The 13th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
4-6 September, 2025, Gliwice, Poland

A Comprehensive Review of Public Datasets for Machine Learning-Based Intrusion Detection in IoT and OT Networks

Muhammad Afaq[1] , Jaafer Rahmani[2] , Axel Sikora[3]
[1] Department of Computer Engineering and Research Center for Intelligent Secure Systems,
King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, afaq24@gmail.com
[2, 3] Institute of Reliable Embedded Systems and Communication Electronics (ivESK), Offenburg University
Offenburg, Germany, jaafer.rahmani@hs-offenburg.de, axel.sikora@hs-offenburg.de

***Abstract* — The rapid growth of networks for Operational Technology (OT) and the Internet of Things (IoT) has increased their susceptibility to cyberattacks. Since many years, traditional intrusion detection systems (IDS) are in place. However, they are falling short, as they are insufficient against dynamic threats. However, ML-based models are only as good as their training datasets, so these datasets are of paramount importance. Consequently, this paper reviews publicly accessible datasets for anomalies in OT- and IoT-networks and evaluates their applicability to ML-based IDS. It discusses the strengths and weaknesses of the datasets, including data imbalance, oversimplified attack scenarios, and restrictions on protocol diversity.**

In order to increase the effectiveness of ML-based IDS and strengthen cybersecurity in OT and IoT environments, the paper suggests increasing dataset complexity and realism.

Keywords* — *IoT, OT, Intrustion Detection Systems, IDS, Machine Learning ML, Dataset.

I. Introduction

The rapid growth of networks for Operational Technology (OT) and the Internet of Things (IoT) has increased their susceptibility to cyberattacks [1]. Since many years, traditional intrusion detection systems (IDS) are in place. However, they are falling short, as they are insufficient against dynamic threats. The fast growth in volume and complexity of IoT and OT networks creates new vulnerabilities and makes them more appealing targets for cybercriminals. Recently, regulations, like Cyber-Resilience Act (CRA), Network- and Information Systems (NIS2), or Radio Equipment Directive (RED) mandate high security standards for devices and networks, which makes security a legal obligation.

Conventional monitoring approaches for network security, such as signature-based and rule-based Intrusion Detection Systems are ineffective against the dynamic and novel threats because they are static in nature and rely on the signatures. Consequently, ML-based IDS have become essential for protecting networks connected to IoT and OT against advanced cyberattacks [2], as they can dynamically detect anomalous behaviors including threats that were previously unknown [3].

However, the effectiveness of Machine-Learning (ML) techniques strongly depends on the size and the quality of the training datasets. A major challenge that researchers and practitioners face is the limited availability of publicly accessible datasets targeted for attacks on OT and IoT networks. This limited availability currently restricts the ability to build, train, and evaluate the performance of ML models [4] [5].

To our best knowledge, this paper presents the first systematic review of current publicly available datasets to be used to train and evaluate ML-based IDS in IoT and OT network environments. It examines notable datasets, highlights their structural characteristics, feature sets, attack types covered, and discusses their applicability to real-world IoT and OT scenarios. The paper aims to not only provide guidelines to choose appropriate datasets by highlighting their limitations, weaknesses, and strengths, but also focusing on developing more representative and realistic datasets.

The remainder of the paper is structured as follows: Section II provides an overview of intrusion detection in IoT and OT networks, Section III offers a detailed review and comparative analysis of publicly available datasets, Section IV identifies gaps and limitations, and presents future directions, Section V concludes the paper.

II. Overview of Intrusion Detection in IoT and OT Networks

IoT networks are made up of physical objects that are connected to one another and have sensors, software, and network connectivity to allow data collection and sharing. These networks are distinguished by a variety of communication protocols, heterogeneous devices, and constrained computing power. Due to their wide-ranging and diverse deployment, i.e., from wearables and smart homes to smart cities and healthcare, the IoT environment poses special security challenges, including resource limitations, a variety of attack vectors, including botnets,

979-8-3315-8045-2/25/$31.00 ©2025 IEEE

Table I. SUMMARY OF PUBLICLY AVAILABLE DATASETS FOR ML-BASED INTRUSION DETECTION IN IOT AND OT NETWORKS.

* ESTIMATED BASED ON SCENARIO COUNT, DATASET SIZE, OR SUPPORTING DOCUMENTATION WHEN OFFICIAL SAMPLE COUNTS ARE UNAVAILABLE. REFERENCES ARE GIVEN IN THE CORRESPONDING TEXT

Dataset	Protocols								Attacks					Features			Metadata		
	HTTP	MQTT	Modbus	CoAP	Zigbee	DNS	FTP	SSH	Botnet	DDoS	APT	MITM	Spoofing	Logs	PCAP	Sensors	Samples	Size	Complexity
IoT-23	✓	✓				✓			✓	✓					✓		1-3 million*	20 GB	Single-stage
SWaT (ICS)			✓							✓		✓	✓		✓	✓	946,722	11 days	Multi-stage
TON-IoT	✓	✓	✓				✓			✓				✓	✓	✓	27,520,260	25 GB	Multi-stage
CICIDS2017	✓					✓	✓	✓		✓				✓	✓		3,119,345	50 GB	Single-stage
Edge-IIoTset	✓	✓	✓	✓					✓	✓			✓	✓	✓	✓	1,909,671	20 GB	Single-stage
CIC APT IIoT 2024 (ICS)	✓	✓	✓	✓						✓	✓			✓	✓	✓	4-5 million*	65 GB	APT
Gotham Testbed (ICS)	✓	✓		✓									✓	✓			1-2 million*	10 GB	Simulated
N-BaIoT		✓		✓					✓					✓			7,062,606	15 GB	Single-stage
WADI (ICS)			✓							✓				✓	✓	✓	1,221,372	16 days	Multi-stage
EPIC (ICS)															✓	✓	500K-1 million*	5 GB	Single-stage
LBNL-ETA/Brick (ICS)			✓											✓		✓	1+ million*	2.38 GB	N/A
Power Consumption														✓			2.3 million*	3.5 GB	N/A
Blaq_0 Hackathon (ICS)															✓		0.5-1 million*	Varied	Simulated
IoT Honeypot															✓		300K-1 million*	5 GB	Single-stage

ransomware, and Distributed Denial of Service (DDoS), and problems with device authentication [6] [7].

Conversely, IoT and OT networks are largely responsible for managing physical processes in home and building automation, industrial automation, and public infrastructure, including manufacturing systems, water treatment facilities, and power grids and other critical infrastructure. With strict specifications for real-time performance and dependability, IoT and OT networks usually place a high priority on availability and safety. IoT and OT networks are more vulnerable to cyberattacks that have historically been limited to traditional IT environments as they become more integrated with IT infrastructures to increase efficiency and monitoring. Effective intrusion detection is crucial because intrusions in IoT and OT networks have the potential to cause serious operational disruptions as well as physical harm [8].

In both IoT and OT networks, IDS seek to identify abnormal activity or malicious actions that could be signs of a cyberattack. ML-based IDS provide a number of advantages over traditional techniques by offering adaptive detection capabilities that identify variations from known behavior patterns, including zero-day attacks without known signatures. In order for ML-based IDS to perform better in these specialized networks, training datasets need to faithfully capture realistic traffic, a variety of attack scenarios, and the particular operational contexts of IoT and OT environments.

III. REVIEW AND COMPARATIVE ANALYSIS OF PUBLICLY AVAILABLE DATASETS

As illustrated in Table I, SWaT and WADI are two of the most widely used ICS datasets, both originating from iTrust Labs at the Singapore University of Technology and Design. SWaT simulates attacks on a secure water treatment facility using Modbus protocol, providing 11 days of sensor and actuator data with physical-state realism and labeled cyber-physical interactions [9]. WADI complements this with a 16-day capture simulating attacks on a water distribution system [9]. EPIC focuses on power system integrity and simulates synchronization and load manipulation attacks, reflecting actuator command failures and grid-level response [9]. LBNL-ETA/Brick, while not an intrusion dataset, provides operational time-series sensor data structured with the Brick schema [10], useful for unsupervised anomaly detection and smart building diagnostics.

IoT-23, developed by Stratosphere Labs, includes labeled PCAPs across multiple botnet and malware families such as Mirai and Okiru, distributed over HTTP, MQTT, and DNS protocols [11]. It provides scenario-specific metadata, making it suitable for malware traffic modeling. N-BaIoT is tailored for IoT botnet detection using behavioral features and network patterns extracted from infected consumer IoT devices [12]. It provides over 7 million labeled records. The IoT Honeypot dataset captures opportunistic real-world attack behavior from honeypot deployments and reflects reconnaissance and probing activity [13]. Among these, IoT-23 and N-BaIoT are more structured and labeled, while the IoT Honeypot captures stealthy adversarial behavior.

TON-IoT, released by UNSW Canberra, is a large-scale dataset offering over 27 million records from MQTT, Modbus, and telemetry sources [14]. It spans smart home, industrial, and edge scenarios and includes attacks such as ransomware and data exfiltration. Edge-IIoTset complements this with a focus on edge-device constraints and includes CoAP alongside Modbus and MQTT, while capturing attacks like spoofing and malware with sensor-augmented data [15]. CICIDS2017 [16] provides comprehensive enterprise-like flow statistics and labeled packet data for classical attacks like brute-force and DDoS. While widely used as a benchmark, CICIDS2017 lacks IoT/OT specificity and cyber-physical metrics, making it better suited for traditional IT security evaluation.

CIC APT IIoT 2024 is among the most advanced datasets for IoT/ICS, simulating multi-phase attacks such as Advanced Persistent Threats (APT), command injection, and lateral movement across MQTT, Modbus, CoAP, and BACnet [17]. It includes rich behavioral logs for anomaly detection. Gotham Testbed is a synthetic IoT testbed for federated learning research, simulating

anomalous behavior over MQTT, CoAP, and HTTP protocols [18]. The Blaq_0 Hackathon dataset comprises creative, competition-generated ICS attacks and offers varying realism based on attacker strategies [19]. These datasets stand out in terms of adversarial complexity and experimental flexibility, especially for testing anomaly-aware or unsupervised IDS models.

The Building Power Consumption dataset [20] includes historical energy use across hundreds of buildings, annotated with weather, holiday, and occupancy metadata. Though not designed for intrusion detection, it is valuable for behavioral modeling and semi-supervised anomaly detection. Similarly, LBNL-ETA/Brick provides detailed operational profiles of building equipment, structured by type and space, which can assist in modeling normal operations and spotting anomalies or faults.

In summary, datasets such as IoT-23, TON-IoT, and Edge-IIoTset offer strong protocol diversity and network-layer realism for IoT use cases. OT-specific datasets like SWaT, WADI, and EPIC provide high-fidelity cyber-physical representations essential for evaluating OT intrusion detection. More advanced datasets like CIC APT IIoT 2024 and Gotham address the growing need for multi-stage and APT scenarios. However, Table I also reveals clear gaps — many datasets still lack operationally critical ICS metrics or comprehensive multi-stage attack coverage. Datasets tend to emphasize either realism or structure, but few balance both, limiting their standalone utility for generalized IDS benchmarking.

To systematically address these limitations, we evaluate the datasets against three additional criteria derived from their operational and technical constraints:

- **Data Freshness**: Temporal relevance and sampling frequency.
- **Labeling Quality**: Granularity and consistency of attack annotations.
- **Scalability**: Adaptability to large-scale or cross-domain deployments.

Our analysis reveals that while **IoT-23** offers rich botnet variants (e.g., Mirai, Gafgyt), it demonstrates inconsistent malware stage tagging (labeling quality) and limited scalability due to fixed device profiles. The **SWaT** dataset achieves high labeling fidelity through precise actuator/sensor logs, though its freshness is constrained by static attack scripts across its 11-day duration. Similarly, **WADI** captures extensive sensor/actuator data (123 nodes over 16 days) but suffers from scalability limitations due to its exclusive Modbus/TCP protocol focus.

EPIC's focus on synchronization and load attacks comes at the expense of real-time streaming support, impacting freshness for dynamic threat detection. While **CICIDS2017** provides broad network coverage, it lacks IoT/OT protocol specificity (e.g., no MQTT/Modbus) and shows coarse-grained labeling quality issues. The **TON-IoT** dataset effectively balances freshness (cross-domain telemetry) and scalability (25+ GB) but lacks operational context in its cyber-physical event correlations. Specialized datasets reveal specific trade-offs: **N-BaIoT** excels in botnet traffic patterns but shows poor labeling quality with imbalanced benign/malicious samples and limited attack diversity, while **Edge-IIoTset** delivers high labeling quality through per-device logs but suffers from freshness limitations due to synthetic attack emulation.

Recent datasets demonstrate evolving capabilities with new constraints: **CICIoT2023** advances protocol diversity (Zigbee, MQTT) but compromises labeling quality through oversimplified attack dynamics (e.g., single-stage DDoS). **CIC APT IIoT 2024** introduces valuable APT and lateral movement scenarios but requires improved labeling quality with finer stage-specific annotations. The **Gotham Testbed** prioritizes scalability through synthetic data suitable for federated learning, but this comes at the cost of labeling quality (generic anomaly tags). Protocol-specific collections like **KNX Dataset Bundle** achieve precise spoofing/replay labels (labeling quality) but fundamentally limit scalability through their narrow protocol focus.

Operational datasets present unique profiles: **HAI 1.0** delivers high-fidelity actuator-state logs (labeling quality) but lacks freshness due to limited longitudinal data (~12 GB). The **Tennessee Eastman Process** which is a simulated chemical process model widely used as a benchmark for studying various process control anomaly detection techniques [21], provides exceptional operational fidelity through multivariate time-series data but shows limited cybersecurity relevance due to missing cyberattack labels. More specialized collections like **LBNL-ETA/Brick** (HVAC analytics) and **Power Consumption (Kaggle)** (energy metrics) fall short on all three criteria by lacking attack annotations and protocol data entirely. Finally, **Blaq_0 Hackathon** and **IoT Honeypot Dataset** demonstrate fundamental limitations in scalability (task-specific attacks) and labeling quality (unstructured malware logs) respectively, despite their niche value propositions.

These findings underscore the need for future datasets to harmonize protocol diversity, attack complexity, and operational fidelity. For instance, while SWaT and HAI 1.0 excel in cyber-physical interactions, they lack protocol breadth, whereas CICIoT2023 and TON-IoT prioritize scalability at the cost of attack granularity. The Tennessee Eastman Process demonstrates the value of high-resolution process data but highlights the need to integrate cyberattack context with operational anomalies. Therefore, future efforts should prioritize comprehensive protocol coverage, enhanced realism in attack simulations, and thorough metadata annotations.

IV. Identified Gaps, Limitations and Future Directions

Upon reviewing the datasets, several critical limitations and gaps emerge. Firstly, existing datasets often lack

comprehensive protocol diversity, particularly regarding emerging IoT and industrial protocols such as OPC UA, BACnet, or KNX, despite their growing relevance. For example, KNX Dataset Bundle and LBNL-ETA/Brick focus narrowly on KNX and BACnet, respectively, while Edge-IIoTset and CIC APT IIoT 2024 partially address this with multi-protocol support. The Tennessee Eastman Process dataset, while valuable for process fault analysis, does not contribute to protocol diversity as it focuses on simulated chemical process data rather than network protocols.

Legacy protocol representation is equally critical: Modbus (covered in SWaT, HAI 1.0) and CAN-bus variants remain underrepresented, as do smart grid protocols like Wireless M-Bus. Similarly, while datasets like CIC APT IIoT 2024 include advanced persistent threats (APTs), most (e.g., IoT-23, N-BaIoT) focus on single-stage attacks, limiting their utility in evaluating real-world threat complexity. The Tennessee Eastman Process dataset's fault injection scenarios provide valuable anomaly detection cases but lack explicit cyberattack context. Data imbalance issues remain prevalent in N-BaIoT and CICIDS2017, where benign samples dominate, skewing model performance.

Furthermore, real-time operational data capturing realistic industrial scenarios, such as those involving high-fidelity cyber-physical interactions, remain underrepresented except in SWaT and HAI 1.0. The Tennessee Eastman Process dataset excels in operational fidelity with its detailed process simulations but lacks real-time cyberattack integration. Even these datasets lack continuous updates, reducing freshness for evolving threat landscapes.

To bridge these gaps, we propose the following design principles for future dataset development:

- **Time-Series Analysis**: Integrate time-stamped, high-frequency data streams (e.g., EPIC's sensor measurements and Tennessee Eastman's process variables) with contextual metadata (device state, process phase) to detect multi-stage APTs.
- **Real-Time Threat Detection**: Embed streaming formats (e.g., Apache Kafka pipelines) and sub-second latency labeling, addressing SWaT's static scripts and IoT Honeypot's unstructured logs while building on Tennessee Eastman's process-level temporal resolution.
- **Stratified Sampling**: Balance benign and attack samples proportionally, as seen in Edge-IIoTset, to counter imbalance in N-BaIoT and CICIDS2017.
- **Protocol-Agnostic Features**: Encode attributes like packet header semantics (CIC APT IIoT 2024) and cyber-physical process correlations (SWaT and Tennessee Eastman) to bridge legacy (Modbus) and emerging (OPC UA) protocol gaps.

Guided by these principles, future dataset development must prioritize the following objectives to address existing gaps and enhance practical utility:

- **Protocol Diversity**: Expand coverage to KNX, Wireless M-Bus, and OMS, building on CIC APT IIoT 2024's multi-protocol groundwork.
- **Attack Complexity**: Simulate APTs with stage-specific labels (e.g., reconnaissance, lateral movement) as partially achieved in CIC APT IIoT 2024, while incorporating process-level impacts demonstrated in Tennessee Eastman.
- **Data Balance**: Adopt Edge-IIoTset's stratified sampling to mitigate skew in N-BaIoT and CICIDS2017.
- **Operational Fidelity**: Integrate SWaT's actuator-sensor causality with TON-IoT's cross-domain telemetry and Tennessee Eastman's process dynamics for holistic cyber-physical datasets.
- **Community Collaboration**: Foster initiatives like Blaq_0 Hackathon but with standardized tasks to enhance reproducibility.

V. Conclusion

This paper reviewed publicly available datasets for ML-based intrusion detection systems in IoT and OT networks, highlighting their features, strengths, and limitations. While notable advancements exist—such as SWaT's cyber-physical fidelity, CIC APT IIoT 2024's APT coverage, TON-IoT's scalability, and Tennessee Eastman Process's detailed process anomaly detection—critical gaps remain in protocol diversity (e.g., KNX, OMS), attack complexity (e.g., multi-stage APTs), data balance, and operational freshness.

By adopting the proposed design principles including time-series analysis (demonstrated by Tennessee Eastman), real-time streaming, stratified sampling, and protocol-agnostic features, future datasets can address these gaps. For instance, combining SWaT's actuator-sensor causality with CIC APT IIoT 2024's multi-protocol coverage, TON-IoT's freshness, and Tennessee Eastman's process-level anomaly detection would yield a benchmark dataset for cross-domain IDS evaluation. Collaborative efforts to generate such harmonized datasets will be crucial for advancing ML-based IDS efficacy in rapidly evolving IoT and OT ecosystems.

Additionally, the authors are currently preparing simulated and real (automated) lab setups to generate datasets addressing these weaknesses, with a focus on APT emulation (inspired by CIC APT IIoT 2024), protocol-agnostic feature extraction (building on Edge-IIoTset), real-time cyber-physical interaction logging (extending SWaT's approach), and process-level impact analysis (following Tennessee Eastman's methodology).

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: grammar and spelling check, paraphrase and reword. After using this service, the authors

reviewed and edited the content as needed and take full responsibility for the publication's content.

REFERENCES

[1] A. Sikora, A. Walz, and L. Zimmermann, "Research Aspects for Secure Communication in the Industrial Internet of Things," *2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, 2020, pp. 284-289*.

[2] B. R. Kikissagbe and M. Adda, "Machine Learning-Based Intrusion Detection Methods in IoT Systems: A Comprehensive Review," *Electronics*, vol. 13, no. 18, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/18/3601

[3] M. Selem, F. Jemili, and O. Korbaa, "Deep Learning for Intrusion Detection in IoT Networks," *The Journal of Supercomputing*, 2025.

[4] M. Samantaray, R. C. Barik, and A. K. Biswal, "A Comparative Assessment of Machine Learning Algorithms in the IoT-Based Network Intrusion Detection Systems," *Decision Analytics Journal*, vol. 11, p. 100478, 2024.

[5] Z. Chen, J. Liu, Y. Shen, M. Simsek, B. Kantarci, H. T. Mouftah, and P. Djukic, "Machine Learning-Enabled IoT Security: Open Issues and Challenges Under Advanced Persistent Threats," *arXiv preprint arXiv:2204.03433*, 2022.

[6] L. T. Rajesh, T. Das, R. M. Shukla, and S. Sengupta, "Give and Take: Federated Transfer Learning for Industrial IoT Network Intrusion Detection," *arXiv preprint arXiv:2310.07354*, 2023.

[7] X. Sáez-de Cámara, J. L. Flores, C. Arellano, A. Urbieta, and U. Zurutuza, "Clustered Federated Learning Architecture for Network Anomaly Detection in Large Scale Heterogeneous IoT Networks," *Computers & Security*, vol. 131, p. 103299, 2023.

[8] U. P. D. Ani, H. He, and A. Tiwari, "Review of Cybersecurity Issues in Industrial Critical Infrastructure: Manufacturing in Perspective," *Journal of Cyber Security Technology*, vol. 1, no. 1, pp. 32–74, 2017.

[9] iTrust Labs, "iTrust Dataset Portal (SWaT, WADI, EPIC)," https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/, 2021, accessed: 2025-03-24.

[10] R. Arghandeh *et al.*, "LBNL-ETA Brick Schema Time Series Building Dataset," https://www.nature.com/articles/s41597-022-01257-x, 2022, accessed: 2025-03-24.

[11] S. Labs, "IoT-23 Dataset," https://www.stratosphereips.org/datasets-iot23, 2021, accessed: 2025-03-24.

[12] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, and Y. Elovici, "N-BaIoT: Network-based Detection of IoT Botnet Attacks," https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT, 2018, accessed: 2025-03-24.

[13] iTrust Labs, "Iot honeypot dataset," https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/, 2023, accessed: 2025-03-24.

[14] N. Moustafa, "TON-IoT Datasets," https://ieee-dataport.org/documents/toniot-datasets, 2020, accessed: 2025-03-24.

[15] M. A. Ferrag and L. Shu, "Edge-IIoTset: A New Benchmark Dataset for Edge-Based Intrusion Detection in IIoT Networks," https://www.kaggle.com/datasets/sibasispradhan/edge-iiotset-dataset, 2022, accessed: 2025-03-24.

[16] C. I. for Cybersecurity, "CICIDS2017 Dataset," https://www.unb.ca/cic/datasets/ids-2017.html, 2017, accessed: 2025-03-24.

[17] ——, "CIC APT IIoT Dataset 2024," https://www.unb.ca/cic/datasets/iiot-dataset-2024.html, 2024, accessed: 2025-03-24.

[18] X. Sáez-de Cámara, J. L. Flores, C. Arellano, A. Urbieta, and U. Zurutuza, "Gotham Testbed: A Reproducible IoT Testbed for Security Experiments and Dataset Generation," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 1, pp. 186–203, 2024.

[19] iTrust Labs, "Blaq_0 Hackathon ICS Dataset," https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/, 2023, accessed: 2025-03-24.

[20] A. Nickabadi, "Building Power Consumption Dataset," https://www.kaggle.com/datasets/arashnic/building-sites-power-consumption-dataset, 2021, accessed: 2025-04-14.

[21] C. Reinartz, M. Kulahci, and O. Ravn, "An Extended Tennessee Eastman Simulation Dataset for Fault-Detection and Decision Support Systems," *Computers Chemical Engineering*, vol. 149, p. 107281, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098135421000594