

<Conference abbreviation>

<Conference Series name>

<Volume number and Year> <DOI Number>

Behavior-Aware Machine Learning Approaches for Detecting False Data Injection and Ensuring Data Integrity in IoT Networks

S.Vijaya¹ V. Devi^{2*} A. Ambeth Raja³

¹Research Scholar ²Professor & Principal ³Associate Professor

^{1&2&3}PG & Research Department of Computer Science

Thiruthangal Nadar College, Chennai.

¹vijaya.samydurai1979@gmail.com²drdevi.venkatesan@gmail.com³arajacs1983@gmail.com

Abstract:

The increasing deployment of Internet of Things (IoT) systems in smart grids and cyber-physical infrastructures has made them highly vulnerable to false data injection (FDI) and data integrity attacks. In these attacks, adversaries manipulate sensor readings or network traffic to mislead monitoring and control systems, leading to incorrect decisions and potential physical damage. Traditional rule-based and signature-based security mechanisms are not effective against such stealthy and evolving threats. To address this problem, this paper proposes a behavior-aware hybrid machine learning framework for detecting FDI and data integrity attacks using only publicly available datasets. The proposed system follows a five-layer architecture that performs data preprocessing, behavior-based feature extraction, baseline behavior modeling, hybrid attack detection, and final threat classification. Two public datasets are used: a Smart Grid False Data Injection dataset containing voltage, phase angle, and power flow measurements, and the CIC-IoT-2023 dataset containing IoT network traffic with data-tampering attacks. Statistical, temporal, physical, and network-level features are extracted to model normal device behavior and identify deviations. A combination of Isolation Forest, Autoencoder, Random Forest, and Support Vector Machine is used to detect both unknown and known attacks. Experimental results show that the proposed framework achieves 97.81% accuracy, 97.34% precision, 98.12% recall, and a 2.56% false alarm rate, demonstrating its effectiveness and reliability for securing IoT systems.

Keywords: Machine learning, attacks, data security, data integrity, Internet of Things

Introduction

The Internet of Things (IoT) represents a paradigm shift in the connectivity of devices, enabling seamless communication among sensors, actuators, and computing systems. From smart homes and industrial automation to healthcare monitoring and intelligent transportation systems, IoT networks facilitate real-time data collection and analytics that support informed decision-making. According to recent estimates, the number of IoT devices worldwide is projected to exceed 30 billion by 2030, emphasizing the critical role

of these networks in modern digital infrastructure. However, the widespread deployment of IoT devices also introduces significant security and privacy challenges. Many IoT devices are resource-constrained, with limited processing power, memory, and battery capacity, which makes them particularly vulnerable to malicious attacks [1-4].

Despite the growing adoption of ML-based security solutions, several challenges remain in detecting FDI attacks and ensuring data integrity in IoT networks. First, the volume and velocity of IoT data pose significant computational and storage challenges for real-time analysis. Second, the heterogeneity of IoT devices and communication protocols complicates the generalization of ML models across different environments[5-7]. Third, attackers often employ adaptive strategies, manipulating data in subtle ways that evade conventional anomaly detection algorithms. Finally, there is a pressing need to minimize false positives, as excessive alerts can overwhelm security administrators and reduce system reliability. Addressing these challenges requires a comprehensive, behavior-aware ML framework capable of continuously learning from device interactions, contextual data, and network traffic to identify threats accurately and efficiently[8-11]. This research focuses on designing, implementing, and evaluating such behavior-aware ML frameworks for IoT security, contributing to safer and more reliable digital ecosystems.

Research Gap: While several ML-based methods have been proposed for IoT security, most existing approaches focus solely on network traffic or device-level features, ignoring the behavioral patterns that provide early indicators of attacks. Additionally, few studies comprehensively evaluate both FDI and data integrity threats in heterogeneous IoT environments using behavior-aware learning techniques. There is a pressing need for a framework that integrates behavioral analysis with machine learning models to detect attacks proactively and ensure the integrity of IoT data streams.

Research Objective: The primary objective of this study is to design a behavior-aware machine learning framework capable of:

- Detecting false data injection attacks across heterogeneous IoT devices.
- Ensuring data integrity in real-time network environments.
- Reducing false positives while maintaining computational efficiency.
- Evaluating the framework's effectiveness using benchmark datasets and performance metrics such as accuracy, precision, recall, F1-score, and computational overhead.

By achieving these objectives, the proposed work aims to provide an effective, scalable, and intelligent solution for securing IoT networks against evolving cyber threats.

Proposed Work

This research proposes a Behavior-Aware Machine Learning Framework for detecting False Data Injection (FDI) attacks and data integrity violations in Internet of Things (IoT) networks. The key idea of the proposed work is that malicious data manipulation alters the normal behavioral patterns of IoT devices, even when the injected data appears statistically valid. The proposed framework addresses this limitation by combining behavioral modeling with hybrid machine learning, enabling both known and unknown attack patterns to be detected with high reliability. The framework is designed to operate entirely on public benchmark datasets, ensuring reproducibility and eliminating dependency on real-time data collection.

Proposed Architecture

The proposed system follows a five-layer behavior-aware architecture designed to transform raw IoT data into reliable and explainable attack decisions. The core objective of this architecture is to detect false data injection (FDI) and data integrity attacks by learning the normal operational behavior of IoT devices and identifying deviations caused by malicious manipulation. Unlike traditional intrusion detection systems that rely on static rules or signatures, the proposed framework adopts data-driven behavior modeling combined with hybrid machine learning to achieve high detection accuracy and low false alarm rates.

Let the overall system be represented as a mapping function: $F: X \rightarrow Y$ where X denotes raw IoT data (sensor readings and network traffic), and $Y \in \{Normal, FDI, Integrity Attack\}$ is the predicted security state.

The five-layer architecture ensures that raw, noisy IoT data is progressively refined into discriminative behavioral patterns and then converted into reliable threat decisions.

Layer 1 – Dataset Layer

The Dataset Layer forms the foundation of the proposed architecture. It loads the datasets containing IoT sensor data and network traffic. These datasets include both normal operating conditions and maliciously manipulated samples caused by FDI and data integrity attacks.

Let the dataset be denoted as: $D = \{(x_i, y_i)\}_{i=1}^N$ where x_i represents a data record consisting of sensor and network attributes, y_i is the corresponding label indicating whether the record is normal or attacked, and N is the total number of samples

Layer 2 – Preprocessing and Feature Engineering Layer

Raw IoT data is noisy, incomplete, and heterogeneous. Therefore, this layer performs cleaning, normalization, and behavior-based feature extraction.

Data Cleaning and Normalization

Missing values are handled using mean or median imputation. Let x_{ij} be the j^{th} feature of sample i . Normalization is applied using min-max scaling: $x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$

This ensures that all features lie within $[0,1]$, preventing bias toward high-magnitude variables.

Layer 3 – Behavior Modeling Layer

This layer learns the normal operational profile of each IoT device.

Let the set of normal samples be D_n . The mean behavior vector is: $B_d = \frac{1}{|D_n|} \sum_{i \in D_n} F_i$

When a new observation F_t arrives, its deviation from normal behavior is computed as:

$$\delta = \|F_t - B_d\|^2$$

If δ exceeds a learned threshold, the data is suspicious. This allows detection of:

- False sensor readings

- Data manipulation
- Abnormal transmission behavior

Layer 4 – Hybrid Machine Learning Detection Layer

This layer applies both **unsupervised** and **supervised learning**.

Unsupervised Detection

Isolation Forest and Autoencoders learn normal behavior. Reconstruction error for Autoencoder: $E = \|F_t - \hat{F}_t\|^2$

High reconstruction error indicates anomalies.

Supervised Classification

Random Forest and SVM classify samples using labeled attack data. SVM decision function:

$$f(x) = w^T x + b$$

Probability of attack: $P_s = \sigma(f(x))$

Unsupervised anomaly score P_u and supervised probability P_s are fused.

Layer 5 – Decision and Alert Layer

Final threat score:

$$T = \alpha P_u + \beta P_s$$

If

$$T > \theta$$

the sample is classified as attack.

Classification output:

$$Y = \{Normal, FDI, Integrity Attack\}$$

Algorithm 1: Dataset Preparation and Feature Extraction

Input:

Public IoT datasets D (Smart Grid FDI, CIC-IoT-2023)

Output:

Behavior feature matrix F

Steps:

1. Load dataset D
2. Remove missing and corrupted records
3. Normalize all numerical features using min–max scaling
4. For each device d and time window t:
 - a. Compute statistical features (mean, variance, entropy)
 - b. Compute temporal features (rate of change, burstiness)
 - c. Compute physical features (voltage drift, signal deviation)
 - d. Compute network features (packet rate, delay, retransmissions)
5. Combine all extracted features to form behavior vector F_t
6. Store all behavior vectors in feature matrix F
7. Return F

Algorithm 2: Behavior Profile Learning

Input:

Behavior feature matrix F , normal data subset F_n

Output:

Baseline behavior profile B_d for each device

Steps:

1. For each IoT device d :
2. Select all normal samples F_n^d
3. Compute mean behavior vector: $B_d = \frac{1}{|F_n^d|} \sum_{i \in F_n^d} F_i$
4. Store B_d as baseline profile
5. Return all B_d

Algorithm 3: Unsupervised Anomaly Detection**Input:**

Behavior vectors F , baseline profiles B_d

Output:

Anomaly score P_u

Steps:

1. Train Isolation Forest and Autoencoder using normal data
2. For each new behavior vector F_t
 - a. Compute reconstruction error using Autoencoder
 - b. Compute anomaly score using Isolation Forest
3. Normalize anomaly scores
4. Combine scores to obtain unsupervised probability P_u
5. Return P_u

Experimental Evaluation.

Table 1 presents the overall performance of the proposed hybrid behavior-aware framework on the two public datasets. The results show that the integration of behavioral modeling with hybrid ML leads to strong detection capability for both FDI and data integrity attacks.

Table 1. Overall Performance of the Proposed Framework

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Detection Rate (%)	False Alarm Rate (%)	Processing Time (ms)
Smart Grid FDI	97.84	97.12	98.41	97.76	98.41	2.14	7.8
CIC-IoT-2023	96.93	96.45	97.21	96.83	97.21	2.97	8.5
Combined Dataset	97.38	96.79	97.81	97.29	97.81	2.56	8.1

Table 2. Ablation Study Results

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FAR (%)
Full Proposed Framework	97.38	96.79	97.81	97.29	2.56
Without Behavior Modeling	93.44	92.71	94.13	93.41	5.87
Without Unsupervised Module	94.36	93.58	95.01	94.29	4.99
Without Supervised Module	92.81	91.96	93.44	92.69	6.56
Only Traditional ML (No Hybrid)	91.75	90.88	92.11	91.49	7.89

Conclusion

This study presented a comprehensive behavior-aware machine learning framework for detecting false data injection (FDI) and data integrity attacks in Internet of Things (IoT) networks using only publicly available datasets. By integrating statistical, temporal, physical, and network-level behavioral features with hybrid machine learning models, the proposed system achieved robust discrimination between normal behavior, FDI attacks, and integrity violations. Overall, the proposed framework provides a scalable, data-driven, and reproducible solution for securing IoT infrastructures against sophisticated data manipulation attacks without relying on real-time or proprietary data. This makes the system suitable for deployment in smart grid, smart home, and industrial IoT environments where data integrity and trustworthiness are mission-critical. Future work can extend this framework by incorporating federated learning and blockchain-based auditability to further enhance resilience, privacy, and trust in distributed IoT security systems.

References

- [1] Kesici, M., Pal, B., Yang, G.: Detection of false data injection attacks in distribution networks: A vertical federated learning approach. *IEEE Transactions on Smart Grid* **15**(6), 5952–5964 (2024). <https://doi.org/10.1109/TSG.2024.3399396>
- [2] Dahiya, A., Dhingra, A., Sindhu, V., Sangwan, A.: A comparative analysis of detection and mitigation of false data injection attacks in the Internet of Things. In: Luhach, A.K., Jat, D.S., Ghosh, U., Gao, X.Z., Lingras, P. (eds.) *Advanced Informatics for Computing Research (ICAICR 2023)*. Communications in Computer and Information Science, vol. 2073, pp. 353–366. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-84059-3_28
- [3] Shahid, M.A., Ahmad, F., Albogamy, F.R., Hafeez, G., Ullah, Z.: Detection and prevention of false data injection attacks in the measurement infrastructure of smart grids. *Sustainability (Switzerland)* **14**(11), 1–26 (2022). <https://doi.org/10.3390/su14116407>
- [4] Pedroso, C., Santos, A.: Dissemination control in dynamic data clustering for dense IIoT against false data injection attack. *International Journal of Network Management* **32**(5) (2022). <https://doi.org/10.1002/nem.2201>
- [5] Qi, R., Rasband, C., Zheng, J., Longoria, R.: Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning. *Information* **12**, 328 (2021).
- [6] Alwageed, H.S.: Detection of cyber-attacks in smart grids using SVM-boosted machine learning models. *Service Oriented Computing and Applications* **16**, 313–326 (2022).

- [7] Song, H., Fink, G.A., Jeschke, S.: Detecting data integrity attacks in smart grid. In: *Security and Privacy in Cyber-Physical Systems: Foundations, Principles, and Applications*. Wiley-IEEE Press, Hoboken (2017).
- [8] Awad, M., Fraihat, S.: Recursive feature elimination with cross-validation with decision tree: Feature selection method for machine learning-based intrusion detection systems. *Journal of Sensor and Actuator Networks* **12**, 67 (2023).
- [9] Murugesan, R., Gunasekaran, P., Jeyabal, S., et al.: AI-based false data injection attack detection in smart grid advanced metering infrastructure. *Electrical Engineering* **107**, 14517–14538 (2025). <https://doi.org/10.1007/s00202-025-03276-4>
- [10] Hu, J., Yang, S., Caballero-Águila, R., Dong, H., Wu, B.: Mixed static–dynamic protocol-based Tobit recursive filtering for stochastic nonlinear systems against random false data injection attacks. *IEEE Transactions on Signal and Information Processing over Networks* **10**, 445–459 (2024).
- [11] Du, Z., Yan, Z., Xu, Y.: A dimensional augmentation-based data-driven method for detecting false data injection in smart meters. *IEEE Transactions on Smart Grid* **15**(1), 1180–1183 (2024).

Biographies



Mrs. S. Vijaya received her Bachelor of Science (B.Sc.) degree from Bharathidasan University in 2001, her Master of Science (M.Sc.) degree from Bharathidasan University in 2003, and her Master of Philosophy (M.Phil.) degree from Periyar University in 2007. She is currently working as an Assistant Professor in the Department of Computer Science at Sri Muthukumaran Arts and Science College, Chennai, and is also pursuing her Ph.D. as a Research Scholar at Thiruthangal Nadar College, Chennai. Her research interests include Artificial Intelligence, Machine Learning, Deep Learning, Data Mining, Data Warehousing, Internet of Things (IoT), Cyber Security, and Behavior-Aware Machine Learning Approaches for Detecting False Data Injection and Ensuring Data Integrity in IoT Networks.



Dr. V. Devi, Principal – Thiruthangal Nadar College, Chennai, and Senate Member in University of Madras. She received her undergraduate degree, Master degree and Ph. D from University of Madras. She published 60 research papers/article in peer review indexed Journals, Books and Book chapters in reputed publishers. She has more than 25 years of experience in teaching and research field. She also presented various academic as well as research-based papers at several national and international conferences in South East Asian Countries. Her research area includes Data Science & Analytics, Cloud Computing, Software Engineering, Digital Image Processing, and IoT Systems and Information System Management.