highlights the algorithm's sensitivity to changes in the system conditions, emphasizing its potential for enhancing the security and resilience of power systems against malicious attacks.

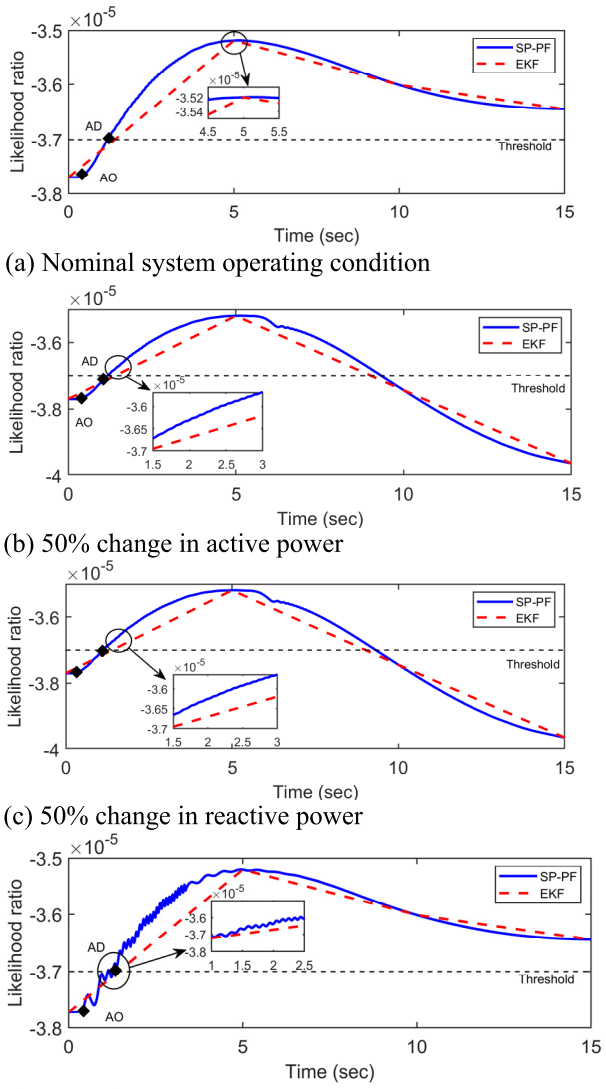
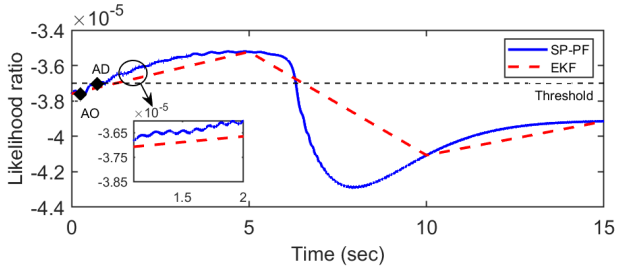


Fig. 3. Detection of FDI attack under changes in operating conditions.

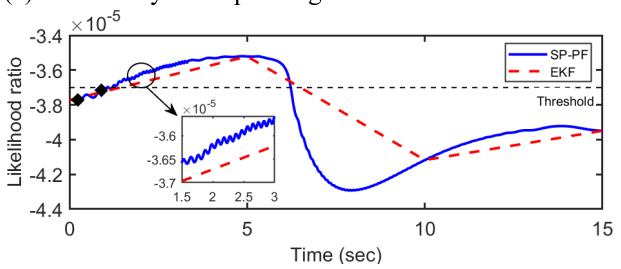
B. Detection of Replay Attack

This section presents the assessment of SP-PF detection algorithm against replay attack targeted on exciter #1 of generator #1. A systematic test is conducted for replay attack under diverse operating conditions to thoroughly assess the effectiveness of the algorithm. In Fig. 4(a), results are depicted for base operating, when the replay attack is initiated at 0.01 sec. The state-dependent likelihood ratio dynamically changes following the onset of attack. The detection is indicated as the detector index surpasses the predefined threshold of 3.7×10^{-5} determined from eqn. (29). A randomly value of a= 0.002 is chosen. The results are also compared with EKF. The likelihood ratio remains higher than those calculated by

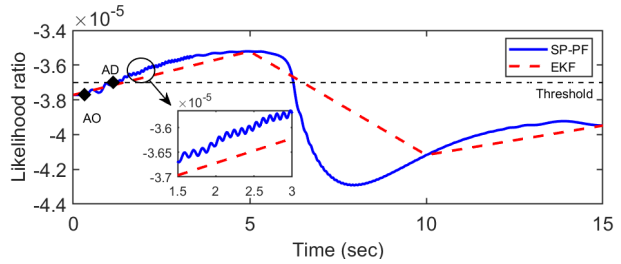
Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.



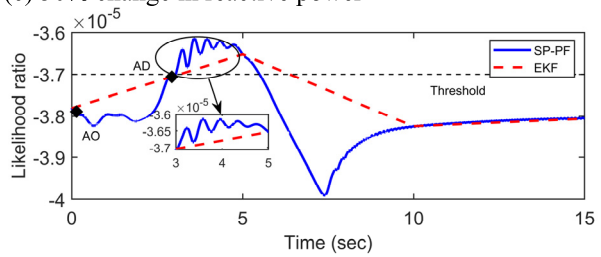
(a) Nominal system operating condition



(b) 50% change in active power

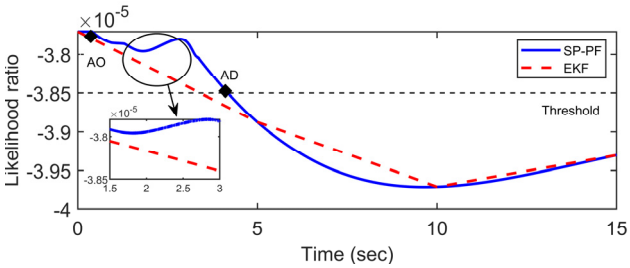


(c) 50% change in reactive power

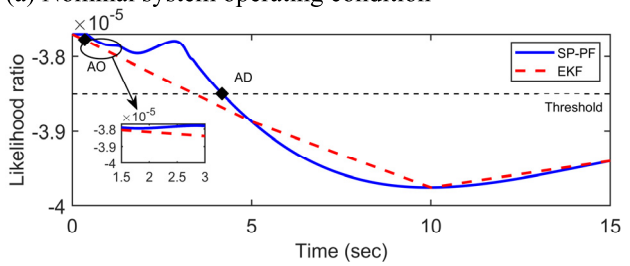


(d) 30% change in reactive power

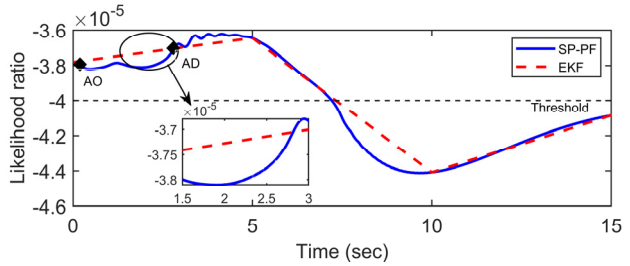
Fig. 4. Detection of replay attack under changes in operating conditions.



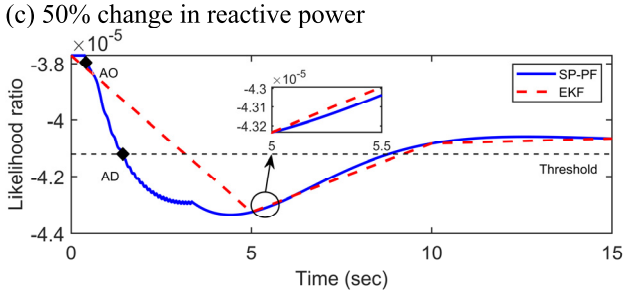
(a) Nominal system operating condition



(b) 50% change in active power



(c) 50% change in reactive power



(d) 30% change in reactive power

Fig. 5. Detection of hybrid attack under changes in operating conditions.

EKF, indicates a comparatively higher margin with respect to threshold. Thus, proposed algorithm suggests robustness in detection on changes in attack parameter, i.e. a. Next for changes in operating condition in the power network, Figs. 4(b)-(d) illustrates the evaluation of the algorithm. The successful detection of the replay attack using the SP-PF algorithm under these adjusted operational parameters is evident. Note that in Fig. 4(d), after the launch of attack, the likelihood significantly decades with negative slope, but recovers with a surge to cross the threshold to indicate the detection. After the elapse of attack duration, the index decreases below threshold. The attack originates at 0.01 seconds and is detected at AD point.

C. Detection of Hybrid Attack

Next, the effectiveness of SP-PF detection in the context to hybrid attack launched on the governor #1 of generator #1 is considered. In case of hybrid attack the value of a is chosen as 0.01 for all four conditions. The SP-PF algorithm generates a likelihood ratio under varying operational conditions, offering insights into its performance over time. In Fig. 5(a), the likelihood ratio, a crucial output of the SP-PF algorithm, dynamically evolves and upon crossing the predefined threshold value (set at 3.85×10^{-5}), indicates the attack event. The results are also compared with EKF under this condition. Similar to the above scenarios, Figs. 5(b)-(d) shows the likelihood variation for changes in power flow conditions and can be noted that computed threshold value is $3.85 \times 10^{-5}, 3.85 \times 10^{-5}, 4.0 \times 10^{-5}$ and 4.12×10^{-5} respectively and attack time duration is 6.0 sec for case (a)-(b), 7.0 sec for case (c) and 3.1 sec for case (d).

Further it is observed that the performance of proposed detection algorithm remains consistence as discussed above.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

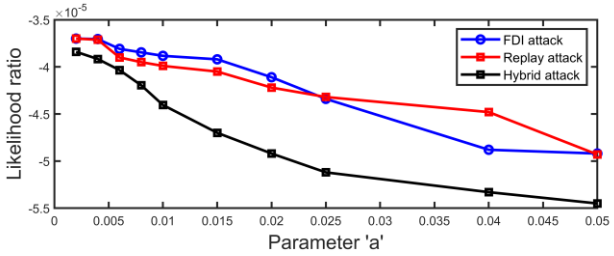
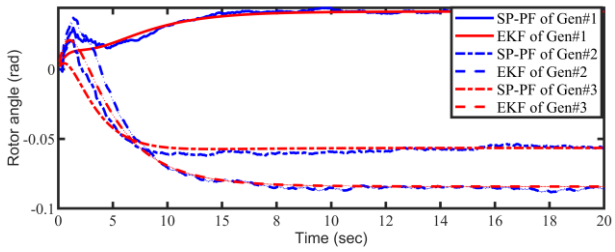
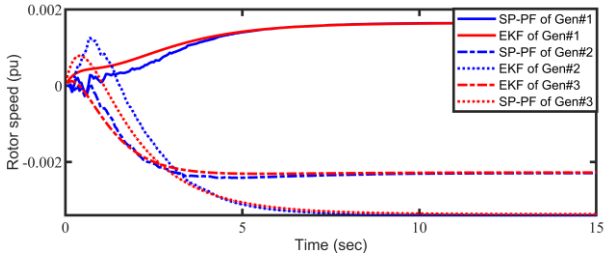


Fig. 6. Changes in likelihood ratio for parameter 'a'.



(a) Estimation of rotor angle



(b) Estimation of rotor speed

Fig. 7. State estimation analysis for nominal system operating condition under hybrid attack.

D. Analysis

The analysis of three types of attacks concerning the intruder-dependent parameter 'a', eqn. (44) is illustrated in Fig. 6. For simplicity, the parameter 'a' is considered as intensity of the attack. The likelihood ratio decreases as the attack intensity increases. The replay attack replicates packets at each interval, thus increasing the attack intensity more significantly compared to FDI attack. The hybrid attack, that combines the elements of both FDI and replay attacks, shows changes in likelihood ratio in their accordance until $a = 0.025$. Between 0.025-0.05, the negative slope of likelihood changes reduces. The estimation results obtained from SP-PF and EKF for all generators #1, #2 and #3 are shown in Fig. 7. As seen in Fig. 7(a), the SP-PF can accurately estimate the states for nominal system operating condition in case of hybrid attack. The proposed algorithm can successfully estimate the state dynamics during the time period of CPA. On other hand, those estimated using EKF, mainly during 1-7 sec, including CPA duration, are associated with large error.

However, in later part, with non-existence of attack, the estimated states using both techniques follow similar trend. Similar observations can be observed for rotor speed under nominal system operating condition in case of hybrid attack, as shown in Fig. 7(b).

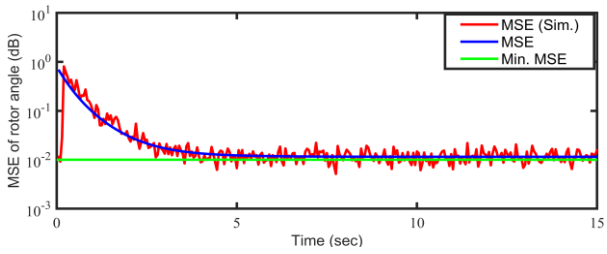


Fig. 8. State estimation analysis for base operating condition under hybrid attack.

The estimation accuracy is quantified using mean square error (MSE), which measures the average squared difference between the estimated states and their true values, defined as $MSE = (1/n)\sum(\hat{x} - x_i)^2$, where $\hat{x}$ represents the estimated state, $\hat{x}$ is the true state, and n is the number of observations.

The computed MSE from the performance of SP-PF for rotor angle variable under hybrid attack is shown in Fig. 8. The minimum MSE scale as shown typically refers to the theoretical MMSE that could be achieved if the particle filter algorithm remains perfect. However, due to finite number of particles, resampling issues and model inaccuracies are introduced. In contrast, for the same data set, the SP-PF does not show significantly large estimation errors, indicating the SP-PF's improved performance. Further as indicated, MSE computed from estimates using EKF tends to oscillate around the mean value of convergence. On other hand, those from SP-PF follows a smooth convergence to the minimum MSE value. The mean square error between the rotor angle of generator #1 and the estimated rotor angle using SP-PF is 0.001929.

VII. Conclusion

In this work, a novel CPA detection algorithm based on SP-PF, aimed at addressing the persistent challenge of adaptability posed by different attack strategies was presented. The formulation of detection algorithm was structured around two fundamental steps: firstly, the estimation of SP for the state variables, and secondly, the utilization of likelihood as detection criteria. By dividing the state space into small, manageable blocks, the SP enabled independent execution of correction and resampling steps, with enhancement of computational efficiency. As a result, partitioning step effectively groups the highly correlated state variables, facilitating more precise detection. The algorithm adjusts the detection threshold in response to variable operating conditions.

The experimental demonstration of proposed CPA detection algorithm in real-time on cyber-physical platform ensured its realistic and comprehensive evaluation. The results obtained underscore the algorithm's robustness and effectiveness in detecting a wide array of attack scenarios, thus offering promising prospects for enhancing the security and resilience of power system networks against evolving threats. The SP-PF algorithm has undergone rigorous testing, demonstrating its efficacy in identifying various attack types: FDIAs, replay attacks, and hybrid attacks.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.

VIII. Acknowledgment

This research has been performed using the ERIGrid 2.0 Research Infrastructure. The support of European Research Infrastructure ERIGrid 2.0 and its partner: Electrical Sustainable Power Lab (TU Delft) is very much appreciated.

References

[1] C. Peng, H. Sun, M. Yang, and Y.-L. Wang, "A survey on security communication and control for smart grids under malicious cyber attacks," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 8, pp. 1554–1569, Aug. 2019.

[2] M. M. S. Khan, J. A. Giraldo, and M. Parvania, "Attack detection in power distribution systems using a cyber-physical real-time reference model," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1490–1499, Mar. 2022.

[3] S. Yadav, N. Kishor, S. Purwar, V. S. Rajkumar, and A. Ştefanov, "A state partition particle filter based approach for detection of switching attack," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Norway, Sep. 2024, pp. 135–140.

[4] S. Pérez-Vieites, J. Vilà-Valls, M. F. Bugallo, J. Míguez, and P. Closas, "Second order subspace statistics for adaptive state-space partitioning in multiple particle filtering," in *Proc. IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2019, pp. 609–613.

[5] X. Zhang, D. J. Kang, and S. Choi, "Evaluation of false data injection attacks on automatic generation control considering wind power dynamics," *Int. J. Electr. Power Energy Syst.*, vol. 143, Dec. 2022, Art. no. 108400.

[6] H.-Y. Tran, J. Hu, X. Yin, and H. R. Pota, "An efficient privacy-enhancing cross-silo federated learning and applications for false data injection attack detection in smart grids," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2538–2552, 2023.

[7] J. Xu, Z. Wu, T. Zhang, Q. Hu, and Q. Wu, "A secure forecasting-aided state estimation framework for power distribution systems against false data injection attacks," *Appl. Energy*, vol. 328, p. 120107, Dec. 2022.

[8] K. A. Rahman, K. S. Balagani, and V. V. Phoha, "Snoop-forge-replay attacks on continuous verification with keystrokes," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 528–541, Mar. 2013.

[9] Y. Yu, W. Yang, W. Ding, and J. Zhou, "Reinforcement learning solution for cyber-physical systems security against replay attacks," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2583–2595, 2023.

[10] S. Yadav, N. Kishor, S. Purwar, S. Chakrabarti, P. Raussi, and A. Kumar, "Real-time implementation for vulnerability of power components under switching attack based on sliding mode," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 9, no. 4, pp. 375–391, Dec. 2024.

[11] S. Yadav, N. Kishor, S. Purwar, and S. Chakrabarti, "Indirect cyber-physical attack with combined circuit breaker and excitation system," in *Proc. IEEE EUROCON 20th Int. Conf. Smart Technol.*, Jul. 2023, pp. 204–209.

[12] M. N. Kurt, Y. Yilmaz, and X. Wang, "Real-time detection of hybrid and stealthy cyber-attacks in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 498–513, Feb. 2019.

[13] Q. Liu, V. Hagenmeyer, and H. B. Keller, "A review of rule learning-based intrusion detection systems and their prospects in smart grids," *IEEE Access*, vol. 9, pp. 57542–57564, 2021.

[14] Y. N. Soe, Y. Feng, P. I. Santosa, R. Hartanto, and K. Sakurai, "Rule generation for signature based detection systems of cyber attacks in IoT environments," *Bull. Netw., Comput., Syst., Softw.*, vol. 8, no. 2, pp. 93–97, 2019.

[15] S. Pan, T. H. Morris, and U. Adhikari, "A specification-based intrusion detection framework for cyber-physical environment in electric power system," *Int. J. Netw. Secur.*, vol. 17, no. 2, pp. 174–188, 2015.

[16] A. Hussain, E. Marín Tordera, X. Masip-Bruin, and H. C. Leligou, "Rule-based with machine learning IDS for DDoS attack detection in cyber-physical production systems (CPPS)," *IEEE Access*, vol. 12, pp. 114894–114911, 2024.

[17] H. Karimipour and V. Dinavahi, "Robust massively parallel dynamic state estimation of power systems against cyber-attack," *IEEE Access*, vol. 6, pp. 2984–2995, 2018.

[18] E. Ghahremani and I. Kamwa, "Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2556–2566, Nov. 2011.

[19] X. Qing, H. R. Karimi, Y. Niu, and X. Wang, "Decentralized unscented Kalman filter based on a consensus algorithm for multi-area dynamic state estimation in power systems," *Int. J. Electr. Power Energy Syst.*, vol. 65, pp. 26–33, Feb. 2015.

[20] Z. Huang, K. Schneider, J. Nieplocha, and N. Zhou, "Estimating power system dynamic states using extended Kalman filter," in *Proc. IEEE PES Gen. Meeting — Conf. Expo.*, Jul. 2014, pp. 1–5.

[21] J. Zhao and L. Mili, "Robust unscented Kalman filter for power system dynamic state estimation with unknown noise statistics," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1215–1224, Mar. 2019.

[22] Z. Jin, J. Zhao, L. Ding, S. Chakrabarti, E. Gryazina, and V. Terzija, "Power system anomaly detection using innovation reduction properties of iterated extended Kalman filter," *Int. J. Electr. Power Energy Syst.*, vol. 136, p. 107613, Mar. 2022.

[23] Y. Li, Z. Li, and L. Chen, "Dynamic state estimation of generators under cyber attacks," *IEEE Access*, vol. 7, pp. 125253–125267, 2019.

[24] S. Nath, I. Akingeneye, J. Wu, and Z. Han, "Quickest detection of false data injection attacks in smart grid with dynamic models," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 1, pp. 1292–1302, Feb. 2022.

[25] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 580–591, Mar. 2014.

[26] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2505–2516, May 2017.

[27] J. Zhao et al., "Power system dynamic state estimation: Motivations, definitions, methodologies, and future work," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019.

[28] K. Hamedani, L. Liu, R. Atat, J. Wu, and Y. Yi, "Reservoir computing meets smart grids: Attack detection using delayed feedback networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 734–743, Feb. 2018.

[29] K. Hamedani, L. Liu, S. Hu, J. Ashdown, J. Wu, and Y. Yi, "Detecting dynamic attacks in smart grids using reservoir computing: A spiking delayed feedback reservoir based approach," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 3, pp. 253–264, Jun. 2020.

[30] H. Karimipour, A. Dehghantanha, R. M. Parizi, K.-K.-R. Choo, and H. Leung, "A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids," *IEEE Access*, vol. 7, pp. 80778–80788, 2019.

[31] D. An, Q. Yang, W. Liu, and Y. Zhang, "Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach," *IEEE Access*, vol. 7, pp. 110835–110845, 2019.

[32] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5174–5185, Sep. 2019.

[33] M. Higgins, F. Teng, and T. Parisini, "Stealthy MTD against unsupervised learning-based blind FDI attacks in power systems," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1275–1287, 2021.

[34] Y. Cui and R. Kavasseri, "A particle filter for dynamic state estimation in multi-machine systems with detailed models," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3377–3385, Nov. 2015.

[35] N. Zhou, D. Meng, and S. Lu, "Estimation of the dynamic states of synchronous machines using an extended particle filter," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4152–4161, Nov. 2013.

[36] R. Wang, D. B. Work, and R. Sowers, "Multiple model particle filter for traffic estimation and incident detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3461–3470, Dec. 2016.

[37] H. Wang, X. Wen, Y. Xu, B. Zhou, J. Peng, and W. Liu, "Operating state reconstruction in cyber physical smart grid for automatic attack filtering," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 2909–2922, May 2022.

[38] Y. Zhai, M. B. Yeary, S. Cheng, and N. Kehtarnavaz, "An object-tracking algorithm based on multiple-model particle filtering with state partitioning," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 5, pp. 1797–1809, May 2009.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.



Seema Yadav (Student Member, IEEE) received the B.Tech. degree in electrical engineering from the United Institute of Technology, India, in 2014, and the M.Tech. degree in power electronics and drives from the Motilal Nehru National Institute of Technology Allahabad, India, in 2018, where she is currently pursuing the Ph.D. degree in cyber-physical security issues in power systems. Her broad research interests include cyber-physical security in power systems, real-time validation of controller hardware-in-the-loop using RTDS, and nonlinear control.



Nand Kishor (Student Member, IEEE) received the Ph.D. degree from Indian Institute of Technology Roorkee, India, in 2006. He is currently a Professor with the Department of Engineering Sciences, Østfold University College, Fredrikstad Campus, Norway. From August 2013 to October 2013, he was a Marie Curie Experienced Researcher (a Marie Curie Fellow) with the Electrical Engineering Department, Aalto University, Espoo, Finland. His research interests include AI applications in power systems, distributed generation with renewable resources, WAMS, and smart grid technologies. He serves as an Editor for *Sustainable Energy, Grids and Networks* (Elsevier).



Shubhi Purwar received the Ph.D. degree in electrical engineering from Indian Institute of Technology Delhi, New Delhi, India. She is currently a Professor with the Department of Electrical Engineering, MNNIT Allahabad, Prayagraj, India. Her research interests include cyber-physical systems, multi agent systems, nonlinear control, and intelligent control.



Vetrivel Rajkumar received the B.Eng. degree in electrical engineering from Anna University, India, in 2013, and the M.Sc. and Ph.D. degrees in electrical engineering from Delft University of Technology, The Netherlands, in 2019 and 2025, respectively. He is currently a Policy and Process Advisor with the System Operations Department, TenneT TSO B.V., Dutch transmission service operator. His broad research interests include cyber security of power grids and power system dynamics.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 08,2026 at 07:55:21 UTC from IEEE Xplore. Restrictions apply.