IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

Behavioral Fingerprinting of IoT Devices for Post-Attack Forensic Identification: A Review of Recent Advances, Datasets and Open Challenges

Jayashree Pandurang Darandale[1] andVina N. Chaudhary[2]
[1] M.Tech. Student, Department of Electronics and Telecommunication Engineering
[2]Assistant Professor, Department of Electronics and Telecommunication Engineering
G. H. Raisoni University, Amravati, India
[1] khushi.darandale1@gmail.com,[2] vinakadam17@gmail.com

Abstract : *The Internet of Things (IoT) has moved beyond home, hospital, and factory deployments to reach critical infrastructure, but most connected devices still lack instrumentation, have limited resources, and are vulnerable to compromise. Investigators trying to trace back what happened during an incident would have to piece together information about events from data gathered by different devices, gateways, and cloud services, where even static identifiers like IP and MAC addresses could be faked or randomized. Behavioral fingerprinting is an approach to identify devices and detect changes due to compromise through analysis of the statistical, temporal, and protocol regularities of a device's network traffic. This paper gives a structured survey on the use of behavioral fingerprinting for forensic identification of compromised IoT devices, based on forty-four peer-reviewed and preprint publications between 2024 and 2026. We divide the literature into five categories: traffic-based device identification; supervised and deep learning intrusion detection systems; unsupervised, one-class, and drift-sensitive anomaly detection; federated, explainable, and large-language model-aided approaches; and frameworks for IoT forensics. We introduce a behavioral features taxonomy, provide formal definitions of window aggregation, isolation score, and behavioral drift, and compare benchmark datasets and evaluation protocols. It is observed that there is considerable focus on real-time detection, whereas phase-based timeline reconstruction, attribution of attacks, detection of residual compromises, and independent device validation are underdeveloped. The proposed framework for forensic analysis of compromised devices is presented in terms of a generic phase-aware pipeline for partitioning the traffic into pre-attack, attack, and post-attack phases. Some of the open challenges, namely, the issue of admissibility of evidence, distinguishing between benign drift and compromise, adversarial mimicry, and cross-device generalization, are addressed.*

Keywords: Internet of Things, behavioral fingerprinting, digital forensics, device identification, intrusion detection, anomaly detection, isolation forest, behavioral drift.

I. INTRODUCTION

The Internet of Things (IoT) has emerged as a pervasive computing layer that connects common objects, clinical equipment, and industrial controllers to the Internet through their interaction with data and control services available online, and the number of these devices is expected to reach 41 billion in 2027 [1]. Smart cameras, voice assistants, wearables, PLCs, and sensors used in building automation systems not only generate data but can act upon it with minimal or no human intervention. While this growth has brought real-time automation, remote diagnostics, and intelligent control, it has also led to the creation of an attack surface like never before. IoT devices are usually constrained by limited computing power, small amounts of RAM, and battery life; moreover, most devices come with

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

default or weak credentials, while updates to firmware are rare, at best [2], [3]. These compromised devices are recruited into botnets used for large-scale DDoS attacks, port scanning, and C&C operations.

An extensive amount of literature addresses this issue with IDS that detect malicious traffic as it takes place. Detection, however, is just one of several questions that need to be answered in the wake of any security event. The other questions include determining the compromised node, start and end times of the compromise, type of attack performed, and persistence of infection even after the malicious activity ceases. Forensic analysis is the way to find answers to these questions. In regulated industries like healthcare and critical infrastructure, the answers also have implications regarding the breach-notification responsibilities, liability and remediation needs. Digital forensics, in its traditional form, is reactive and relies on the assumption of central logging, post-incident artifact acquisition, and the analysis of storage media; none of which apply to embedded IoT devices [5]. A comparison of IoT forensic frameworks led to the conclusion that none of the frameworks currently available meet all the requirements for investigation and that mechanisms for artifact acquisition and analysis are not yet standard [6].

The challenge is exacerbated by the fragmented nature of the evidence available throughout the IoT architecture, as shown in Fig. 1. At the device level, the memory is volatile, log entries on local storage are infrequent, and logs are often overwritten; at the cloud level, records related to the account and telemetry information are kept by the vendor and may not be available for investigation; and at the edge or gateway level, traffic and TLS handshake records, flow records and DNS transactions can be obtained through passive means. Furthermore, IoT installations are diverse from a protocol, operating system, and vendor perspective, meaning that specialized tools for acquisition of devices are rare. Consequently, the network view is the most reliable source of evidence and is also the least susceptible to tampering by the compromised device. Several systematic reviews have confirmed the increased adoption of machine learning in malware analysis, but have noted that the lack of forensic labels and anti-forensic techniques pose constant challenges [7], [8].

In this context, the use of static attributes like IP addresses, MAC addresses, and device self-naming is poorly supported by evidence since these are prone to be spoofed, randomized, or renumbered. The approach of behavioral fingerprinting aims to compensate for the drawbacks of the former technique by describing the device based on its activities instead of based on its claimed attributes: packet sizes, directions, intervals between arrival times, periodicity of keep-alives, combination of used protocols and ports, and connection state transitions. The said regularities emerge due to specific rules of firmware logic, hardware limitations, and the functional role of the device and are therefore hard to replicate. Recent research has demonstrated the ability to extract such behavioral fingerprints at terabit-per-second line speed using programmable switches [9], based solely on periodic key packets characterized by their size and direction [10], learned automatically from traffic grayscale sequences [11], expressed in terms of graphs of application layer patterns [12], and applied to new unseen devices in a zero-shot way [13].

In this context, statically assigned identities such as IP addresses, MAC addresses, and device names are of little evidentiary significance since they can be spoofed, randomized, or changed. Behavioral fingerprinting comes in handy in overcoming the deficiency associated with static IDs through profiling the device not based on what it is but how it behaves. The characteristics include the sequence of packet sizes and directions, inter-arrival timing, periodic keep-alive messages, protocol-port combination, and state transition properties. The aforementioned traits result from firmware logic, hardware limitations, and functional requirements of the device and are hence very hard to replicate. Behavioral fingerprints have been demonstrated to be achievable even at terabit per second data speeds in programmable switches [9], from periodic key packets which are defined in terms of their sizes and directions [10], automatically learned from traffic grayscale images [11], through graph representations of application layer patterns [12] and through devices seen for the first time through zero-shot learning [13].

Tapping into this potential imposes some special needs on the process of analysis. To estimate the baseline, one needs to use traffic prior to the occurrence of the attack; hence, the order of the observations cannot be scrambled but should be maintained in time. Attribution needs models capable of differentiating between attack families, and not just distinguishing the malicious from the benign traffic. The residual analysis calls for detectors that can detect behavior

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

that is unseen during training because, after an attack, the traffic might behave differently than during the attack. Lastly, the conclusions need to generalize to devices that were not included in model construction because an investigator typically does not have their history labeled.

This paper attempts to fill this gap via an organized survey of research papers written during the period between 2024 and 2026. The selection of suitable papers was done via IEEE Xplore, ScienceDirect, SpringerLink, MDPI, and arXiv, where keywords such as IoT, device fingerprinting, device identification, intrusion detection, anomaly detection, concept drift, and forensics were used to obtain possible papers. Papers were included based on whether they presented any technique related to device behavior detection, intrusion detection, or analysis of IoT traffic, leading to a total of forty-four papers distributed as shown in Fig. 2.

The contributions of this paper are fivefold. Firstly, it formalizes behavioral fingerprinting from a forensics perspective, including a feature taxonomy and the mathematical definitions of window aggregation, isolation-based anomaly score, and behavioral drift. Secondly, it provides an in-depth analysis of recent literature under five thematic aspects. Thirdly, it surveys the available benchmark datasets and evaluation protocols, especially focusing on data leakage and the leave-one-device-out approach. Fourthly, it introduces a general framework for phase-aware behavioral forensics, including three steps: Detection, Attribution, and Residual Anomalies Analysis. Lastly, it proposes open challenges and future research directions.

The rest of this paper is structured as follows. In Section 2, we introduce the basic concepts of behavioral fingerprinting. Section 3 covers recent literature. In Section 4, we present the datasets, evaluation protocols, and reference framework. Section 5 covers open challenges, and Section 6 concludes the paper.

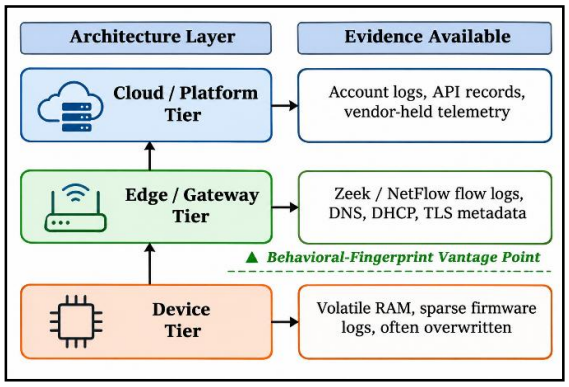

Fig. 1. Distribution of forensic evidence across the IoT architecture. The gateway tier offers the most consistently available, device-independent vantage point for behavioral fingerprinting.

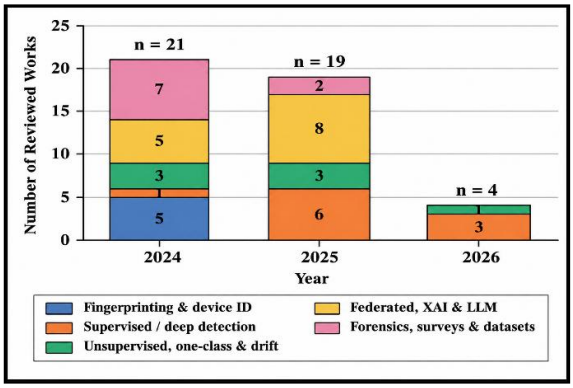

Fig. 2. Distribution of the forty-four works reviewed in this paper by research theme and publication year (2026 covers publications up to November 2026).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

II. FUNDAMENTALS OF BEHAVIORAL FINGERPRINTING

Prior to discussing each study, it is necessary to define what a behavioral fingerprint is, what observable characteristics it is composed of, and how it may be formulated in a way that is usable for forensic purposes. This chapter begins with a discussion comparing behavioral signatures to static identifiers, followed by a categorization of behavioral characteristics and the formulation of the quantities used in the rest of the review.

2.1 From Static Identifiers to Behavioral Signatures

Static identifiers are assigned as opposed to demonstrated. MAC addresses may be either cloned or randomized, IP addresses are dynamic due to DHCP leasing and NAT, and any self-identified information such as hostnames or user agents is entirely within the hands of whoever owns the device. On the contrary, behavioral signatures appear as the result of the interplay between firmware, hardware, and functionality. The smart plug reports state changes at a regular interval; the camera demonstrates steady and high volumes of upstream traffic; the voice assistant demonstrates bursts of cloud communications activated via wake-word detection. As those patterns are the product of normal operations of the device, in order to emulate the identity of a particular device, an attacker should not only forge the static identifiers but also mimic the traffic patterns, which is much more difficult [10], [11]. The very same attribute turns the behavioral signature into an important element for analysis of post-compromise communication, since the compromised device will demonstrate both its legitimate traffic and malicious traffic.

2.2 Behavioral Feature Taxonomy

The behavioral attributes may be classified under the attribute families outlined in Table 1 below. The volumetric and temporal attributes are obtainable from any network flow exporter and form the foundation of most flow-based detection methods; packet sequence attributes allow for more granular device-specific regularities and are the basis for identification with very high accuracy [10], [11]; finally, the application layer metadata attributes such as TLS handshake attributes and user-agent attributes have found their way into dual-use datasets used for both identification and anomaly detection [14].

Table 1. Taxonomy of behavioral features used for IoT fingerprinting and their forensic value.

Family	Representative features	Forensic value
Volumetric	Bytes and packets per flow (each direction), flow rate	Reveals floods, exfiltration, and traffic spikes
Temporal	Duration, inter-arrival time, periodicity, burstiness	Exposes heartbeats and timing changes after compromise
Packet sequence	Ordered packet sizes and directions, key packets	Fine-grained device identity; detects firmware tampering
Protocol/port	Transport protocol, service, destination port entropy	Flags anomalous ports and protocol misuse
Connection state	Zeek conn_state, history flags, failed handshakes	Characterizes scanning and rejected connections
Destination structure	Unique peers, fan-out, geographic and ASN spread	Identifies C&C contact and propagation
Application metadata	DNS queries, TLS handshake fields, user-agent	Links behavior to services and malware families

2.3 Formal Problem Formulation

Suppose that device k generates a time-series of flows, such that each flow i is represented by an m-dimensional vector of the attributes in Table 1:

$$x_i = \left[x_{i,1}, x_{i,2}, \ldots, x_{i,m}\right]^T \in \mathbb{R}^m \qquad (1)$$

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

Since individual flow values are noisy, behavior is typically characterized in windows. The window descriptor for the w-th window Wk,w for the device k is:

$$z_{k,w} = \frac{1}{|W_{k,w}|}\sum_{i\in W_{k,w}} x_i \qquad (2)$$

Phase-aware analysis assigns each window to one of three phases, $\phi \in$ {pre, att, post}, relative to the first and last labeled malicious events of the device. The phase centroid is then defined over the set Φk of windows in phase ϕ:

$$\mu_k^{(\varphi)} = \frac{1}{|\Phi_k^{(\varphi)}|}\sum_{w\in\Phi_k^{(\varphi)}} z_{k,w} \qquad (3)$$

Unsupervised residual detection commonly relies on the Isolation Forest, which isolates observations through random axis-parallel partitions. An anomalous window requires fewer partitions to be isolated, so its expected path length E[h(z)] across the ensemble is short. For a training sample of size n, the anomaly score is

$$s(z,n) = 2^{-\frac{E[h(z)]}{c(n)}} \qquad (4)$$

where c(n) is the average path length of an unsuccessful search in a binary search tree and H(·) denotes the harmonic number:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \qquad (5)$$

Numbers closer to 1 represent an anomaly, while numbers significantly lower than 0.5 represent normal behavior. In cases where the forest is trained using only pre-attack windows, high numbers in the post-attack window represent proof of residual compromise. Lastly, the amount of behavioral change caused by compromise can be measured using the Euclidean distance between pre- and post-attack centroids:

$$D_k = \|\mu_k^{(post)} - \mu_k^{(pre)}\|_2 = \sqrt{\sum_{j=1}^{m}\left(\mu_{k,j}^{(post)} - \mu_{k,j}^{(pre)}\right)^2} \qquad (6)$$

Dk becomes relevant only through standardization using statistical data obtained from the pre-attack phase because features have different scales. When a drift score tends towards zero, it means that the device reverted to its baseline state, while when it remains very high, it means that the identity has changed.

III. REVIEW OF RECENT LITERATURE (2024–2026)

These reviewed papers have been categorized into five themes, whose respective importance is illustrated in Fig. 2. Identification of devices ensures that behavior is distinctive; supervised and deep detection techniques ensure that compromise can be detected; unsupervised and drift-detection approaches deal with unknown and changing behavior; federated, explainable, and language model-based techniques deal with deployment, trust, and reasoning; and forensic approaches deal with evidentiary procedures. These themes are explored further in what follows, and some representative papers are compared in Table 2.

3.1 Traffic-Based Device Identification

Device fingerprinting scalability has become an issue of paramount importance. PoirIoT offloads fingerprint extraction to the data plane of programmable switches, thus achieving instant identification up to terabits per second throughput without additional latency overhead for standard flows [9]. DeviceRadar addresses Internet service provider (ISP) networks, where middleboxes mask five-tuples and flow features and resort to periodic key packets characterized only by their sizes and directions, encoded using a packet-size model and matched on programmable switches [10]. Feng et al. eliminate feature engineering by transforming device traffic into a sequence of grayscale images from which a deep model learns spatio-temporal fingerprints [11]. In contrast, GoNP represents device traffic by graph-based network patterns for UDP application-layer protocols [12]. Almost all identification models are based on the assumption that the set of devices is closed; ZEST addresses this restriction via a self-attention feature extractor and zero-shot learning to classify unseen devices [13]. These studies provide evidence that behavioral fingerprints are discriminative and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

computationally efficient. Nevertheless, these techniques have been developed primarily for device inventory and access control purposes, and no study considers the transformation of a fingerprint after the device compromise.

3.2 Supervised and Deep-Learning Detection

Supervised learning has been the prevalent choice for detecting attacks in IoT. In the IoT-23 dataset, Akif et al. applied a voting algorithm of Random Forest, XGBoost, k-nearest neighbors, and AdaBoost. They claimed that the hybrid models performed better than their single models for both binary and multiclass classification [15]. Gradient boosting is a desirable approach for digital forensics since it can deal with heterogeneous tabular features and provides feature importance. AlFuraih et al. proposed CNN feature extraction with an XGBoost classifier and applied leakage-aware feature selection with SHAP explanation on the CIC IoT-DIAD 2024 dataset [14].

There is extensive research regarding deep architectures on CICIoT2023. For example, Hizal et al. compared DNN, CNN, LSTM, and RNN based on binary and multi-class DDoS attack classification [16]; Al-Ghamdi and Alansari compared anomaly detection using CNN and RNN based on 33 attacks in the CICIoT2023 dataset [17]; Erskine used deep learning along with a multilayer perceptron for real-time intrusion detection and prevention systems [18]; and Andrade et al. presented lightweight CNN and LSTM detectors for large IoT networks [19]. Furthermore, Zhang et al. utilized a transformer alongside CNN-BiLSTM for solving problems such as class imbalance and feature extraction issues [20]; and Wasswa et al. used a vision transformer for classification by changing the flow features into single-channel images [21]. The feasibility of deep architectures on hardware was also analyzed: for instance, Mazinani et al. evaluated intrusion detection on constrained IoT devices using deep networks [22].

The comparison study by Wasswa et al. provides a helpful caveat to be aware of regarding the presumption of improved detection based on increased architectural complexity [23]. As seen in Fig. 3, the graph-based encoders performed worse than a simple variational autoencoder encoder with a multilayer perceptron. From the perspective of forensics, another limitation can be noted in this category of machine learning models: most of these classify flow or window independently; thus, they tell if there was an attack but not how the behavior of a certain device changed.

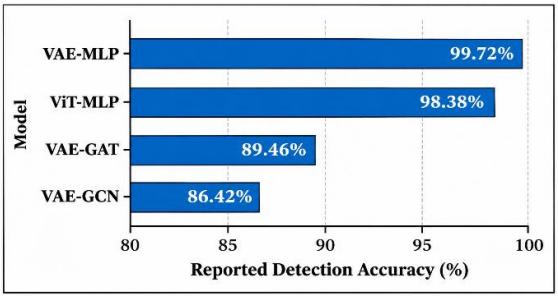

Fig. 3. Reported IoT botnet detection accuracy of four deep architectures, redrawn from the data reported by Wasswa et al. [23].

3.3 Unsupervised, One-Class and Drift-Aware Detection

Supervised detectors do not detect attacks that do not belong to any of the attack classes present during training data acquisition. This is problematic because the investigator does not know what has happened beforehand. One-class detectors only learn from legitimate behavior. Shi et al. used Isolation Forest and deep autoencoder techniques with only benign NetFlow examples in the training phase. They achieved perfect recall with precision exceeding 80%, showcasing the flexibility of one-class learners to changing malware [24]. Zahoor et al. evaluated Isolation Forest, One-Class SVM, and a hybrid scoring approach on the TON_IoT dataset and concluded that unsupervised lightweight approaches perform well for anomaly detection [25]. Sami and Abid reported a significantly lower computational burden of Isolation Forest in comparison to One-Class SVM on the TON_IoT thermostat dataset [26]. At the same time, Luthfi and Emigawaty performed Isolation Forest analysis on smart home traffic [27]. Jamshidi et al. implemented Isolation Forest scoring with autoencoders for efficient edge gateways [28].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

Changes in behavior occur for benevolent reasons too, including the update of firmware, changes in patterns of use, or seasonal activity. Chu et al. detected concept-drift points in IoT data streams using a combination of non-parametric detector ensembles, online Isolation Forest anomaly detection, and XGBoost classification optimized by whales and model update at detected points of drift [29]. Beshah et al. developed an online DDoS detection system based on detection and adaptation to drift points in IoT data streams [30]. In the mentioned works, drift is viewed as something that should be accommodated. But from a forensic point of view, any drift occurring in conjunction with attacks is a piece of evidence, and the task is to separate the two types of drift.

3.4 Federated, Explainable and LLM-Assisted Approaches

Privacy and data locality are the driving forces behind federated learning, where devices/gateways locally train the model and share the parameters alone. Vyas et al. reviewed privacy-preserving federated intrusion detection systems using federated learning techniques along with homomorphic encryption, differential privacy, and secure multi-party computation [31], while Isma'ila et al. listed limited training data and heterogeneous datasets among the major challenges in federated-based anomaly detection systems [32]. Federated learning was applied along with an optimized Isolation Forest ensemble for edge anomaly detection by Babu and Bagubali [33]. Federated solutions are suited for forensic readiness owing to localization of evidence, but pose issues with respect to chain of custody.

Explainability is mandatory in evidence that has to be explained to a court or incident response board. Gaspar et al. analyzed LIME and SHAP in a multilayer perceptron-based IoT intrusion detection system [34]; Wang et al. studied explainable techniques to help make decisions of the IoT intrusion detection system understandable by non-expert users [35]; Kumar and Thing studied advanced IoT intrusion detection systems based on networks with TRUSTEE and SHAP [36]; Gyawali et al. used counterfactual SHAP for actionable insight [37]; and Loi et al. employed Random Forest and SHAP for industrial IoT traffic [38]. In Sharma et al., explanations were used as part of defense, i.e., detecting adversarial inputs through SHAP attribution fingerprints [39].

Recently, large language models (LLMs) joined the fray. Al-Hammouri et al. employed signature detection in conjunction with GPT-2 for detecting zero-day attacks [40], Diaf et al. leveraged pretrained LLMs to predict upcoming attacks [41], and Jamshidi et al. utilized lightweight machine learning detectors alongside LLM reasoning at the edge gateway, maintaining under 75 J energy consumption [42]. According to Sudasinghe et al., a QLoRA-tuned LLaMA-1B model achieved an F1-score of 0.7124 on the CICIoT2023 dataset, which was on par with a Random Forest baseline at 0.7159, and retrieved augmentation allowed 42.63% accuracy on unseen attacks without further training [43]. It appears that LLMs are still not better detectors, but can serve as a means of constructing a forensic narrative.

3.5 IoT Forensic Frameworks and Synthesis

Forensic research in process-oriented investigation has primarily focused on preparation and procedure. The analysis performed by Mahmood et al. showed that the current frameworks fail to cover all the forensic needs of the Internet of Things [6], and Kuku supported the idea of real-time forensic analysis at the edge instead of only relying on post-hoc data acquisition [5]. General surveys on machine learning-based IoT security [1], [3] and deep learning malware forensics [7], [8] have mentioned the forensic aspect but approached it mostly as an extension of detection. Dataset initiatives like CICIoMT2024 have started profiling device states explicitly [44].

Table 2 provides a summary of representative studies. Three points emerge from this review. Firstly, identification and detection are almost never performed in a single framework for the evolving behavior of a device. Secondly, the temporal phase structure is underutilized since the order of data points is lost after shuffling before division, and this leads to an overestimation of performance. Lastly, the residual behavior of the device after the attack and the quantitative shift have not been quantified in the reviewed detection studies.

Table 2. Comparative summary of representative studies (2024–2026). ID = identification, Det = detection, Expl = explanation; F = forensic support.

Ref.	Technique	Data	Scope and forensic gap
[6]	Comparison of IoT forensic frameworks	Literature	Procedural; no behavioral analytics

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

[9]	In-network fingerprinting on switches	Home IoT traffic	ID at Tbps; no compromise analysis
[10]	Key packets (size, direction) + embedding	ISP traffic	Online ID; F not addressed
[11]	Traffic grayscale image sequences, DL	Public ID sets	Automatic fingerprints; ID only
[13]	Self-attention, zero-shot learning	IoT traffic	Unseen devices; ID only
[14]	CNN–XGBoost, leakage-aware, SHAP	CIC IoT-DIAD 2024	Det + Expl; no drift quantification
[15]	Voting RF / XGBoost / KNN / AdaBoost	IoT-23	Binary and multi-class Det; random split
[16]	DNN, CNN, LSTM, RNN	CICIoT2023	DDoS Det; no device attribution
[20]	Transformer + CNN-BiLSTM	IoT IDS sets	Imbalance handling; Det only
[23]	VAE-GNN vs. ViT-MLP vs. VAE-MLP	Botnet sets	Complexity vs. accuracy; Det only
[24]	One-class IF and autoencoder	NetFlow sets	Benign-only training; no timeline
[25]	IF, OCSVM, and fused CSAD	TON_IoT	Lightweight Det; no phases
[29]	Drift localization, online IF, WOA-XGBoost	Edge-IIoTset	Adapts to drift; drift not evident
[33]	Federated optimized IF ensemble	Edge IoT	Privacy-preserving Det; custody unclear
[38]	Random Forest + SHAP	CIC IoT-DIAD 2024	Det + Expl; no residual analysis
[43]	QLoRA LLM with RAG	CICIoT2023	Zero-shot Det; limited accuracy

IV. DATASETS, EVALUATION PROTOCOLS AND A REFERENCE PIPELINE

The credibility of any fingerprint claim is contingent upon both the data behind the claim and the way in which that data is divided into partitions for evaluation. This section contrasts the benchmark datasets employed by the literature under review, evaluates evaluation strategies and measures, and discusses a reference pipeline which combines the attributes highlighted in Section 3.

4.1 Benchmark Datasets

Table 3 presents an overview of IoT security datasets that are commonly used and their potential for behavioral forensics. The dataset named IoT-23, published by the Stratosphere Laboratory, consists of twenty-three captures of attacks and normal traffic by IoT devices recorded from 2018 to 2019 [15]. This dataset contains timestamps and retains continuity of scenarios, which makes it very appropriate for phase segmentation. The CICIoT2023 dataset offers scale and variety of attacks [16], [17], and CIC IoT-DIAD 2024 was designed specifically for device identification and anomaly detection [14], [38]. The CICIoMT2024 extends the coverage to medical devices, different protocols, and describes device lifecycle states [44].

Table 3. Comparison of benchmark datasets for IoT behavioral fingerprinting and forensics (compiled from dataset documentation and the cited studies).

Dataset	Scope	Attacks	Forensic suitability
IoT-23 [15]	23 captures of real devices, 2018–19; Zeek logs	Mirai, Okiru, Torii, C&C, DDoS, port scan	High: ordered per-scenario timelines
CICIoT2023 [16], [17]	Large lab topology of IoT devices	33 attacks in 7 categories	Medium: scale, weak device continuity

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

CIC IoT-DIAD 2024 [14], [38]	Dual-purpose ID and anomaly sets	DDoS, DoS, recon, web, brute force, spoofing, Mirai	High: joint ID and detection labels
CICIoMT2024 [44]	IoMT over Wi-Fi, MQTT and BLE	18 attacks in 5 categories	Medium–high: lifecycle profiles
TON_IoT [25], [26]	Telemetry, network and OS logs	Scanning, DoS, ransomware, injection, etc.	Medium: multi-source evidence
Edge-IIoTset [29]	Industrial IoT testbed	14 attacks in 5 threat groups	Medium: drift studies

4.2 Evaluation Protocols and Metrics

Detection quality is normally expressed using precision P, recall R, and the harmonic average of both, known as the F1-score, and can be determined by:

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN} \qquad (7)$$

$$F_1 = \frac{2PR}{P+R} \qquad (8)$$

Accuracy is not the right metric in IoT data, as the malicious flows can greatly outnumber the benign flows or vice versa; thus, macro-average F1 score and confusion matrix by class are better. Of no less significance is the protocol for splitting the data into train and test parts. With random splitting at the level of flows, almost identical flows that were captured from one device will appear in both train and test parts, yielding overly optimistic estimates; leakage-aware feature selection in [14] is a sign of increased awareness about this problem. The LODO protocol for generalization in forensics is:

$$A_{LODO} = \frac{1}{K}\sum_{k=1}^{K} A(f_{-k}, \mathcal{D}_k) \qquad (9)$$

where f−k denotes the model trained without device k, and $\mathcal{D}$k is that device's data. LODO approximates the realistic case in which an investigator examines a device that was never observed during model development, and it complements the zero-shot identification setting explored by ZEST [13].

4.3 Phase-Aware Forensic Reference Pipeline

Combining all the best from the discussed strands, Fig. 4 illustrates a generalized pipeline with phase awareness for the analysis of the aftermath of attacks. The flow logs are chronologically sorted, encoded, and cleaned, and segmented by devices into pre-attack, attack, and post-attack phases according to the start and end of malicious activities.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

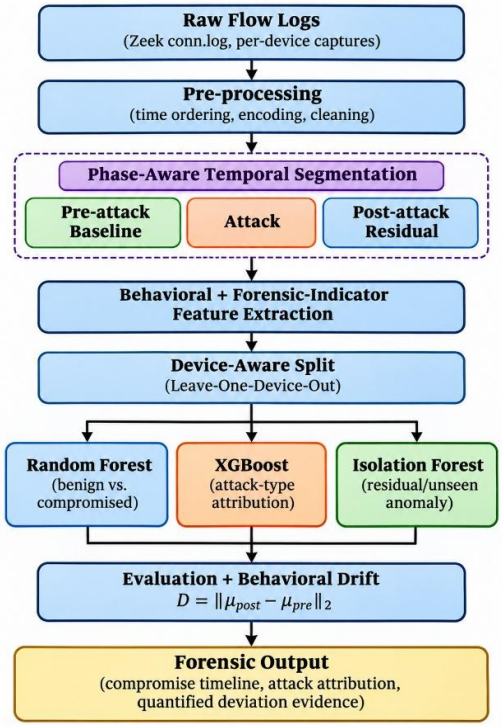

Fig. 4. Generalized phase-aware behavioral fingerprinting pipeline for post-attack forensic identification of IoT devices, combining supervised detection, attack attribution, one-class residual analysis and drift quantification.

Behavioral features presented in Table 1 are enhanced with explicit forensic features such as anomalous ports, traffic spikes, and abnormal connection states, which are analyzed in LODO. Then, three complementary models are run in parallel. Firstly, the Random Forest is employed to perform the binary classification of benign versus compromised, benefiting from the robustness of tree ensemble models shown in IoT-23 [15]. Secondly, the XGBoost classifier assigns attack category labels to compromised windows, including DDoS, C&C, C&C heartbeat, and port scanning attacks [14]. Thirdly, the Isolation Forest, trained on pre-attack windows alone, based on the one-class idea [24], [25], detects residual and novel anomalies after the attack. Finally, the drift score defined in Eq. (6) measures the identity change of each device. This is how we get the forensic report containing the compromise timeline, attack attribution, and evidence of deviation.

V. OPEN CHALLENGES AND FUTURE DIRECTIONS

While the above-mentioned studies provide very good building blocks, turning behavioral fingerprints into forensic use remains difficult due to many open questions. These will be outlined in the following subsections, together with potential research directions.

5.1 Evidential Admissibility and Explainability

The results of forensics must be replicable, interpretable, and defendable against challenges. Explanation methodologies post hoc, like SHAP and LIME, make interpretation better [34], [35], [38]; but interpretation itself varies run by run and by model [36]. Further research needs to develop criteria for stability of interpretations, cryptographically timestamping flow logs and model versions for continuity of chain of evidence and analysis pipelines with forensic process standards.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

5.2 Separating Benign Drift from Compromise

The drift-adaptive detectors adapt to changes in behavior to maintain their accuracy [29], [30]; however, in a forensics system, the reason for such a change must be explained. To distinguish between updates to the firmware, changes in behavior, seasonal effects, and changes resulting from the compromise and therefore drift, one needs a reference dataset for benign change and also a labeled dataset.

5.3 Cross-Device Generalization and Unseen Devices

Most detectors are trained using random splits, while the attackers come across detectors for which there is no training data. LOPO evaluations, zero-shot recognition [13], and baselines at the device type level should be made standard practices. Federated learning may enable the training on more diverse devices without centralized raw data [31]–[33].

5.4 Adversarial Mimicry and Anti-Forensics

A skillful adversary could modify the malicious traffic to appear as if it matches the device's legitimate fingerprint or even remove any fingerprints completely. Fingerprinting through explanations [39] and anti-forensic analysis [7] can serve as a base for further research, but a systematic evaluation of fingerprinting robustness is yet to be performed.

5.5 Resource Constraints and LLM-Assisted Investigation

Edge-side forensic readiness needs constant logging and lightweight scoring. The approaches that make use of isolation or autoencoders combined with isolation forest have low computational overhead [22], [26], [28]. LLMs have potential in translating phase-by-phase evidence into an investigator's language [42], [43]; however, because they tend to output false statements, grounding in flow evidence is necessary.

VI. CONCLUSION

The review of behavioral fingerprinting as a basis for post-attack forensic identification of IoT devices has been carried out based on 44 peer-reviewed and preprint publications from 2024 to 2026. The literature review shows that device behavior is distinctive enough for scalable identification, that supervised and deep learning models achieve high accuracy in attack detection, that one-class models are able to reveal novel anomalies, and that explainability, federation, and the use of language models start addressing the issues of trust, privacy, and reasoning. Nonetheless, the research efforts stay focused on real-time detection. The temporal structure of phases, the behavior that follows the attack, quantization drift, and device-independent validation are hardly ever addressed, despite being precisely what is needed for incident reconstruction. The framework of formalization, dataset comparison, and phase-aware pipeline proposed in this paper can be used to bridge that gap.

REFERENCES

[1] A. Alfahaid, E. Alalwany, A. M. Almars, F. Alharbi, E. Atlam, and I. Mahgoub, "Machine learning-based security solutions for IoT networks: A comprehensive survey," Sensors, vol. 25, no. 11, Art. no. 3341, May 2025, doi: 10.3390/s25113341.

[2] K. Kaur, A. Kaur, Y. Gulzar, and V. Gandhi, "Unveiling the core of IoT: Comprehensive review on data security challenges and mitigation strategies," Front. Comput. Sci., vol. 6, Art. no. 1420680, Jun. 2024, doi: 10.3389/fcomp.2024.1420680.

[3] A. Ghaffari, N. Jelodari, S. Pouralish, N. Derakhshanfard, and B. Arasteh, "Securing internet of things using machine and deep learning methods: A survey," Cluster Comput., vol. 27, 2024, doi: 10.1007/s10586-024-04509-0.

[4] M. Gelgi, Y. Guan, S. Arunachala, M. S. S. Rao, and N. Dragoni, "Systematic literature review of IoT botnet DDoS attacks and evaluation of detection techniques," Sensors, vol. 24, no. 11, Art. no. 3571, Jun. 2024, doi: 10.3390/s24113571.

[5] O. Kuku, "Real-time forensic analysis in internet of things," BOHR Int. J. Internet Things Artif. Intell. Mach. Learn., vol. 4, no. 1, pp. 29–38, 2025, doi: 10.54646/bijiam.2025.30.

[6] H. Mahmood, M. Arshad, I. Ahmed, S. Fatima, and H. ur Rehman, "Comparative study of IoT forensic frameworks," Forensic Sci. Int.: Digit. Investig., vol. 49, Art. no. 301748, Jun. 2024, doi: 10.1016/j.fsidi.2024.301748.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

[7] S. U. Qureshi, J. He, S. Tunio, N. Zhu, A. Nazir, A. Wajahat, F. Ullah, and A. Wadud, "Systematic review of deep learning solutions for malware detection and forensic analysis in IoT," J. King Saud Univ. – Comput. Inf. Sci., vol. 36, no. 8, Art. no. 102164, Oct. 2024.
[8] R. Darwish, M. Abdelsalam, and S. Khorsandroo, "Deep learning based XIoT malware analysis: A comprehensive survey, taxonomy, and research challenges," arXiv:2410.13894, Oct. 2024.
[9] C. Kuzniar, M. Neves, V. Gurevich, and I. Haque, "PoirIoT: Fingerprinting IoT devices at Tbps scale," IEEE/ACM Trans. Netw., 2024, doi: 10.1109/TNET.2024.3395278.
[10] R. Li, Q. Li, T. Lin, Q. Zou, D. Zhao, Y. Huang, G. Tyson, G. Xie, and Y. Jiang, "DeviceRadar: Online IoT device fingerprinting in ISPs using programmable switches," IEEE/ACM Trans. Netw., 2024, doi: 10.1109/TNET.2024.3398778.
[11] Y. Feng, Y. Zhang, H. He, W. Zhang, and D. Wang, "An IoT device identification method using extracted fingerprint from sequence of traffic grayscale images," IEEE Trans. Dependable Secure Comput., 2024, doi: 10.1109/TDSC.2024.3383159.
[12] L. Medury and F. Kandah, "GoNP: Graph of network patterns for device identification using UDP application layer protocols," in Proc. IEEE 49th Conf. Local Comput. Netw. (LCN), 2024, doi: 10.1109/LCN60385.2024.10639659.
[13] B. Wu, P. Gysel, D. M. Divakaran, and M. Gurusamy, "ZEST: Attention-based zero-shot learning for unseen IoT device classification," arXiv:2310.08036v2, Jan. 2024.
[14] D. AlFuraih, L. Mhamdi, and A. S. Karar, "Explainable hybrid CNN–XGBoost framework for multi-class IoT intrusion detection with leakage-aware feature selection," Appl. Syst. Innov., vol. 9, Art. no. 49, 2026.
[15] M. A. Akif, I. Butun, A. Williams, and I. Mahgoub, "Hybrid machine learning models for intrusion detection in IoT: Leveraging a real-world IoT dataset," arXiv:2502.12382, Feb. 2025.
[16] S. Hizal, U. Cavusoglu, and D. Akgun, "A novel deep learning-based intrusion detection system for IoT DDoS security," Internet Things, vol. 28, Art. no. 101336, Dec. 2024, doi: 10.1016/j.iot.2024.101336.
[17] A. M. Al-Ghamdi and M. M. Alansari, "Enhancing IoT security: A comparative study of CNN and RNN-based anomaly detection using the CICIoT2023 dataset," IAENG Int. J. Comput. Sci., vol. 52, no. 5, pp. 1424–1441, May 2025.
[18] S. K. Erskine, "Real-time large-scale intrusion detection and prevention system (IDPS) CICIoT dataset traffic assessment based on deep learning," Appl. Syst. Innov., vol. 8, no. 2, Art. no. 52, Apr. 2025, doi: 10.3390/asi8020052.
[19] I. Andrade, S. S. Mahadik, M. Mukherjee, P. M. Pawar, and R. Muthalagu, "Deep learning based intelligent IDS for large-scale IoT networks," arXiv:2603.16342, Mar. 2026.
[20] C. Zhang, J. Li, N. Wang, and D. Zhang, "Research on intrusion detection method based on transformer and CNN-BiLSTM in internet of things," Sensors, vol. 25, no. 9, Art. no. 2725, Apr. 2025, doi: 10.3390/s25092725.
[21] H. Wasswa, T. Lynar, A. Nanyonga, and H. Abbass, "IoT botnet detection: Application of vision transformer to classification of network flow traffic," arXiv:2504.18781, Apr. 2025.
[22] A. Mazinani, D. Antonucci, L. Davoli, and G. Ferrari, "Performance assessment of DL for network intrusion detection on a constrained IoT device," Future Internet, vol. 18, no. 1, Art. no. 34, 2026, doi: 10.3390/fi18010034.
[23] H. Wasswa, H. Abbass, and T. Lynar, "Are GNNs worth the effort for IoT botnet detection? A comparative study of VAE-GNN vs. ViT-MLP and VAE-MLP approaches," arXiv:2505.17363, May 2025.
[24] T. Shi, R. A. McCann, Y. Huang, W. Wang, and J. Kong, "Malware detection for internet of things using one-class classification," Sensors, vol. 24, no. 13, Art. no. 4122, 2024, doi: 10.3390/s24134122.
[25] A. Zahoor, W. Abbasi, M. Z. Babar, and A. Aljohani, "Robust IoT security using isolation forest and one class SVM algorithms," Sci. Rep., vol. 15, Art. no. 36586, Oct. 2025, doi: 10.1038/s41598-025-20445-4.
[26] M. S. A. Sami and M. Abid, "Unsupervised anomaly detection for smart IoT devices: Performance and resource comparison," arXiv:2511.21842, Nov. 2025.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

[27] A. Luthfi and Emigawaty, "Isolation forest-based anomaly detection in IoT smart home network traffic," ILKOM J. Ilmiah, vol. 18, no. 1, pp. 43–57, Apr. 2026, doi: 10.33096/ilkom.v18i1.3156.43-57.
[28] S. Jamshidi, F. Erfan, O. Abdul-Wahab, M. Bellaiche, and F. Khomh, "Lightweight autoencoder-isolation forest anomaly detection for green IoT edge gateways," arXiv:2511.18235, Nov. 2025.
[29] R. Chu, P. Jin, H. Qiao, and Q. Feng, "Intrusion detection in the IoT data streams using concept drift localization," AIMS Math., vol. 9, no. 1, pp. 1535–1561, 2024, doi: 10.3934/math.2024076.
[30] Y. K. Beshah, S. L. Abebe, and H. M. Melaku, "Drift adaptive online DDoS attack detection framework for IoT system," Electronics, vol. 13, no. 6, Art. no. 1004, Mar. 2024, doi: 10.3390/electronics13061004.
[31] A. Vyas, P.-C. Lin, R.-H. Hwang, and M. Tripathi, "Privacy-preserving federated learning for intrusion detection in IoT environments: A survey," IEEE Access, vol. 12, pp. 127018–127050, 2024, doi: 10.1109/ACCESS.2024.3454211.
[32] U. A. Isma'ila, K. U. Danyaro, A. A. Muazu, and U. D. Maiwada, "Review on approaches of federated modeling in anomaly-based intrusion detection for IoT devices," IEEE Access, vol. 12, pp. 30941–30961, 2024, doi: 10.1109/ACCESS.2024.3369915.
[33] A. Babu and A. Bagubali, "Federated learning with sailfish-optimized ensemble models for anomaly detection in IoT edge computing environment," IEEE Access, vol. 13, pp. 53171–53187, 2025, doi: 10.1109/ACCESS.2025.3554301.
[34] D. Gaspar, P. Silva, and C. Silva, "Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron," IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3368377.
[35] Y. Wang, M. A. Azad, M. Zafar, and A. Gul, "Enhancing AI transparency in IoT intrusion detection using explainable AI techniques," Internet Things, vol. 33, Art. no. 101714, 2025.
[36] A. Kumar and V. L. L. Thing, "Evaluating the explainability of state-of-the-art machine learning-based IoT network intrusion detection systems," arXiv:2408.14040, Aug. 2024.
[37] S. Gyawali, J. Huang, and Y. Jiang, "Leveraging explainable AI for actionable insights in IoT intrusion detection," in Proc. 19th Annu. Syst. Syst. Eng. Conf. (SoSE), 2024, pp. 92–97.
[38] P. Loi, D. Canavese, L. Regano, D. Maiorca, and G. Giacinto, "SHAP happens: An explainable IDS for industrial IoT networks," in Proc. IEEE 9th Forum Res. Technol. Soc. Ind. (RTSI), Tunis, Tunisia, 2025, pp. 71–76.
[39] D. P. Sharma, L. Xue, and X. Sun, "Enhancing adversarial robustness of IoT intrusion detection via SHAP-based attribution fingerprinting," arXiv:2511.06197, Nov. 2025.
[40] M. F. Al-Hammouri, Y. Otoum, R. Atwa, and A. Nayak, "Hybrid LLM-enhanced intrusion detection for zero-day threats in IoT networks," arXiv:2507.07413, Jul. 2025.
[41] A. Diaf, A. A. Korba, N. E. Karabadji, and Y. Ghamri-Doudane, "BARTPredict: Empowering IoT security with LLM-driven cyber threat prediction," arXiv:2501.01664, Jan. 2025.
[42] S. Jamshidi, A. Nikanjam, N. Shahabi, K. W. Nafi, F. Khomh, S. Keivanpour, and R. Herrero, "Think fast: Real-time IoT intrusion reasoning using IDS and LLMs at the edge gateway," arXiv:2511.18230, Nov. 2025.
[43] P. B. Sudasinghe, K. S. K. Liyanage, and H. S. G. Pussewalage, "Lightweight LLMs for network attack detection in IoT networks," in Proc. 7th Comput., Commun. IoT Appl. Conf. (ComComAp), Madrid, Spain, Dec. 2025 (arXiv:2601.15269).
[44] S. Dadkhah, E. C. P. Neto, R. Ferreira, R. C. Molokwu, S. Sadeghi, and A. A. Ghorbani, "CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT," Internet Things, vol. 28, Art. no. 101351, Dec. 2024.