

StateFi : Effectively Identifying Wi-Fi Devices through State Transitions

Abstract

Randomized MAC addresses aim to prevent passive device tracking, yet Wi-Fi management frames still leak structured behavioral patterns. Prior work has relied primarily on syntactic probe-request features such as Information Elements (IEs), sequence numbers (SEQ), or RSSI correlations, which degrade in dense environments and fail under aggressive randomization. We introduce StateFi, a fingerprinting framework that models device behavior as finite-state machines (FSMs), capturing both structural transition patterns and temporal execution logic. These FSMs are embedded into compact feature vectors that support efficient similarity computation and supervised classification. Across five heterogeneous campus environments, StateFi achieves **94–97%** accuracy for in-network fingerprinting using full management-frame FSMs. With probe-only FSMs, it re-identifies devices under MAC randomization with up to **97%** accuracy across large public datasets comprising more than a million frames. When looking at the discrimination accuracy of the model, StateFi reaches **98%**, outperforming the strongest prior signature by up to **17 percentage points**. These results demonstrate that FSM-level behavioral dynamics form a powerful and largely unmitigated side channel, stable enough to defeat randomization and expressive enough for robust, scalable device identification.

1 Introduction

Wi-Fi devices continuously emit unencrypted management frames for discovery, authentication, and association. Although essential for connectivity, these frames leak structured behavioral patterns that enable device fingerprinting and tracking even in the absence of explicit identifiers [3, 6, 9]. To mitigate such threats, modern operating systems deploy *MAC address randomization* [19], frequently rotating locally administered MACs during probe-based discovery. Yet randomization obscures only identifiers, not the underlying behavioral logic encoded in management-frame sequences, timing, and retry dynamics [13, 18]. Consequently, randomized identifiers provide incomplete protection against behavioral re-identification.

Prior efforts to defeat MAC randomization have relied predominantly on probe-request based signatures derived from Information Elements (IEs), timing cues, sequence numbers (SEQ), or RSSI correlations [13, 18, 19]. Although these features can perform well in controlled settings, their discriminative power degrades sharply in realistic environments with high contention, mobility, and heterogeneous AP deployments, and they become increasingly brittle under aggressive randomization [12]. Moreover, probe activity is inherently sparse and highly variable, further limiting the stability of such syntactic signatures.

In contrast, this paper introduces StateFi, which shifts from content-based features to behavioral dynamics and analyzes the full spectrum of management frames, not just

probe requests. We model device behavior as a *finite-state machine* (FSM) that captures how management-frame subtypes evolve over time, including both structural transition patterns (e.g., retries, self-transitions) and temporal execution logic (e.g., inter-state delays, scan pacing). These FSMs are then embedded into compact feature vectors that support scalable similarity computation and supervised classification.

Our key insight is that management-plane behavior arises from deep firmware and chipset design choices: scan strategies, association sequences, backoff timing, that remain stable across MAC randomization events and are difficult to obfuscate without harming performance. Through extensive real-world evaluation, we show that these FSM-derived behavioral signatures are both distinctive and persistent.

We evaluate StateFi in two operational regimes. In a *network-internal* setting where devices could even remain associated, full-management FSMs achieve **94–97%** accuracy across five heterogeneous campus environments, enabling high-precision behavioral integrity monitoring and intrusion detection. In the adversarial *pre-association* setting where devices emit randomized MACs, probe-only FSMs re-identify devices with **92–97%** accuracy across three large public datasets containing millions of management frames captured over long durations. Using the Infocom’21 [18] discrimination-accuracy metric, StateFi attains **93–98%**, outperforming the strongest published signature (IE+SEQ+RSSI) by up to **17%**. These gains persist across all sizes of time-window in which we perform MAC de-randomization, demonstrating robustness to contention, mobility, and burst fragmentation.

This paper makes the following contributions:

- **Behavioral Modeling via FSMs:** We introduce the first fingerprinting framework that captures both structural and temporal management-frame dynamics using finite-state machine modeling.
- **Compact and Scalable Embeddings:** We derive an efficient vector representation of FSMs that supports scalable distance-based matching and supervised classification.
- **Comprehensive Real-World Evaluation:** Across university campus captures and large public datasets, StateFi achieves high accuracy for in-network fingerprinting and under MAC randomization, improving on the state-of-the-art by up to 17%.
- **Revealing a Behavioral Side Channel:** We show that MAC randomization fails to conceal implementation-level behavioral fingerprints that could be exploited through FSMs, motivating the need for behavioral obfuscation in future wireless standards.

2 Related Work

Early work on Wi-Fi device fingerprinting focuses predominantly on extracting *content-based* features from probe requests, with particular emphasis on Information Elements (IEs). Approaches

such as [1, 10, 11, 14, 17, 18, 20] analyze transmission rates, SSIDs, vendor tags, and other IE fields to construct device signatures. Vanhoef et al. [20] systematically evaluate the discriminability and stability of these fields, proposing strategies for selecting robust IE subsets. Despite their historical prominence, IE-based techniques suffer from well-documented limitations: devices from the same vendor frequently share identical fields, operating system updates alter emitted IEs, and field-level identifiers can be easily obfuscated. Our work departs from these brittle metadata fingerprints by instead capturing persistent behavioral signals expressed through management-frame dynamics.

Beyond content analysis, several studies explore *timing-based* fingerprinting. Clock-skew approaches [5, 7] exploit low-level oscillator imperfections but require precise measurement hardware and extended observation intervals, reducing their practicality in real-world deployments. Other methods [8, 19] extract statistical timing features between frames, yet their performance degrades significantly when probe activity is sparse or when channel contention introduces jitter. In contrast, our framework captures temporal consistency across the *full management lifecycle*: authentication, association, and scanning, yielding fingerprints that are more stable and robust across heterogeneous environments.

Sequence-number-based correlation has also been examined as a means of linking frames emitted by the same device [18, 20]. The state-of-the-art Infocom’21 framework [18] combines sequence-number continuity, IE fields, and multi-sniffer RSSI observations using a minimum-cost flow formulation to associate randomized MAC addresses with physical devices. While effective in controlled settings, these methods depend on fragile syntactic fields and multi-sensor coordination. Other lines of research circumvent MAC randomization through protocol-level or cryptographic side channels (e.g., UUID-E reversal [20]), though such attacks rely heavily on specific implementation flaws.

Our approach diverges fundamentally from prior content-, timing-, and sequence-based techniques by modeling *behavioral dynamics* using finite-state machines (FSMs). Rather than just isolating individual frame types or metadata fields, we characterize the complete structure and timing of management-plane interactions. This includes probe exchanges as well as authentication and association behaviors, providing a holistic view of how devices execute the Wi-Fi management protocol in practice.

3 Data Collection and Preprocessing

Our evaluation uses eight datasets spanning both *in-house campus captures* and *external large-scale public traces*. The campus datasets consist of passive recordings of all management frames across five heterogeneous environments: Café, Laboratory, Classroom, Residence, and Restaurant, chosen to reflect diverse device densities, mobility patterns, and infrastructure setups. To assess robustness under strong MAC randomization, we additionally include three widely used public datasets: **Infocom’21** [18], **Cagliari** [16], and **MITIK** [2]. Together, these datasets enable a comprehensive evaluation under both *network-internal* and *pre-association* threat models.

3.1 Capture Settings and Dataset Overview

Campus captures were collected fully passively, ensuring device behavior remained unmodified. Sessions lasted 30–95 minutes and yielded over 2.6M packets. Table 1 summarizes the five internal environments that provide a realistic distribution of contention and mobility.

Table 1: Summary of the Campus dataset

Env.	Duration (min)	MACs	Packets	Mgmt.
Residence	80	1,526	304k	51k
Cafe	68	1,782	80k	21k
Laboratory	66	2,448	321k	35k
Laboratory (small)	30	227	34k	7k
Classroom	93	4,726	541k	59k
Public Sq.	95	16,645	52k	25k
Restaurant	85	4,735	207k	43k
Total	–	31,862	2.6M+	235k

To evaluate performance under randomized MACs, we also use three external benchmark datasets:

- **Infocom’21** [18]: high-density probe requests collected via 21 ceiling-mounted sensors on 2.4 GHz channel 1 across an ~8000 m² shopping mall floor. A single business hour contains ~20k frames and ~5k unique MACs.
- **Cagliari** [16]: a labelled “Individual Devices” subset with probe requests captured on channels 1/6/11 from 18 devices across six operating modes, each in 30-minute sessions.
- **MITIK** [2]: four 60-minute passive experiments using five supersniffers (25 radios) at La Rochelle University, producing 2.65M probe requests and responses and 5,425 unique MACs.

These external traces form the basis of our probe-only evaluations in Section 6.2, where the goal is to link randomized MAC addresses across bursts in the pre-association phase.

3.2 Privacy Considerations and Processing

All campus traces undergo a unified privacy-preserving sanitization workflow aligned with institutional ethics protocols and local data-protection regulations. We strip all personally identifiable fields: including MAC addresses, SSIDs, and absolute timestamps, retaining only anonymized metadata such as management-frame subtype and inter-arrival timing. Malformed or unrecognized frames are discarded, and only valid management, control, and data frames are preserved. To ensure that our analysis focuses exclusively on client-side behavior, we apply a multi-stage filtering pipeline: 20,615 of the 32,310 observed MACs (63.8%) are flagged as randomized based on local-administration bits and short-lived rotation patterns; and 265 MACs transmitting Beacon or Association Response frames, indicative of access-point behavior, are removed. This processing ensures that all subsequent fingerprinting models operate solely on client-side, implementation-driven behavioral characteristics.

3.3 Isolating Bursts and Ground Truth

Following the methodology from Infocom’21 [18] and BLEACH [13], we define a *burst* as a sequence of packets emitted by the same

MAC address where inter-packet gaps never exceed one second. This models natural scanning activity, where probe requests and related management frames occur in tightly spaced clusters.

Because the true device identity is not observable in any dataset, we rely on persistent MACs as ground truth proxies. For each such MAC, we treat every burst as an independent behavioral instance of the same physical device. Bursts belonging to the same MAC are assigned a unique anonymized identifier, enabling supervised learning and cross-burst consistency evaluation.

This burst-anchored ground truth is used uniformly across all datasets except **Cagliari** (where we already have the controlled ground-truth) and evaluation settings—supporting *full-management FSM analysis* for internal fingerprinting (Section 6.1) and *probe-only FSM analysis* for randomized MAC re-identification (Section 6.2).

4 FSM-Based Device Characterization

Wi-Fi communication follows the IEEE 802.11 state machine [4], which prescribes the logic governing device discovery, authentication, association, and roaming. These interactions unfold through management frames that remain unencrypted even under modern standards (e.g., WPA3) and are therefore observable in passive captures. Crucially, while the standard specifies *allowable* sequences, it does not dictate precise ordering, timing, or retry behavior. Vendors thus implement this logic differently to optimize power consumption, connection latency, roaming responsiveness, or regulatory compliance [20]. These implementation-level choices introduce subtle but stable variations in management-frame behavior, producing a rich and persistent behavioral footprint. Our framework models these behaviors using **Finite State Machines (FSMs)** to capture both structural and temporal dynamics that remain stable even under MAC randomization.

4.1 Modeling Device Behavior as an FSM

Let \mathcal{S} denote the set of all management-frame subtypes observed across the eight datasets (Section 3). For a device d , we formalize its behavior as a finite state machine:

$$\mathcal{F}_d = (Q, \Sigma, \delta, q_0),$$

where:

- $Q \subseteq \mathcal{S}$ is the set of states corresponding to management frame subtypes (e.g., ProbeReq, Auth, AssocReq),
- Σ is the timestamp-ordered sequence of frames emitted during a burst,
- $\delta : Q \times Q \rightarrow \mathbb{N}$ is a transition function recording observed transitions between successive subtypes,
- q_0 is the initial state, set to ProbeReq in probe-only FSMs and to the first observed subtype in full-management FSMs.

Transitions are extracted directly from passive traces: a directed edge $q_i \rightarrow q_j$ is recorded whenever frame subtype j immediately follows subtype i for the same device within a burst. Because our model is entirely empirical, capturing what the device *actually* does rather than what the protocol allows, it naturally encodes vendor- and OS-specific logic, retry patterns, scanning aggressiveness, timing behavior, and deviations introduced by energy-saving strategies.

Empirically, we observe significant structural variation across vendors and device classes. For vendor v , let μ_v denote the average number of transitions per FSM; differences in μ_v across the dataset reflect heterogeneous scanning intervals, probing frequency, and connection logic. These distinctions persist even when MAC addresses rotate rapidly, supporting the robustness of FSM-based characterization.

4.2 Constructing FSMs from Bursts

FSMs are built at the granularity of *bursts* (Section 3.3), which serve as short, coherent behavioral units. This construction supports both evaluation regimes in the paper:

- **Full-management FSMs (Section 6.1):** When MAC addresses remain stable (e.g., after association), we can capture complete management-plane interactions: probe traffic, authentication/association exchanges, reassociation attempts, and deauthentication logic. These FSMs enable high-precision fingerprinting and anomaly detection *within the network*.
- **Probe-only FSMs (Section 6.2):** Under MAC randomization, devices broadcast primarily probe requests and related scanning frames. To remain aligned with the adversarial pre-association setting, we construct FSMs solely from these probe-phase frames (requests and responses). Despite the reduced alphabet, probe-only FSMs retain rich behavioral structure, ordering of IEs, scanning patterns, bursts of identical probes, and chipset-specific retry dynamics. These fingerprints remain stable across randomized MAC addresses, which we will validate in the results section (Section 6).

For each burst, we parse the ordered sequence of management subtypes and create a directed, weighted transition graph. Consecutive occurrences of the same subtype naturally generate self-transitions (e.g., repeated ProbeReq frames), signaling differences in scanning aggressiveness or network-unavailability handling. These per-burst FSMs encode both structural and temporal consistency: in addition to transition frequencies, we record inter-state delays, which reflect implementation-level timing strategies.

Despite sharing the same protocol specification, devices produce structurally distinct per-burst FSMs due to chipset logic, firmware behavior, scheduling policies, and radio coexistence features. These subtle but stable differences form the foundation of StateFi’s ability to (i) fingerprint associated devices with high precision and (ii) re-identify devices across randomized MAC addresses. The burst-level FSM is therefore the core abstraction powering the accuracy and robustness demonstrated in Sections 6.

5 FSM-Based Behavioral Comparison and Fingerprinting

To compare device behavior across heterogeneous environments, StateFi models each device’s Probe Request sequence as an FSM. However, directly comparing FSM graphs is computationally expensive: graph isomorphism and subgraph matching require combinatorial exploration of node mappings, and differences in state cardinality, transition multiplicity, and burst length further complicate alignment. Such operations scale poorly with the number of devices and the thousands of FSMs produced in realistic traces.

Table 2: FSM-derived features.

Feature	Description
x_1	# unique FSM states (frame subtypes)
x_2	# total state transitions
x_3	# self-transitions ($s_i \rightarrow s_i$)
x_4	Entropy of transition probabilities
x_5	Transition rate ($x_2/\Delta t$)
x_6	Mean inter-burst time gap
x_7	Max sequence-number gap across bursts
x_8	Binary encoding of Probe Request IEs

To enable large-scale analysis, we instead embed each FSM into a compact numerical representation that preserves its key structural and temporal properties, allowing efficient similarity computation and seamless integration with supervised classifiers.

5.1 FSM Feature Representation

Each FSM is embedded into a real-valued feature vector summarizing its state structure, transition dynamics, and temporal behavior. Table 2 lists the features extracted for every aggregated burst.

Let $\mathbf{x}_i \in \mathbb{R}^d$ denote the FSM vector for the i -th burst group. Each vector contains seven FSM-derived features and a binary IE vector of a randomly chosen frame in the burst. This hybrid representation captures both temporal dynamics and protocol-level semantics.

5.2 Device-Pair Classification

To evaluate how reliably we predict that the two bursts originate from the same physical device, StateFi adopts a supervised learning pipeline. We construct pairwise similarity features between FSM vectors.

For two bursts with vectors \mathbf{x}_i and \mathbf{x}_j , we compute:

$$\mathbf{f}_{ij} = [\|\mathbf{x}_i - \mathbf{x}_j\|_2, 1 - \cos(\mathbf{x}_i, \mathbf{x}_j), \|\mathbf{x}_i - \mathbf{x}_j\|_1, \text{IE_sim}(\mathbf{x}_i, \mathbf{x}_j)],$$

where IE_sim is the normalized overlap of the IE bit vectors. These features provide complementary views of behavioral similarity: magnitude, direction, structural deviation, and protocol syntax.

Positive and Negative Pair Construction. Training data is created by pairing bursts associated with persistent (non-randomized) MAC addresses. For each MAC with at least two bursts, we randomly sample up to five *within-MAC* pairs as positives. Negative examples are created by pairing bursts from different MACs. We balance the dataset by subsampling both categories to a maximum of 5,000 total pairs per scenario.

Model Training and Hyperparameter Selection. We evaluate three standard binary classifiers:

- **Random Forest (RF)** with balanced class weights,
- **Support Vector Machine (SVM)** with RBF kernel,
- **Logistic Regression (LR)** with ℓ_2 regularization.

SVM and LR operate on standardized features using scikit-learn’s StandardScaler, while RF operates directly on raw features. For each scenario and burst-aggregation parameter P , we perform hyperparameter optimization via GridSearchCV. All classifiers (SVM, Logistic Regression, and Random Forest) and preprocessing tools

are implemented using the scikit-learn machine learning library [15]. The final model is trained on 70% of the constructed pair samples and tested on the remaining 30%, and we report accuracy, balanced accuracy, confusion matrices, and per-class precision/recall.

Summary. This accuracy-focused supervised learning framework enables StateFi to evaluate device-pair classification across heterogeneous datasets while avoiding the prohibitive cost of direct FSM graph comparison. We next present the detailed accuracy results for RF, SVM, and LR.

6 Results

We evaluate StateFi across complementary settings that reflect the two principal goals of behavioral fingerprinting: (i) *fingerprinting associated devices and detecting behavioral anomalies within a network*, and (ii) *re-identifying devices across MAC address randomization during pre-association scanning*. In all cases, accuracy is defined as the proportion of correctly matched communication bursts originating from the same physical device. All methods operate on FSM-based fingerprints constructed from individual bursts. Together, the following subsections demonstrate that StateFi provides both high-precision intrusive-detection capabilities *after* devices join the network and strong re-identification capabilities *before* association, where MAC randomization is intended to protect user privacy.

6.1 Accuracy In Fingerprinting devices

We first evaluate StateFi in a *network-internal* setting, where the goal is to fingerprint devices and detect anomalous or spoofed behavior even after *after* the association. In this scenario, MAC addresses are relatively stable (per SSID or session), and the objective is not to defeat address randomization, but to determine whether two FSMs correspond to the same associated device and to enable fine-grained behavioral intrusion detection.

To this end, we instrument five heterogeneous campus environments: Café, Laboratory, Classroom, Residence, and Restaurant; and construct FSMs from the *full set of observable management frames* associated with long-lived MAC addresses. This includes probe traffic, but also authentication, association, (re)association, and deauthentication frames exchanged as devices join and maintain connectivity with the campus network. Each FSM thus captures the complete management-plane behavior of a device within a given time window.

For every scenario, we derive pairwise distance features between FSM embeddings (Section 5.2) and train three supervised classifiers: Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR). The resulting accuracies are shown in Fig. 1.

Across all environments, RF yields the strongest performance, achieving approximately 94–97% accuracy, while SVM and LR lag slightly behind. These results demonstrate that full management-frame FSMs provide a stable and highly discriminative behavioral signature once a device has associated with an access point. In other words, even when MAC addresses are only session- or SSID-stable, StateFi can reliably distinguish devices and detect deviations in their management behavior, making it well-suited for network-side

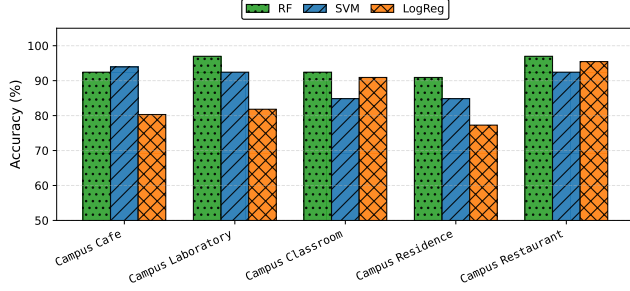


Figure 1: Fingerprinting accuracy across heterogeneous campus scenarios using RF, SVM, and LR models. Here, FSMs are constructed from the full management-frame exchange of associated devices, enabling high-precision behavioral fingerprinting for anomaly and intrusion detection.

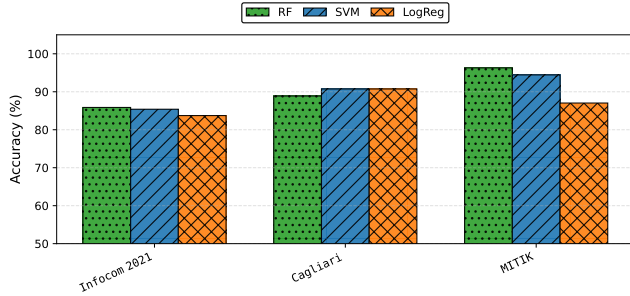


Figure 2: Accuracy in defeating MAC randomization for large public datasets (Infocom’21, Cagliari, MITIK) using probe-only FSMs. StateFi maintains high accuracy even when operating solely on pre-association probe behavior, illustrating its ability to re-link devices across randomized MAC addresses.

intrusion detection, device impersonation detection, and integrity monitoring.

6.2 Accuracy in Defeating MAC Randomization

We next turn to the setting most relevant to MAC address randomization: the *pre-association discovery phase*, where devices broadcast probe requests using frequently changing, locally administered MAC addresses. In this regime, devices are not yet associated with any AP, and the dominant packet type is the Probe Request. To align with this adversarial model, we restrict our FSMs to *probe requests and responses*, and ask whether StateFi can still link bursts generated by the same physical device across randomized MACs.

We evaluate this capability on three large public datasets: Infocom’21, Cagliari, and MITIK, which capture large numbers of mobile devices in realistic public settings. For each trace, we construct probe-only FSMs by utilizing individual bursts and derive the same pairwise distance features as in the campus experiments. We then train RF, SVM, and LR classifiers to predict whether two probe-only FSMs originate from the same underlying device, despite potential changes in the source MAC address. The resulting accuracies are reported in Fig. 2.

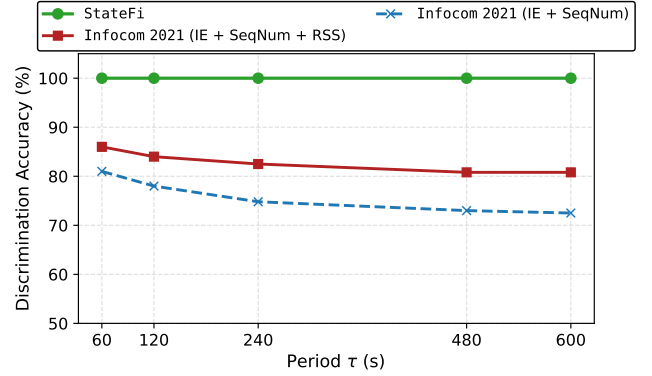


Figure 3: Discrimination accuracy comparison on the Infocom’21 dataset for $\tau \in \{60, 120, 240, 480, 600\}$ seconds. StateFi substantially outperforms IE- and SEQ-based baselines, including the full IE+SEQ+RSS signature.

The trends mirror those observed in the campus setting: RF consistently achieves the highest accuracy (typically 92–97%), with LR following closely and SVM underperforming both. Crucially, these results are obtained using *only* probe request and response behavior, without relying on stable association phases or long-lived MACs. This shows that the FSM of probe activity, i.e., how often, in what patterns, and with which information elements a device scans the environment, is sufficiently distinctive to re-identify devices across randomized MAC identifiers.

6.3 Comparison with State-of-the-Art

To contextualize the performance of StateFi within the broader literature on MAC address association, we compare our results with the Infocom’21 framework [18], which represents the most comprehensive and highest-performing approach reported to date. Infocom’21 integrates *all major feature families* used in prior work, including Information Elements (IEs), Sequence Numbers (SEQ), and multi-sniffer RSSI correlations, into a unified association strategy. As shown in [18], this combination outperforms techniques relying solely on timing, IEs, sequence numbers, or RSSI individually, and is widely recognized as the state-of-the-art for probe-to-device association under long-lived or session-stable MAC addresses. For these reasons, Infocom’21 serves as the strongest available baseline for evaluating whether behavior-based FSM fingerprints provide measurable improvements.

To enable a faithful comparison, we adopt the *discrimination accuracy* metric introduced in Infocom’21. Given a probe request p_i at time t_i , the metric evaluates whether a method can correctly associate p_i to a previous probe from the same physical device within the temporal window $[t_i - \tau, t_i]$. Using the same dataset utilized by [18], we report the accuracies for their IE+SEQ+RSS signature and the IE+SEQ signature, and compare them to our FSM fingerprint. Figure 3 presents the side-by-side results across standard window sizes $\tau \in \{60, 120, 240, 480, 600\}$ seconds.

StateFi delivers a substantial improvement over both Infocom’21 signatures. Across all time windows, our probe-only FSM fingerprint achieves **93–98% discrimination accuracy**, demonstrating both high precision and temporal stability. In comparison, the strongest Infocom’21 baseline (IE+SEQ+RSS) plateaus at **86%**, while IE+SEQ achieves at most **81%**, with both degrading as τ increases. This widening performance gap at longer windows reflects the sensitivity of IE and SEQ fields to burst boundaries, channel contention, and scanning dynamics; factors that introduce drift and fragmentation in syntactic signatures.

Overall, this comparison highlights a key advantage of StateFi as FSM-based behavioral modeling provides a stable, implementation-driven signature that retains discriminative power under realistic, noisy, and adversarial operating conditions.

7 Discussion

Behavioral Basis of FSM Fingerprints. Unlike approaches that rely heavily on volatile syntactic fields such as MAC addresses, IEs, or sequence counters, FSM-based modeling leverages deeper behavioral signals: the order in which devices traverse protocol states and the timing characteristics of those transitions. These patterns arise from firmware logic, chipset timing, scanning policies, and retransmission strategies, all of which remain relatively stable across environments and sessions. By encoding both transition structure and inter-state timing into unified feature vectors, FSMs capture robust behavioral traits that persist even when identifiers are frequently randomized.

Limitations and Countermeasures. FSM fingerprints could be weakened by protocol-level obfuscation (e.g., randomized scanning sequences, injected timing noise, or homogenized state-transition profiles) or by reducing visibility into management frames through emerging WPA3 extensions. However, these defenses impose real costs: excessive randomness may slow association, impair roaming, increase power consumption, or conflict with vendor-specific optimizations that improve performance. The practical feasibility of masking behavioral signatures without degrading user experience remains uncertain.

8 Conclusion

We introduced *StateFi*, a finite state machine (FSM) based fingerprinting framework that captures structural and temporal management frame behavior to derive stable device signatures, even under strong MAC randomization. Unlike prior syntactic approaches relying on brittle fields such as IEs or sequence numbers, *StateFi* leverages implementation-driven behavioral transitions that are significantly harder to mask or randomize. Across diverse real-world environments, full-management FSMs achieve **94–97%** in-network identification accuracy, while probe-only FSMs attain **92–97%** accuracy under randomized MACs. *StateFi* further reaches **93–98%** discrimination accuracy, outperforming the strongest existing signature by up to **12–17%** without requiring multi-sniffer deployments. These results position FSM-level behavioral modeling as a robust and scalable foundation for Wi-Fi fingerprinting, while underscoring the persistence of behavioral side channels and the

need for future privacy defenses that reduce FSM-level leakage without degrading network performance.

References

- [1] Mathieu Cunche, Mohamed-Ali Kaafar, and Roksana Boreli. 2014. Linking wireless devices using information contained in Wi-Fi probe requests. *Pervasive and Mobile Computing* 11 (2014), 56–69.
- [2] Fernando Dias de M. Silva, Fernando Molano Ortiz, Abhishek Kumar Mishra, Antoine Huchet, Mohammad Imran Syed, Luis Henrique M. K. Costa, Anne Fladenmuller, Aline Carneiro Viana, Yacine Ghamri-Doudane, and Nadjib Achir. 2025. *MITIK Dataset of WiFi pseudo-anonymised public management frames captured at La Rochelle University*. doi:10.57745/8LLY13
- [3] Baoqi Huang, Guoqiang Mao, Yong Qin, and Yun Wei. 2021. Pedestrian Flow Estimation Through Passive WiFi Sensing. *IEEE Transactions on Mobile Computing* 20, 4 (2021), 1529–1542. doi:10.1109/TMC.2019.2959610
- [4] Institute of Electrical and Electronics Engineers. 2020. IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.
- [5] Suman Jana and Sneha Kumar Kasera. 2008. On fast and accurate detection of unauthorized wireless access points using clock skews. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*. 104–115.
- [6] Zann Koh, Yuren Zhou, Billy Pik Lik Lau, Chau Yuen, Bige Tuncer, and Keng Hua Chong. 2020. Multiple-Perspective Clustering of Passive Wi-Fi Sensing Trajectory Data. *IEEE Transactions on Big Data* (2020), 1–1. doi:10.1109/TBDATA.2020.3045154
- [7] Felix Lanze, Andriy Panchenko, Thomas Engel, and Matthias Frank. 2012. Clock Skew Based Remote Device Fingerprinting Demystified. In *2012 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 813–818. doi:10.1109/GLOCOM.2012.6503232
- [8] Abhishek Mishra and Mathieu Cunche. 2021. Wi-Fi Fingerprinting and Privacy: A Survey. *IEEE Communications Surveys & Tutorials* 23, 2 (2021), 1234–1256. doi:10.1109/COMST.2021.3058392
- [9] Abhishek Kumar Mishra, Aline Carneiro Viana, Nadjib Achir, and Catuscia Palamidessi. 2021. Public Wireless Packets Anonymously Hurt You. In *2021 IEEE 46th Conference on Local Computer Networks (LCN)*. 649–652.
- [10] Abhishek Kumar Mishra and Mathieu Cunche. 2024. Characterizing probe request bursts to efficiently count Wi-Fi devices with randomized MACs. In *EuCNC-6G Summit-2024 European Conference on Networks and Communications & 6G Summit-Special Sessions*. 1–3.
- [11] Abhishek Kumar Mishra, Samuel Pélissier, and Mathieu Cunche. 2024. Fingerprinting connected Wi-Fi devices using per-network MAC addresses. In *International Symposium on Foundations and Practice of Security*. Springer, 295–303.
- [12] Abhishek Kumar Mishra, Aline Carneiro Viana, and Nadjib Achir. 2023. Introducing benchmarks for evaluating user-privacy vulnerability in WiFi. In *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE, 1–7.
- [13] Abhishek Kumar Mishra, Aline Carneiro Viana, and Nadjib Achir. 2024. Bleach: From WiFi probe-request signatures to MAC association. *Ad Hoc Networks* 164 (2024), 103623.
- [14] Jeffrey Pang, Ben Greenstein, Ramakrishna Gummadi, Srinivasan Seshan, and David Wetherall. 2007. 802.11 user fingerprinting. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. 99–110.
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [16] Lucia Pintor, Alessio Chirigu, and Luigi Atzori. 2025. *A dataset of labelled device Wi-Fi probe requests for MAC address de-randomization - 2024*. doi:10.17632/5tvnwhtsj2p.2
- [17] Pieter Robyns, Bram Bonné, Peter Quax, and Wim Lamotte. 2017. Noncooperative 802.11 mac layer fingerprinting and tracking of mobile devices. *Security and Communication Networks* 2017 (2017).
- [18] Jiajie Tan and S.-H. Gary Chan. 2021. Efficient Association of Wi-Fi Probe Requests under MAC Address Randomization. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*. IEEE, 1–9. doi:10.1109/INFOCOM42981.2021.9488769
- [19] Mathy Vanhoef, Célestin Matte, Mathieu Cunche, Leonardo S. Cardoso, and Frank Piessens. 2016. Why MAC Address Randomization is Not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In *ACM on Asia CCS (Xi’an, China) (ASIA CCS ’16)*. 413–424.
- [20] Mathy Vanhoef and Frank Piessens. 2016. Advanced Wi-Fi attacks using commodity hardware. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, 545–559.