

Detection of False Data Injection Attacks in Nuclear Power Plants Using Particle Filters and Time-Series Transformers

Shashank Padasalagi

shashank06sp@gmail.com

M S Ramaiah University of Applied Sciences

Harshitha Pawar

M S Ramaiah University of Applied Sciences

Research Article

Keywords:

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11077950/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Detection of False Data Injection Attacks in Nuclear Power Plants Using Particle Filters and Time-Series Transformers

Shashank S Padasalagi¹

M S Ramaiah University of Applied Sciences Peenya 4th
Phase, Peenya, Bengaluru, Karnataka 560058

Corresponding author email: shashank06sp@gmail.com

Harshitha U K Pawar²

M S Ramaiah University of Applied Sciences Peenya 4th
Phase, Peenya, Bengaluru, Karnataka 560058

author email: harshiu kp@gmail.com

Abstract

False Data Injection encompasses a class of malicious data attacks that target critical infrastructures controlled by Cyber Physical Systems, specifically threatening the integrity of multiple type sensor data within nuclear power plants . This paper proposes the physics assisted defence framework that combined a non linear Particle Filter with a time series transformer to detect the stealthy false data injection attacks by disclosing the thermodynamic and temporal decoupling they induce across coupled nuclear reactor sensor channels. This approach pairs non linear particle filter thermodynamics with multi head attention to expose the multi channel physical decoupling an attacker cannot fake over time, while conventional detectors that only track static linear residuals are easily bypassed by stealth attacks.. Finally, simulations conducted on MATLAB and PyTorch systems demonstrate that the proposed method effectively detects FDIA in nuclear power plant.

1. Introduction

Smart Infrastructure and services are essential for developing smart cities and industries. Moreover modern critical energy infrastructures have increasingly transitioned into integrated Cyber Physical Systems (CPS) . Application areas are very diverse as autonomous vehicles, civil infrastructure, energy, manufacturing, most of them are security-critical. Within the nuclear power plants, this integration is demonstrated in digital instrumentation and control architectures, where supervisory control and data acquisition (SCADA) networks, distributed control systems (DCS) and smart field sensors interface to maintain safe and stable reactor operation. While it offer uninterrupted operation precision, but it simultaneously increases the physical attack surface by exposing deterministic thermodynamic processes to highly capable cyber adversaries.

The concrete impact of this vulnerability was observed in late 2019 at Kudankulam nuclear power plant in India. The IT network administrative of the facility was breached through a cyber intrusion using the Dtrack malware family, a strain dedicated to state sponsored advanced persistent threat (APT) groups. Moreover official post incident audits confirmed the malicious footprint was restricted to administrative workstations and did not cross the network segregation into the critical reactor protection systems . This security incident highlights the growing threat of false data injection attacks, where adversaries manipulate raw sensor measurements such as flow rates of by-products or temperatures to induce malicious outcomes while remaining consistent with physics-based expectations .

Before describing the problem statement, we review the literature concerning to cyber physical anomaly detection and false data injection attacks in safety critical industrial infrastructure. In the authors apply classical Kalman filtering and chi-squared residual tests to identify sensor anomalies within supervised control architectures. More recently, researchers have deployed data driven machine learning models, such as auto-encoders and deep recurrent networks to monitor dual channel industry telemetry. However these studies predominately rely on quasi-static assumptions that ignore the underlying plant dynamics. In nuclear control loop , a falsified sensor channel such as coolant temperature directly couples to neutron flux, precursor concentrations. The rate of chain reaction can be traced by rate of flow of xenon poisoning over the time. Consequently, attacker can exploit the blind spots of linear estimators to destabilize the core undetected , the non-linear dynamics of point reactor kinetics ensure the continuous stealth attacks inherently violate physical constraints ($\dot{c} \neq A c$). Existing approaches fail to

harness this coupled thermodynamic behavior, creating a critical gap for hybrid frameworks that integrate non-linear particle filtering with cross-attention transformers to detect such physical and temporal decoupling in real time.

In this paper, we present the MATLAB simulation models utilizing the five coupled constraints from point reactor kinetics. Prompt neutron flux (δn), precursor concentrations (δC), fuel temperature (δT_f), Xenon Poisoning (δX), and valve positioning (δV). The machine learning framework pairs a particle filter, which evaluates the thermodynamic mismatch using KL divergence over a physics centred state swarm with a time series transformer that utilizes multi-head cross attention over a sliding window to detect cross-sensor and temporal decoupling.

2. Problem Formulation

In this section, we formulate the nuclear cyber physical system as multi channel control and estimation architecture, containing linearized point reactor kinetics, kalman filter and a chi-squared residual failure detector operating under SCADA telemetry.

2.1 Non Linear Point Reactor Kinetics & Physical State Constraints

To model a cyber physical system of pressurized water reactor (PWR), here are the five fundamental physical states selected under operational transients: prompt relative neutron flux (δn), delayed neutron precursor concentration (δC), fuel temperature deviation (δT_f), Xenon-135 isotopic poisoning concentration (δX), and primary coolant control valve positioning (δV). These five states are the main core multi-physics necessary to expose cyber physical anomalies:

- i) Coupled Neutronic-Thermodynamic feedback: To manipulate single thermal or hydrodynamic channel without modifying the neutron population violates inherent physical reactivity feedback mechanism .
- ii) Multi scale temporal invariants: Integrating both rapid neutronic transients (constrained by δn and δC on the order of milliseconds to seconds) and slow thermodynamic or isotopic drift (constrained by δT_f and δX on the order of minutes to hours) prevents an adversary from disguising long-term sensor spoofing as nominal transient behavior .
- iii) Actuator state causality: Coupling control valve displacement (δV) helps ensure that any independent drift in coolant flow or temperature is cross-checked against physical actuator authority.

A. Linearized state space representation and Jacobian Derivation

The standard linearized continuous time equations of the five physical constraints are:

i) Row 1: Prompt Neutron Flux (δn)

$$\frac{d(\delta n)}{dt} = -\frac{\beta}{\Lambda} \delta n + \lambda \delta C + \frac{\alpha_T}{\Lambda} \delta T_f + \frac{\alpha_X}{\Lambda} \delta X \quad 1$$

ii) Row 2: Precursor Concentration (δC)

$$\frac{d(\delta C)}{dt} = \frac{\beta}{\Lambda} \delta n - \lambda \delta C \quad 2$$

iii) Row 3: Fuel Temperature (δT_f)

$$\frac{d(\delta T_f)}{dt} = \gamma_f \delta n - \tau_c \delta T_f - k_{steam} \delta V \quad 3$$

iv) Row 4: Xenon Concentration (δX)

$$\frac{d(\delta X)}{dt} = \gamma_x \delta n - \lambda_x \delta X \quad 4$$

v) Row 5: Actuator / Valve Dynamics (δV)

$$\frac{d(\delta V)}{dt} = k_{valve} \delta T_f - \tau_v \delta V + u(t) \quad 5$$

Hence Defining the continuous perturbation state vector as $\delta x(t) = [\delta n, \delta C, \delta T_f, \delta X, \delta V]^T \in \mathbb{R}^5$, the linearized continuous time state space equation is:

$$\dot{x}(t) = A_c x(t) + B_c u(t) + w(t) \quad 6$$

where $w(t) \sim \mathcal{N}(0, Q_c)$ represents continuous Gaussian process noise, and $A_c \in \mathbb{R}^{5 \times 5}$ is the system Jacobian matrix:

$$A_c = \begin{bmatrix} -\frac{\beta}{\Lambda} & \lambda & \frac{\alpha_T}{\Lambda} & \frac{\alpha_X}{\Lambda} & 0 \\ \frac{\beta}{\Lambda} & -\lambda & 0 & 0 & 0 \\ \gamma_f & 0 & -\tau_c & 0 & -k_{steam} \\ \gamma_x & 0 & 0 & -\lambda_x & 0 \\ 0 & 0 & k_{valve} & 0 & -\tau_v \end{bmatrix} \quad 7$$

Segmenting the continuous dynamics over a sampling interval Δt yields the discrete-time state-space model implemented in the SCADA estimation framework:

$$x_{k+1} = A x_k + B u_k + w_k, A = e^{A_c \Delta t}, w_k \sim \mathcal{N}(0, Q) \quad 8$$

B. SCADA Architecture

Within a digital nuclear Instrumentation and Control (I&C) system, remote sensor arrays measure physical parameters across the core and transmit them to the Supervisory Control and Data Acquisition (SCADA) network . The discrete measurement model is defined as:

$$y_k = Cx_k + v_k \quad 9$$

where $y_k \in \mathbb{R}^m$ is the received telemetry vector, $C \in \mathbb{R}^{m \times 5}$ maps the state space to the sensor channels, and $v_k \sim \mathcal{N}(0, R)$ denotes Gaussian measurement noise. For maintaining real time awareness, the SCADA architecture retain a discrete linear kalman filter to compute optimal state estimates $\hat{x}_k$:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \quad 10$$

Under nominal conditions, the residual follows a zero-mean Gaussian distribution $r_k \sim \mathcal{N}(0, \Sigma_k)$, where $\Sigma_k = CP_{k|k-1}C^T + R$. An alarm is flagged if the chi-squared metric exceeds a defined threshold γ :

$$z_k = r_k^T \Sigma_k^{-1} r_k \geq \gamma \quad 11$$

2.2 FDIA Problem Formulation

The hacker with knowledge of linear observation matrix C can compromise the communication channels between field sensors and the SCADA server, by injecting a malicious attack vector $a_k \in \mathbb{R}^m$:

$$y_k^a = y_k + a_k = Cx_k + v_k + a_k \quad 12$$

In an algebraic stealth attack, the hacker constructs a_k to lie within the column space of the observation matrix by setting $a_k = Cc_k$, where $c_k \in \mathbb{R}^5$ represents an arbitrary state perturbation. By substituting the corrupted measurement into the Kalman Filter innovation sequence yields:

$$r_k^a = y_k^a - C\hat{x}_{k|k-1}^a = y_k + Cc_k - C(\hat{x}_{k|k-1} + c_k) = r_k \quad 13$$

Because $r_k^a = r_k$, the test statistic remains completely invariant to the injection:

$$z_k^a = (r_k^a)^T \Sigma_k^{-1} r_k^a = z_k < \gamma \quad 14$$

Since the linear residual monitor is rendered blind, allowing the hacker to bias core telemetry arbitrarily without triggering standard SCADA anomaly alarms.

3. Threat Model and Decoupling

Here, we will analyze a hacker attack on the digital communication network that exists between the field instrumentation and supervisory SCADA network of the reactor. It is assumed that the hacker has the full information about the linear observation matrix C and the discrete state transition matrix $A = e^{Ac\Delta t}$. The attacker's objective is to change physical core states toward unsafe thermal limits while manipulating transmitted telemetry y_k so that conventional residual based detectors remain blind in this whole operation process.

Under normal circumstances, the dynamics of the plant evolve in discrete time through:

$$x_{k+1} = Ax_k + Bu_k + w_k \quad 15$$

$$y_k = Cx_k + v_k \quad 16$$

where $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. The SCADA Kalman filter tracks innovations:

$$r_k = y_k - C\hat{x}_{k|k-1} \quad 17$$

To execute an algebraic stealth False Data Injection Attack (FDIA), the hacker injects an additive corruption eigen vector $a_k \in \mathbb{R}^m$ into the true measurements:

$$y_k^a = y_k + a_k \quad 18$$

To bypass residual checks ($z_k = r_k^T \Sigma_k^{-1} r_k < \gamma$), the attacker crafts a_k to lie within the column space of the observation matrix by projecting an unconstrained perturbation vector $c_k \in \mathbb{R}^5$:

$$a_k = Cc_k$$

To spread the latent bias along the discrete state transition matrix:

$$c_{k+1} = Ac_k \quad 19$$

Substituting the corrupted measurements into the supervisory filter yields:

$$r_k^a = y_k^a - C\hat{x}_{k|k-1}^a = (y_k + Cc_k) - C(\hat{x}_{k|k-1} + c_k) = r_k \quad 20$$

Because $r_k^a = r_k$, the linear innovation residual remains completely invariant to the injection, masking the sensor manipulation as a nominal operational within the discrete linear domain.

3.1 Analytical Decoupling Proof: The $c \neq A_c c$ Principle

Since the algebraic stealth attack guarantees residual invariance in a discrete linear estimator, it assumes the underlying physical plant strictly restricts to linear time-invariant dynamics. In a physical reactor, continuous state evolution is governed by the coupled non linear system $\dot{x}(t) = f(x(t), u(t))$.

Theorem 1 (Continuous Decoupling of Discrete Stealth Injections).

Let a nuclear reactor evolve continuously according to coupled Point Reactor Kinetics $\dot{x}(t) = f(x(t), u(t))$. If an hacker constructs an algebraic stealth injection that preserves discrete residual invariance ($r_k^a = r_k$) via $c_{k+1} = A_c c_k$, the continuous time representation of the injection trajectory $c(t)$ inherently violates the continuous governing equations:

$$\dot{c}(t) \neq A_c c(t) \quad 21$$

where A_c is the plant Jacobian evaluated at nominal equilibrium.

Proof.

Let the sensed state under attack be expressed as $\tilde{x}(t) = x(t) + c(t)$, where $x(t)$ denotes the true continuous physical trajectory and $c(t) = [c_n, c_c, c_T, c_X, c_V]^T$ is the continuous latent bias. For the compromised state to remain physically consistent across continuous time, $\tilde{x}(t)$ must satisfy the non-linear governing dynamics:

$$\dot{\tilde{x}}(t) = f(\tilde{x}(t), u(t)) = f(x(t) + c(t), u(t)) \quad 22$$

Expanding $f(x(t) + c(t), u(t))$ via a multi-variable Taylor series about the uncompromised trajectory $x(t)$:

$$\dot{\tilde{x}}(t) = f(x(t), u(t)) + \left. \frac{\partial f}{\partial x} \right|_{x(t)} c(t) + \mathcal{H}(x(t), c(t)) \quad 23$$

where $\mathcal{H}(x(t), c(t))$ represents the non-linear higher-order coupling terms established in Section II. Because $\tilde{x}(t) = x(t) + c(t)$ and $\dot{x}(t) = f(x(t), u(t))$, the physical evolution of the attack vector requires:

$$\dot{c}(t) = \left. \frac{\partial f}{\partial x} \right|_{x(t)} c(t) + \mathcal{H}(x(t), c(t)) \quad 24$$

Hence again the attacker constructs the stealth attack in the discrete domain assuming evolution along the static Jacobian evaluated at nominal equilibrium:

$$c_{linear}(t) = A_c c(t) \quad 25$$

Subtracting this linear assumption from the continuous physical requirement gives us the continuous decoupling residual $\xi(t)$:

$$\xi(t) = \dot{c}(t) - A_c c(t) = \left[\left. \frac{\partial f}{\partial x} \right|_{x(t)} - A_c \right] c(t) + \mathcal{H}(x(t), c(t)) \quad 26$$

Because operational transients shift the instantaneous operating point away from nominal equilibrium ($\frac{\partial f}{\partial x} \big|_{x(t) \neq A_c} \neq 0$) and the non-linear coupling terms cannot vanish under an arbitrary trajectory ($\mathcal{H}(x(t), c(t)) \neq 0$), it follows directly that:

$$\xi(t) \neq 0 \implies c(t) \neq A_c c(t) \quad 27$$

While maintaining the algebraic stealth in the discrete linear domain it inherently violates the continuous non-linear PRK physics. The attacker cannot spoof a single telemetry channel without inducing detectable thermodynamic, kinetic, and actuator contradictions across the remaining coupled states. This continuous physical decoupling forms the theoretical foundation exploited by our dual layer detection framework.

4. Physics Assisted Detection Framework

To disclose the continuous non linear decoupling, we need to deploy a dual layer detection framework, basically a non linear particle filter for statistical divergence tripping and a time series transformer for cross channel temporal satisfaction.

4.1 Non Linear Particle Filter and KL divergence

Unlike classical linear estimators, the particle filter navigates states through non linear point reactor kinetics equations, while retaining continuous thermodynamic coupling except linearization error.

Consider a set of N_p weighted particles $\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^{N_p}$ these represents the state probability distribution. Particles propagate through the non-linear plant dynamics:

$$x_k^{(i)} = f(x_{k-1}^{(i)}, u_{k-1}) + w_k^{(i)}, w_k^{(i)} \sim \mathcal{N}(0, Q) \quad 28$$

Weights are updated upon receiving SCADA measurements y_k :

$$w_k^{(i)} \propto w_{k-1}^{(i)} \cdot \mathcal{N}(y_k; Cx_k^{(i)}, R) \quad 29$$

The normalised particles undergo systematic resampling when the effective sample size drops below $N_p/2$. Under the algebraic stealth FDIA, the manipulated measurements contradicts the underlying kinetics, forcing the posterior $Q_k(x)$ away from the prior $P_k(x)$. To catch this mismatch, KL divergence is computed against each state channel $j \in \{1, \dots, 5\}$:

$$\mathcal{D}_k = \sum_{j=1}^5 D_{KL}(P_k^j \parallel Q_k^j) = \sum_{j=1}^5 \int P_k^j(x) \ln\left(\frac{P_k^j(x)}{Q_k^j(x)}\right) dx \quad 30$$

This initial anomaly flag triggers whenever $\mathcal{D}_k \geq \tau_{KL}$. while flagging kinetic inconsistency for deep temporal classification.

4.2 Time series transformer with cross attention

This layer helps us to process the sliding windows of multivariate telemetry, kalman residuals, and hence particle state estimates to classify the operational condition.

A. Input representation:

At the time of k , an observation window of total length W is estimated using the state estimators like $\hat{x}_t$, sensor telemetry y_t , and linear residuals r_t :

$$z_t = [\hat{x}_t^T, y_t^T, r_t^T]^T \in \mathbb{R}^{15}, t \in [k - W + 1, k] \quad 31$$

Hence the sequence $Z_k \in \mathbb{R}^{W \times 15}$ is mapped to dimension d_{model} through projection matrix W_p and augmented with sinusoidal positional E_{pos} :

$$H_0 = Z_k W_p + E_{\text{pos}} \quad 32$$

B. Multi head cross attention and classification

Each layer applies multi head attention across the temporal and spatial channels. The output of attention head h can be calculated :

$$\text{Attention}(Q_h, K_h, V_h) = \text{softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right) V_h \quad 33$$

The cross attention dynamically weights relationships between the decoupled features like identifying whether temperature which changes the production corresponding to response in prompt flux and Xenon-135 . The output from L blocks at the final time step $H_L[W, :]$ which is passed through a dense layer with a softmax activation:

$$\hat{y} = \text{softmax}(W_c H_L[W, :] + b_c) \in \mathbb{R}^3 \quad 34$$

Thus, this model outputs probabilities across three distinct scenarios, normal(safe), naive attack and stealth attack, shown in table 1, hence it is optimized through categorical cross-entropy loss:

$$\mathcal{L} = - \sum_{c=0}^2 y_c \log(\hat{y}_c) \quad 35$$

Results

A. Data Generation

The experimental dataset was generated in MATLAB by numerically integrating the five state point reactor kinetics system under standard operational transients. Zero mean Gaussian process and measurement noise were introduced for realistic SCADA conditions, yielding up to 629,760 time indexed observations. The designated attack horizons ($t \geq t_{\text{attack}}$), two adversaries were introduced. Firstly, the uncoordinated abrupt offsets (naive) and continuous algebraic temperature drift vectors designed to avoid linear monitoring. The sequences of ground-truth states, corrupted telemetry, and classical Kalman filter innovation residuals were subsequently formatted and exported to PyTorch for benchmarking

B. Simulation

The Fig 1. describes cross channel thermodynamic divergence simulated in MATLAB, this graph highlights the physical constraints of the attack. During nominal operation ($t < 20\text{s}$), telemetry accurately mirrors true reactor states across all channels. At $t \approx 20\text{s}$, the injection of an algebraic negative drift into the temperature channel causes the spoofed SCADA signal to decay toward -0.8 while the true core temperature remains stable . Moreover, the uncompromised neutron flux (δn) and Xenon-135 (δX) readings do not exhibit the corresponding reactivity and feedback adjustments that real thermal changes dictate. This decoupling directly violates continuous reactor kinetics ($\dot{c} \neq Ac$), producing the anomalous cross-channel inconsistencies that trigger the Particle Filter and Transformer detectors.

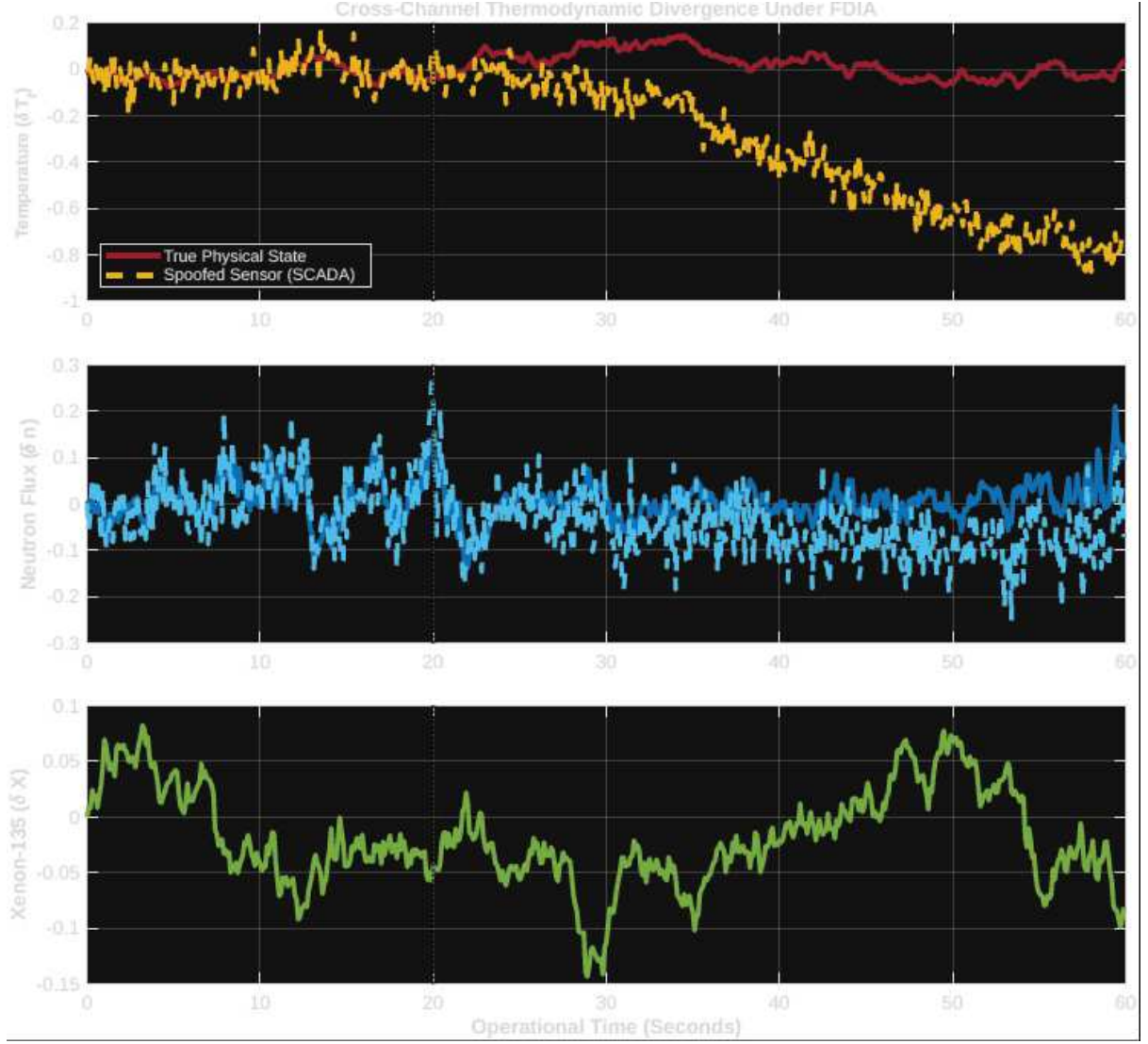


Fig 1: Working functions of the physical constraints in nuclear reactor simulation

C. Transformer Classification and Confusion Matrix Evaluation.

Table 1 shows the time series transformer achieved an overall accuracy and F1-score of 99.90% across 629,760 of evaluated samples. The confusion matrix Fig 2 confirms the perfect isolation of uncoordinated threats, identifying all 191,881 Naive Attack instances with 1.0000 precision and recall. For mathematically masked vectors, the detector correctly classified 186,580 stealth attack samples, attaining a 0.9987 precision and 0.9980 recall. Error rates remained minimal, exclusively 251 false positives and 383 false negatives. Therefore our model confirms that cross attention over non linear states reliably exposes stealthy physical decoupling without generating false alarms during normal operations.

Table 1: Precision, Recall, and F1-Score Evaluation Under Normal and Compromised Operational Conditions

Operational Constraints	Precision	Recall	F1- Score	Support
Normal (safe)	0.9985	0.9990	0.9987	250916
Naive Attack	1.0000	1.0000	1.0000	191881
Stealth Attack	0.9987	0.9980	0.9983	186963
Accuracy	-	-	0.9990	629760
Macro Avg	0.9990	0.9990	0.9990	629760
Weighted Avg	0.9990	0.9990	0.9990	629760

The confusion matrix Fig 2 demonstrates near-ideal classification fidelity, achieving zero misclassifications for Naive Attacks (191,881 true positives) while correctly identifying 250,665 safe operational states and 186,580 stealth attacks. Cross-class confusion is minimal, restricted to only 251 false alarms and 383 missed stealth detections across more than 629,000 evaluated instances.

Table 2: Confusion matrix for FDIA detection

Class	TP	FP	FN	TN
Normal(Safe)	250,665	383	251	378,461
Naive attack	191,881	0	0	437,879
Stealth attack	186,580	251	383	442,546

D. Real time dynamic response under eigenvector drift attack

We evaluate the prior to attack initiation ($t < 16s$) the reported telemetry mirrors actual core conditions within expected noise bounds. At $t = 16s$, the adversary injects an eigenvector drift that drives reported temperatures down to $-34^{\circ}C$ by $t = 60s$, Fig.2 shows prompting automated control compensation that drives true core temperatures into a hazardous climb ($> 2^{\circ}C$). While linear estimators fail to catch this smooth trajectory, the Time-Series Transformer

correlates the fabricated cooling trend against coupled non-linear kinetics, detecting the physical divergence at the injection horizon to trigger an alert before thermal thresholds are exceeded.

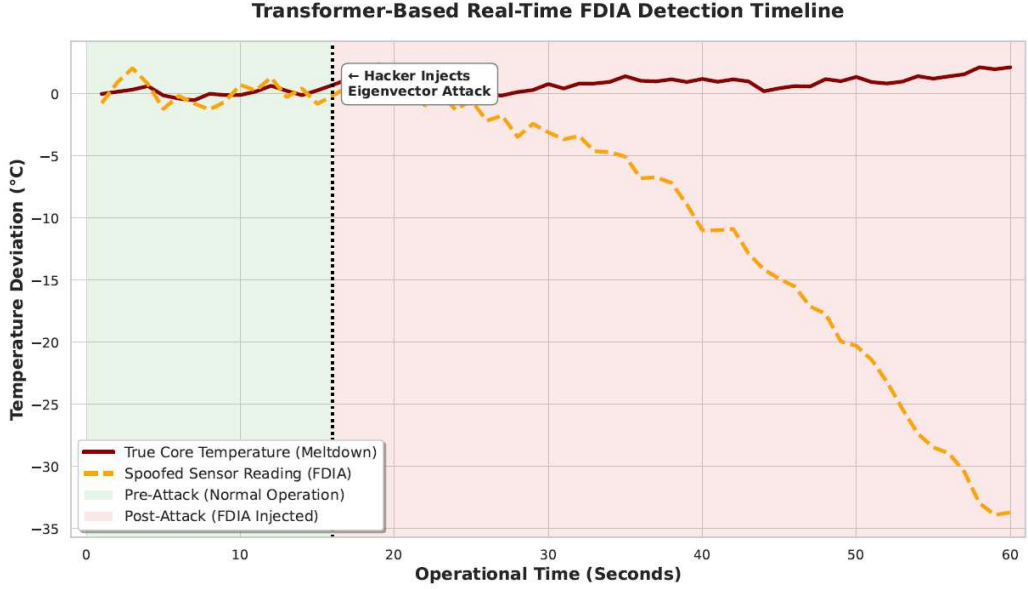


Fig 2: Dynamic tracking timeline of real time FDIA detection

Conclusion

This paper presented a false data injection attack on nuclear power plant and detection methods to override the attack before catastrophic event occurs. The proposed framework uses the 5-D physical constraints matrix to map latent state transitions, effectively identifying stealthy deviations that avoid traditional residual based monitoring. Specifically the particle filter executes non-linear Point Reactor Kinetics across a swarm of states to compute real-time physical expectations and flags thermodynamic divergence using Kullback-Leibler metrics. Concurrently, the time series transformer monitors these states and residuals across a sliding window, leveraging multi-head cross-attention to catch unfeasible cross-sensor and temporal correlations introduced by the attack. Extensive experimental results indicate that an algebraic stealth FDIA designed to deceive linear residual detectors inherently violates the continuous thermodynamic constraints ($c \neq Ac$) of coupled Point Reactor Kinetics. The dual layered model demonstrates that tracking non-linear statistical divergence via particle filtering alongside transformer based cross attention reliably exposes and detects these stealth injections by capturing the physical decoupling so that attackers cannot disguise. This contribution notably advances nuclear cybersecurity.

References

- [1] S. Yoo, G. Mohler, and F. Zhang, “Self-Healing Control of Nuclear Power Plants Under False Data Injection Attacks,” *Nuclear Science and Engineering*, vol. 199, no. 1, pp. 162–175, Jul. 2024, doi: 10.1080/00295639.2024.2372520.
- [2] Y. Li, X. Wei, Y. Li, Z. Y. Dong, and M. Shahidehpour, “Detection of False Data Injection Attacks in Smart Grid: A Secure Federated Deep Learning Approach,” *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4862–4872, Sep. 2022, doi: 10.1109/tsg.2022.3204796.
- [3] H. Kayan, M. Nunes, O. Rana, P. Burnap, and C. Perera, “Cybersecurity of Industrial Cyber-Physical Systems: A Review,” *ACM Computing Surveys*, vol. 54, pp. 1–35, Jan. 2022, doi: 10.1145/3510410.
- [4] S. Saeed *et al.*, “Digital Transformation in Energy Sector: Cybersecurity Challenges and Implications,” *Information*, vol. 15, no. 12, p. 764, Dec. 2024, doi: 10.3390/info15120764.
- [5] P. K. Mallick, “Cyber Attack on Kudankulam Nuclear Power Plant – A Wake Up Call.” Dec. 2019.
- [6] J. Giraldo *et al.*, “A Survey of Physics-Based Attack Detection in Cyber-Physical Systems,” *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–36, Jul. 2018, doi: 10.1145/3203245.
- [7] Y. Li, A. Sundaram, H. S. Abdel-Khalik, and P. Talbot, “Real-Time Monitoring for Detection of Adversarial Subtle Process Variations,” *Nuclear Science and Engineering*, vol. 196, no. 5, pp. 544–567, Jan. 2022, doi: 10.1080/00295639.2021.1997041.
- [8] D. Li, P. Ramanan, N. Gebraeel, and K. Paynabar, “Deep Learning Based Covert Attack Identification for Industrial Control Systems,” pp. 438–445, Dec. 2020, doi: 10.1109/icmla51294.2020.00075.
- [9] I. Filip, M. Abrudean, V. Mureşan, and V. A. Dulgheriu, “Xe-135 Induced Reactor Instability,” pp. 881–886, Oct. 2025, doi: 10.1109/icstcc66753.2025.11240284.
- [10] A. Lu and G. Yang, “False Data Injection Attacks Against State Estimation Without Knowledge of Estimators,” *IEEE Transactions on Automatic Control*, vol. 67, no. 9, pp. 4529–4540, Mar. 2022, doi: 10.1109/tac.2022.3161259.

- [11] Y. Khalili, M. Ahmadi, and M. K. Moraveji, “A physics-informed multi-modal transformer framework for predictive maintenance and remaining useful life estimation of electrical submersible pumps,” *Scientific Reports*, Jun. 2026, doi: 10.1038/s41598-026-56666-4.
- [12] K. Prantikos, S. Chatzidakis, L. H. Tsoukalas, and A. Heifetz, “Physics-informed neural network with transfer learning (TL-PINN) based on domain similarity measure for prediction of nuclear reactor transients,” *Scientific Reports*, vol. 13, no. 1, p. 16840, Oct. 2023, doi: 10.1038/s41598-023-43325-1.
- [13] S. E. Arda and K. E. Holbert, “Nonlinear dynamic modeling and simulation of a passively cooled small modular reactor,” *Progress in Nuclear Energy*, vol. 91, pp. 116–131, Apr. 2016, doi: 10.1016/j.pnucene.2016.03.033.
- [14] P. Desai, S. Mahale, P. M. Desai, and S. Karamchnadani, “Smart SCADA and Automation System in Power Plants,” Jan. 2014, Accessed: Nov. 2025. [Online]. Available: <https://inpressco.com/wp-content/uploads/2014/10/Paper743484-3488.pdf>
- [15] P. Griffioen, B. H. Krogh, and B. Sinopoli, “Ensuring Resilience Against Stealthy Attacks on Cyber-Physical Systems,” *IEEE Transactions on Automatic Control*, vol. 69, no. 12, pp. 8234–8246, May 2024, doi: 10.1109/tac.2024.3401013.
- [16] W. Xu, Z. Wang, L. Hu, and J. Kurths, “State Estimation Under Joint False Data Injection Attacks: Dealing With Constraints and Insecurity,” *Brunel University Research Archive (BURA) (Brunel University London)*, vol. 67, no. 12, pp. 6745–6753, Nov. 2021, doi: 10.1109/tac.2021.3131145.
- [17] G. Dán and H. Sandberg, “Stealth Attacks and Protection Schemes for State Estimators in Power Systems,” pp. 214–219, Oct. 2010, doi: 10.1109/smartgrid.2010.5622046.
- [18] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “A secure control framework for resource-limited adversaries,” *Automatica*, vol. 51, pp. 135–148, Nov. 2014, doi: 10.1016/j.automatica.2014.10.067.
- [19] Y. Liu and L. Cheng, “Completely Stealthy FDI Attack Against State Estimation in Networked Control Systems,” *IEEE Transactions on Circuits & Systems II Express Briefs*, vol. 70, no. 3, pp. 1114–1118, Nov. 2022, doi: 10.1109/tcsii.2022.3217132.

- [20] K. Zhang, C. Keliris, T. Parisini, and M. M. Polycarpou, “Stealthy Integrity Attacks for a Class of Nonlinear Cyber-Physical Systems,” *IEEE Transactions on Automatic Control*, vol. 67, no. 12, pp. 6723–6730, Nov. 2021, doi: 10.1109/tac.2021.3131656.
- [21] S. Kim, K. Park, and C. Lu, “A Survey on Network Security for Cyber–Physical Systems: From Threats to Resilient Design,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1534–1573, Jan. 2022, doi: 10.1109/comst.2022.3187531.
- [22] F. Cadini and E. Zio, “Application of Particle Filtering for Estimating the Dynamics of Nuclear Systems,” *IEEE Transactions on Nuclear Science*, vol. 55, no. 2, pp. 748–757, Apr. 2008, doi: 10.1109/tns.2008.915687.
- [23] S. Islam *et al.*, “A comprehensive survey on applications of transformers for deep learning tasks,” *Expert Systems with Applications*, vol. 241, p. 122666, Nov. 2023, doi: 10.1016/j.eswa.2023.122666.
- [24] G. Žerovnik *et al.*, “An overview of power reactor kinetics and control in load-following operation modes,” *Frontiers in Energy Research*, vol. 11, Feb. 2023, doi: 10.3389/fenrg.2023.1111357.