

Reducing Mobility Trace Identifiability via Stochastic Roaming in IEEE 802.11 Networks

Juan Manuel Montes-Lopez, Isabel Schweim, Marco Gramaglia, Pablo Serrano

Abstract—Wireless networks continuously generate mobility traces through device associations with access points (APs), forming sequential patterns that can uniquely characterize user behavior. This raises significant privacy concerns, as such patterns enable device re-identification even when explicit identifiers are removed. Existing privacy mechanisms predominantly focus on anonymizing identifiers, overlooking information leakage contained in mobility dynamics. In this work, we analyze an stochastic association mechanism that introduces controlled randomness into the AP selection process to reduce the identifiability of mobility traces. The mechanism perturbs transition patterns while preserving realistic connectivity behavior and allows tuning to balance privacy and utility. We evaluate the approach using real mobility traces from approximately 18 500 devices. In the baseline setting, sequence-based re-identification attacks achieve a top-10 accuracy of 26%. When applying the proposed stochastic roaming mechanism, this accuracy decreases to 1% in a controllable manner. These results demonstrate that shaping the association process constitutes an effective and practical strategy to mitigate mobility-based identification risks, moving beyond identifier obfuscation toward protecting the structural information embedded in wireless mobility traces.

Index Terms—wireless networks, mobility privacy, differential privacy

I. INTRODUCTION

Wireless networks have become a fundamental component of modern digital infrastructures, providing continuous connectivity across campuses, enterprises, and urban environments [1]. In these settings, devices associate with multiple access points (AP) as users move through space. As a result, network infrastructures continuously generate large volumes of mobility-related data in the form of AP association logs, which reflect the spatiotemporal behavior of connected devices [2].

While such data is essential for network management, optimization, and security monitoring, it also raises significant privacy concerns. Even when explicit identifiers such as MAC addresses or user identities are removed or pseudonymized, mobility traces may still reveal distinctive behavioral patterns [3], [4]. Prior studies have shown that sequences of visited locations can act as strong behavioral fingerprints, enabling adversaries to re-identify devices or link anonymized traces to previously observed mobility profiles [5], [6].

Existing privacy mechanisms in wireless networks primarily focus on hiding device identifiers. A prominent example is MAC address randomization, which prevents

long-term tracking by periodically changing the device identifier visible to the infrastructure [7]. However, these approaches do not address the information contained in the mobility patterns themselves. As a consequence, an adversary with access to mobility data may still perform behavioral matching attacks based on the similarity of user activity [8] even under MAC address randomization.

Prior work has shown that human mobility traces are highly distinctive due to their regularity and sparsity, enabling the identification and re-identification of users even from a small number of observations [2]–[4], [9]. Building on this insight, subsequent studies have demonstrated that sequences of locations or related signals can act as behavioral fingerprints, allowing adversaries to match anonymized traces to previously observed user profiles [5], [6]. In wireless networks, such mobility patterns naturally arise from device associations with APs, exposing fine-grained spatiotemporal information that can be collected through infrastructure logs or passive monitoring. In particular, profile-based attacks exploit recurring user behavior to link new, unseen traces to known identities [8], illustrating how mobility patterns can serve as reliable behavioral fingerprints. Existing solutions, such as MAC address randomization, primarily aim to obscure device identifiers [7], but do not address the information leakage inherent in mobility patterns. Although prior work has explored controlling AP association decisions [10], [11], these approaches focus on performance objectives (e.g., throughput or load balancing) rather than privacy, leaving this aspect unaddressed.

In this work, we investigate the privacy benefits of modifying the association process in WLANs. The studied mechanism, instead of selecting the next AP according to the device’s default policy, we analyze the a stochastic association mechanism that perturbs transition patterns between APs. This controlled randomization reduces the identifiability of mobility traces while preserving realistic network behavior. We formalize this mechanism as a probabilistic transition model and analyze its privacy guarantees under a local differential privacy framework, highlighting the trade-off between privacy protection and trace fidelity. To evaluate its effectiveness, we apply the mechanism to real mobility traces derived from network logs and assess its robustness against sequence-based re-identification attacks, showing a significant reduction in attack accuracy.

The main contributions of this work are as follows:

- We analyze the privacy risks associated with mobility traces derived from wireless association logs and demonstrate the effectiveness of sequence-based re-identification attacks.
- We analyze how stochastic perturbation of AP transitions reduces the identifiability of mobility pat-

J. M. Montes-Lopez, M. Gramaglia, and P. Serrano are with Departamento de Ingeniería Telemática, Universidad Carlos III de Madrid, 28911 Leganés, Spain. I. Schweim is with IMDEA Networks Institute, 28918 Leganés, Spain. Emails: {jumontes, mgramagl, pablo}@it.uc3m.es, isabel.schweim@networks.imdea.org.

terns, and quantify the resulting privacy gains under a local differential privacy framework.

- We evaluate the privacy and utility trade-offs of the proposed approach using real-world mobility traces from a campus wireless network, by simulating the effect of the mechanism on observed association sequences.

II. DATASET

This section describes the dataset used in this study and the preprocessing steps applied to reconstruct device mobility traces from RADIUS [12] accounting logs collected in a campus WLAN. We first present the data collection process, then describe the preprocessing step used to extract mobility sequences, and finally report the main statistics of the resulting dataset.

A. Data Collection

The dataset used in this study was collected from the **eduroam** wireless network deployed at the Universidad Carlos III de Madrid (UC3M). The network infrastructure records authentication and accounting events through a RADIUS server, which maintains logs of device associations with APs.

Each time a device associates or re-associates with an AP, an accounting record is generated by the RADIUS server. In addition to these association events, the server periodically updates the status of connected devices through accounting messages generated at intervals of at most 15 minutes. These periodic updates report that the device remains connected to the same AP. Consequently, user mobility across the network can be reconstructed. In this setup, transitions between APs correspond to mobility events and 15-minute updates correspond to stationary periods.

Each RADIUS record includes a timestamp, a user identifier, the MAC address of the device, and the MAC address of the AP serving the device. For privacy and ethical considerations, both the user identity and the device MAC address were pseudonymized by the system administrators using an MD5 hashing function before the data was made available for analysis. This transformation preserves the ability to consistently track device activity over time while preventing the direct identification of individual users or devices.

B. Data Preprocessing

To focus on realistic user mobility, the analysis is restricted to devices whose behavior is consistent with user-carried equipment. First, we remove devices that remain associated with a single AP throughout the entire observation period, as these typically correspond to stationary infrastructure such as printers, sensors, or fixed workstations. Additionally, we filter out devices that exhibit predominantly long stationary intervals (around 15 minutes) relative to short transition intervals, since user-carried devices are expected to alternate between movement and shorter association periods across multiple locations.

After applying these preprocessing steps, the dataset is reduced from 19 326 to 18 571 devices, removing 755 static devices, which account for only 4% of the dataset.

Statistic		Value
# Records		5 360 870
# Days		30
# APs (across 7 buildings)		285
Visited APs per device	Per day (avg.)	12.28
	Per week (avg.)	25.48
Visited buildings per device	Per day (avg.)	2.83
	Per week (avg.)	3.94

TABLE I: Summary statistics of device mobility in the dataset.

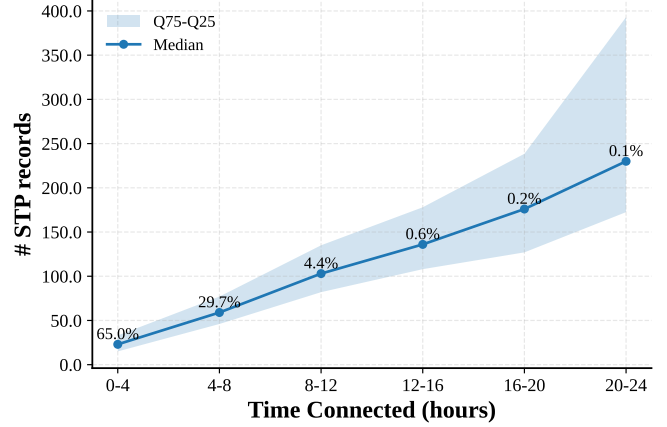


Fig. 1: Time connected vs. number of STP tuples, binned by device-day. Percentages above points indicate bin support.

C. Dataset Statistics

After preprocessing, the dataset consists of mobility traces capturing sequences of AP associations for a set of pseudonymized devices that are likely mobile. Each trace is represented as a sequence of spatiotemporal tuples (STP), where each tuple specifies the serving AP and the corresponding association interval.

The dataset spans 30 days of observations collected from the campus wireless infrastructure, during which we identified 18,571 distinct devices. The deployment includes 285 APs distributed across 7 buildings, covering the main academic and administrative areas of the university.

Table I summarizes key mobility statistics. On average, devices connect to 12.28 APs per day and 25.48 per week, indicating frequent transitions at the access point level. However, this mobility is more constrained at the spatial level: devices visit only 2.83 buildings per day and 3.94 per week on average, out of a total of 7 buildings in the campus. This suggests that, while users exhibit high variability in AP associations, their movement across buildings remains relatively limited, reflecting repetitive mobility patterns.

Figure 1 illustrates the relationship between connection duration and the number of STPs per device-day, where each STP corresponds to an association log.

As expected, the number of STPs increases with connection time. Notably, a substantial portion of the dataset (65.0%) corresponds to devices connected between 0 and 4 hours, which generate around 40 STPs on median. This

indicates that a large fraction of users produce mobility traces of moderate length, containing a non-negligible amount of spatiotemporal information. Variability across users is reflected in the interquartile range, which widens for longer connection durations.

From a privacy perspective, longer connection durations increase exposure by providing richer mobility traces, thereby improving the likelihood of successful identification by an attacker.

III. THREAT MODEL

The open and distributed operation of WLANs inherently exposes communication patterns to observation, enabling a broad range of security and privacy threats, including eavesdropping, spoofing, and rogue APs [13], [14]. Even when strong cryptographic protections are in place, adversaries can still extract sensitive information by passively observing network activity. In particular, a sniffing adversary can collect sequences of AP associations over time and infer mobility patterns from these traces. [15] Such patterns may enable the identification, tracking, or profiling of users, revealing the real identity of the device owner.

In this work, the threat model considers an adversary capable of observing WLAN association requests, which correspond to the sequence of Basic Service Set Identifiers (BSSIDs) to which a device connects over time. These association events form a temporal sequence that reflects the mobility behavior of the device within the network coverage area. The adversary's objective is to link short observed snippets of activity to the device that generated them.

The attacker is assumed to possess historical traces from a large number of devices in the network. Such traces may be obtained through previously collected network logs, monitoring infrastructure, or partial access to network management systems. Using this historical data, the adversary constructs behavioral profiles that capture the typical mobility patterns of each device, represented as sequences of AP associations over time. To emulate this setting, we partition the dataset on a per-user basis using a temporal split, where earlier days are used to build the attacker's knowledge and later days are reserved for evaluation.

When a new sequence of associations is observed, the adversary attempts to determine the originating device by comparing the observed snippet with the stored profiles. This comparison may be performed using similarity measures, probabilistic models, or machine learning techniques that evaluate how likely the snippet is to have been generated by each candidate device. The devices are then ranked according to this likelihood, and the most probable candidate is selected as the predicted origin of the observed activity.

Importantly, the attacker does not rely on explicit identifiers (e.g., MAC addresses) for re-identification. Instead, such identifiers are only used to associate observed network activity with a single device, thereby enabling the extraction of behavioral patterns that serve as the basis for re-identification. As noted in [16], Android devices employ persistent MAC randomization by default, meaning that the same pseudonymous MAC address is

used for a given network. This behavior facilitates long-term tracking and the collection of mobility data. Prior work on mobility trace analysis has shown that such patterns often exhibit strong regularity and uniqueness, making them a powerful basis for re-identification, even when explicit identifiers are removed from the released data [17].

IV. METHODOLOGY

A. Sequence Representation

Given a dataset of traces capturing device associations with different APs as users move through the network, each device can be represented as a spatiotemporal sequence of association events. Formally, for a given device d , we define a sequence

$$S_d = (a_1, a_2, \dots, a_T) \quad (1)$$

where $a_t \in \mathcal{A}$ denotes the AP associated with the device at time step t , and $\mathcal{A}$ represents the set of all APs in the network. These sequences capture the mobility patterns of users within the network infrastructure.

As described in the threat model, given a collection of historical sequences from multiple devices, the goal of the attacker is to identify which device generated an observed subsequence of associations. In practice, the attacker observes a snippet of length L from an unknown device and attempts to match it against the historical behavioral profiles of known devices.

B. Identification Attacks

In this work, we evaluate two identification attacks that exploit the regularity of device mobility patterns.

- **n-gram attack.** This attack builds from the trace of a device 1 the frequency of subsequences of length n (i.e., n -grams) observed in its historical trajectory. Intuitively, these n -grams capture recurring local patterns in device mobility. Given a query snippet, the attacker extracts its n -grams and compares them with the stored profiles of all candidate devices using a similarity measure, (e.g., cosine similarity), following an approach similar to [6]. The device whose profile is most similar to the query is selected as the predicted identity.
- **Markov likelihood attack.** This attack models each device as a first-order Markov process over APs. For each device, the attacker estimates a transition probability matrix from the previously observed traces, capturing how likely it is to move from one AP to another. Given a query sequence, the attacker computes the likelihood of the sequence under each device-specific Markov model. The device whose model assigns the highest likelihood to the observed sequence is selected as the predicted identity. This approach is motivated by prior work showing that transition dynamics capture a significant fraction of the identifying information in mobility traces [17], [18].

To illustrate the identification process, consider an attacker with historical mobility traces from three devices:

- d_1 : frequently transitions between $AP_1 \rightarrow AP_3 \rightarrow AP_2$

- d_2 : frequently transitions between $AP_2 \rightarrow AP_4 \rightarrow AP_6$
 - d_3 : moves more heterogeneously, $AP_1 \rightarrow AP_5 \rightarrow AP_2$
- Then attacker, observes a query snippet:

$$q = (AP_1, AP_3)$$

and aims to infer which device most likely generated it.

n-gram attack example. In this attack Assume $n = 2$ (bigrams). The attacker represents each device by the set (or frequency profile) of observed bigrams and compares them with the query:

- d_1 : contains the bigram (AP_1, AP_3) as part of a frequent pattern $\rightarrow$ *high similarity*
- d_2 : no observed transition $(AP_1, AP_3) \rightarrow$ *low similarity*
- d_3 : shares the starting point AP_1 but not the transition to $AP_3 \rightarrow$ *medium similarity*

Thus, the query directly matches a characteristic bigram of d_1 , while it is absent or only weakly related to the other devices. As a result, d_1 obtains the highest similarity score, and the query is attributed to d_1 .

Markov likelihood attack example. In this other scenario, the attacker models each device as a first-order Markov chain by estimating transition probabilities and evaluates the likelihood of the query:

- d_1 : For this device, $P(AP_3 | AP_1)$ is high due to frequent observations, resulting in *high likelihood*
- d_2 : On the other hand, transition $AP_1 \rightarrow AP_3$ is not observed for this device, resulting in *low likelihood*
- d_3 : For a more mobile device, transition $AP_1 \rightarrow AP_3$ may exist but with low frequency, resulting in a *medium likelihood*

Consequently, the likelihood of observing q under d_1 's model is significantly higher than under the other models. The attacker therefore selects d_1 as the most likely source.

C. Evaluation Metrics

To evaluate the effectiveness of the identification attacks, we measure how accurately an attacker can associate an observed sequence with the correct device. Given a query sequence, the attacker computes a score for each candidate device and produces a ranked list of devices ordered from the most to the least likely identity.

- **Top- k accuracy.** Top- k accuracy measures the fraction of queries for which the correct device appears within the top k positions of the ranked candidate list. Let $\mathcal{Q}$ denote the set of query sequences, and let $\text{rank}(q)$ be the position of the correct device for a query q . The metric is defined as

$$\text{Top-}k = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \mathbf{1}[\text{rank}(q) \leq k],$$

where $\mathbf{1}[\cdot]$ is the indicator function, which equals 1 if its argument is true and 0 otherwise [19]. This metric captures the attacker's ability to narrow the identification to a small set of k candidate devices, even when the exact match is not ranked first.

- **Random baseline.** To contextualize the attack performance, we compare the results with a random baseline in which the attacker ranks devices

uniformly at random. Let N be the total number of candidate devices. Under random ranking, the probability that the correct device appears within the first k positions is

$$\text{Top-}k^{\text{rand}} = \frac{k}{N}.$$

This baseline provides a lower bound for identification performance and assess whether the attacks significantly outperform random guessing.

V. PRIVACY-PRESERVING ASSOCIATION MECHANISM

Now, we present the proposed privacy-preserving association mechanism. We first describe how controlled randomness is introduced into the AP selection process, and then formalize the mechanism as a stochastic transition model governed by the privacy parameter ε . Finally, we analyze its privacy guarantees under local differential privacy.

A. Mechanism Overview

The identification attacks described previously exploit the regularity of user mobility patterns. This is because, as already shown in table I device owners tend to visit a limited set of locations, producing repeatable spatiotemporal fingerprints that can be used to re-identify devices from short sequences of association events.

To mitigate this risk, we analyze the privacy benefits of a *stochastic association mechanism* that injects controlled randomness into the AP selection process. At each association event, the device first determines the *intended next AP* according to its default policy; however, instead of following this choice deterministically, it samples a *released next AP* from a candidate set of feasible alternatives. This approach departs from standard association policy, in which a device associates with the AP that maximizes a connectivity metric, such as received signal strength [20]. While this policy is effective at maximizing performance, repeatedly following this deterministic rule results in highly stable and distinctive mobility traces.

Figure 2 illustrates the studied privacy-preserving association mechanism and its impact on the observed mobility trace. A device d_1 naturally follows a sequence of AP associations (blue trajectory), determined by the default network selection policy. Under or proposed mechanism, instead of directly exposing this sequence, the mechanism introduces controlled randomness (tagged with a $?$) by selecting, at each step, an alternative AP from a set of feasible candidates (orange trajectory). These candidates correspond to nearby and contextually plausible associations, ensuring that the resulting trace remains realistic. As a result, the reported sequence no longer deterministically reflects the true mobility path but rather a perturbed trajectory that preserves local consistency. From the attacker's perspective, this altered trace induces ambiguity: while it remains structurally valid, it can correspond to multiple plausible underlying trajectories, thereby reducing the ability to reliably link or identify users.

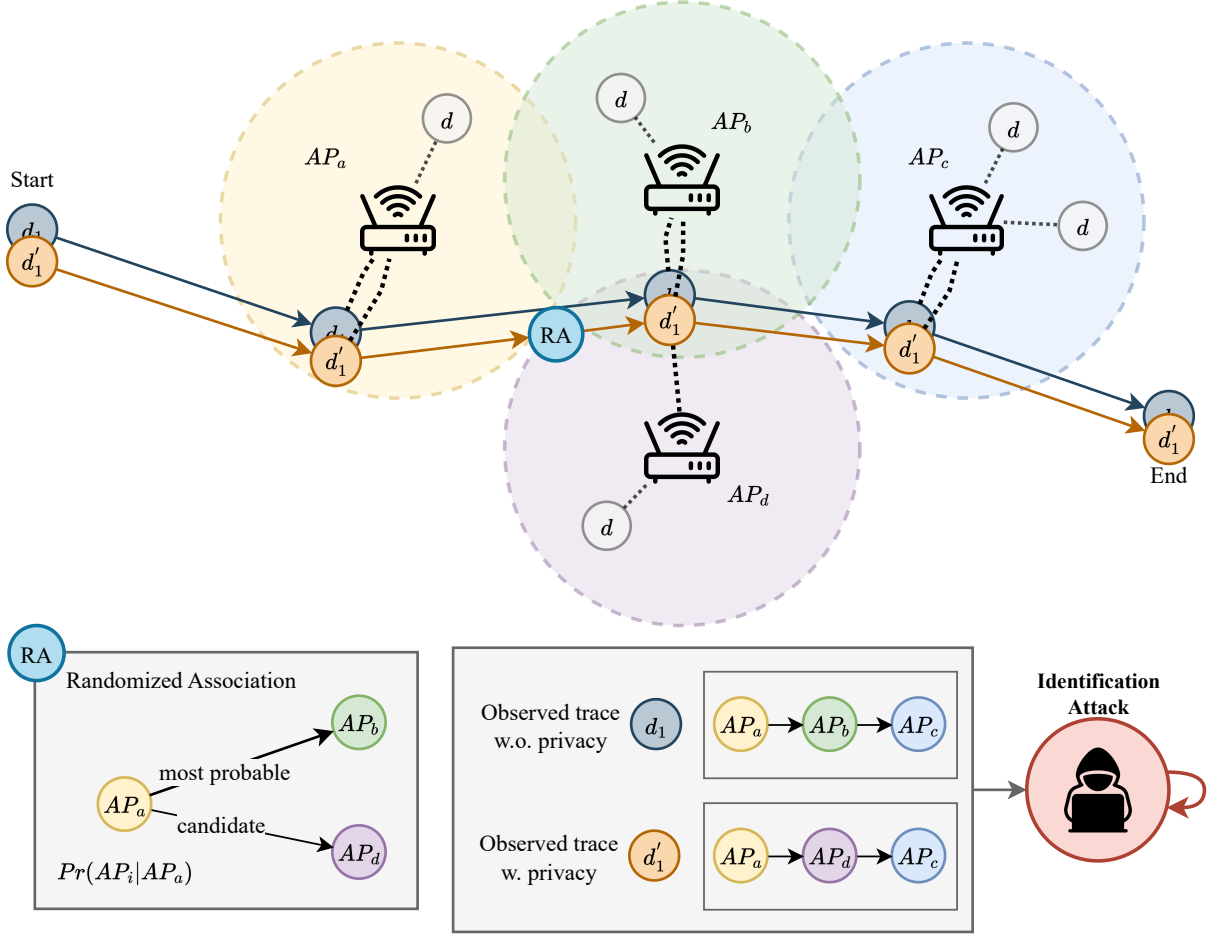


Fig. 2: Privacy-preserving association mechanism. The true AP sequence (blue) is transformed into a reported sequence (orange) by stochastically selecting an AP from the feasible candidate set $C(x)$ at each step. This preserves realistic mobility patterns while introducing ambiguity. The attacker observes the true trace without protection, but only an obfuscated trace when the mechanism is applied.

B. Stochastic Transition Model

Following the sequence representation described in equation (1), for a given AP x , the device exhibits a natural transition behavior captured by $P(y | x)$, which represents the likelihood of moving to a next AP y .

In the default operation, $b \in C(x)$ is the AP selected by the default policy as the next connection, while $y \in C(x)$ denotes the AP that is ultimately chosen after applying the mechanism. Here, $C(x)$ corresponds to the set of feasible next APs from the current AP x .

The Generalized randomized response is a standard mechanism in differential privacy. It outputs the true value with a probability controlled by the privacy parameter ϵ , and an alternative value otherwise. In our setting, this mechanism is applied over the candidate set $C(x)$, producing the following distribution:

$$\Pr[M_x(b) = y] = \begin{cases} \frac{e^\epsilon}{e^\epsilon + |C(x)| - 1} & \text{if } y = b, \\ \frac{1}{e^\epsilon + |C(x)| - 1} & \text{if } y \in C(x) \setminus \{b\}, \\ 0 & \text{if } y \notin C(x) \end{cases}$$

where

- b is the intended next AP,
- $C(x)$ is the set of candidate feasible next associations from AP x ,
- $|C(x)|$ denotes the cardinality of the candidate set $C(x)$,
- $\epsilon \geq 0$ is the privacy parameter controlling the level of randomization introduced by the mechanism.

In practice, the candidate set $C(x)$ is built from mobility statistics and includes transitions that satisfy the following criteria:

- **Minimum support:** transitions whose empirical probability exceeds a threshold τ_p . This filters out rare or noisy associations that are unlikely to reflect stable mobility behavior.
- **Temporal proximity:** transitions that occur within a temporal threshold τ_t . This ensures that only associations corresponding to realistic movement patterns are included, excluding transitions that are temporally implausible or induced by large AP coverage areas.
- **Reachability constraint:** we further restrict the candidate set to transitions that are directly accessible from x . Concretely, we consider the intersection

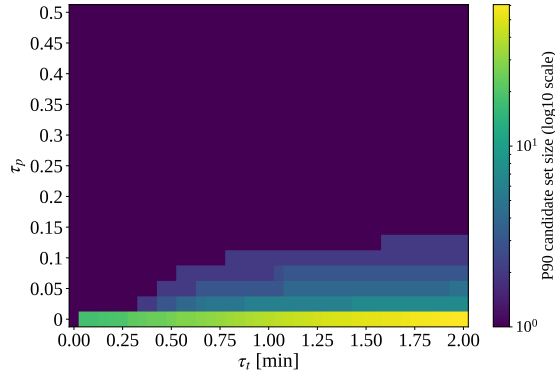


Fig. 3: Candidate set size as a function of the minimum support threshold τ_p and the temporal threshold τ_t (in minutes). The color scale is logarithmic.

between APs observable from x and those reachable from the intended AP b . This ensures that alternative associations remain physically and contextually plausible.

The parameters τ_t and τ_p jointly determine the structure and size of the candidate set. As illustrated in Fig. 3, smaller values of τ_p and larger values of τ_t lead to larger candidate sets, that represent noisy transitions. In practice, selecting τ_p in the range 0.025–0.05 provides a good balance, effectively removing low-probability transitions while stabilizing the number of candidates. For the temporal threshold, values of τ_t in the range of 0.6–2 minutes are appropriate, as they capture realistic movement between access points while excluding transitions that are unlikely to correspond to actual mobility.

Importantly, the applicability of the mechanism depends on the size of the candidate set $C(x)$. When $C(x)$ contains the intended AP b and at least one additional feasible AP, the device can randomize its association decision. In contrast, when $C(x) = b$, no alternative association is available and the mechanism becomes deterministic. To quantify how often this arises, we measure the fraction of traces and observed transitions for which the candidate set is singleton. In our dataset, although 38.7% of traces contain in this sense, they account for only 0.69% of observed transitions.

Under this formulation, the mechanism follows its original association decision with high probability when ϵ is large, concentrating probability mass on the intended AP b . Conversely, as $\epsilon \rightarrow 0$, it approaches a uniform distribution over $C(x)$, increasing the likelihood of selecting an alternative transition. This behavior is illustrated in Fig. 4: for $\epsilon = 0.5$, the distribution is nearly meaning that the probability of choosing another AP increases, whereas for $\epsilon = 4$, it becomes sharply concentrated on b . These examples highlight how, ϵ affects the trade-off between preserving natural mobility behavior and enhancing privacy of the resulting traces.

C. Differential Privacy Analysis

We analyze the privacy guarantees of the proposed stochastic association mechanism within the framework of differential privacy. The goal is to limit an adversary’s ability to infer the intended association decision from the observed sequence of transitions. Informally, differential

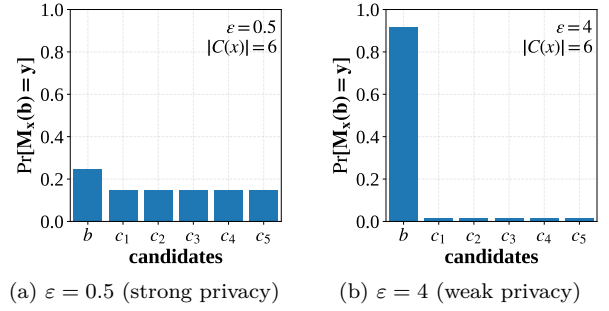


Fig. 4: Distribution induced by the stochastic association mechanism for different privacy levels for a candidate set of 6 APs. Lower ϵ produces a flatter, near-uniform distribution over the candidate set. Higher ϵ concentrates probability mass on the intended association b , preserving sequential structure.

privacy requires that the output of a mechanism remains statistically similar when its input changes [21].

a) *Local Differential Privacy*: A randomized mechanism M satisfies ϵ -local differential privacy if, for any pair of possible inputs b and b' , and any output y ,

$$\Pr[M(b) = y] \leq e^\epsilon \Pr[M(b') = y]. \quad (2)$$

Local differential privacy shifts the trust model from a centralized curator to the data source itself [22], requiring that each user perturbs their data before it is observed. This aligns naturally with our setting, where association decisions are generated at the device level and can be randomized locally before being exposed to the network.

In this context, the mechanism ensures that the probability of observing a connection to AP y is similar regardless of whether the intended association was b or b' . As a result, a passive observer cannot reliably infer the intended AP from the released association.

b) *Privacy guarantee*: The proposed mechanism induces a distribution over the candidate set $C(x)$ that favors the intended AP b while assigning non-zero probability to all alternatives. To quantify its privacy guarantees, we compare the outputs generated by two possible intended associations b and b' , both in $C(x)$.

The distinguishability between these two cases is captured by the likelihood ratio between the corresponding output distributions. This ratio is maximized in the worst-case scenario, when the released AP coincides with one of the intended associations (i.e., $y = b$):

$$\frac{\Pr[M_x(b) = b]}{\Pr[M_x(b') = b]} = \frac{\frac{e^\epsilon}{e^\epsilon + |C(x)| - 1}}{\frac{1}{e^\epsilon + |C(x)| - 1}} = e^\epsilon.$$

This matches the bound in (2), showing that the mechanism satisfies ϵ -local differential privacy.

c) *Sequence-level privacy*: The mechanism is applied independently at each association event in the mobility sequence. Consider a sequence of L associations,

$$S = (a_1, a_2, \dots, a_L).$$

Since each transition is privatized independently using the same mechanism, the overall privacy loss across a sequence follows from the sequential composition property of differential privacy [23]. If each association decision satisfies ϵ -local differential privacy, then the cumulative privacy loss for a sequence of length L is bounded by

$$\epsilon_{\text{seq}} \leq L\epsilon.$$

Overall, this mechanism provides a principled and tunable approach to obfuscating the underlying mobility pattern at the source, reducing the risk of re-identification while maintaining realistic and locally consistent association behavior.

VI. RESULTS

Now, we analyze the re-identification risk in WLAN association traces and evaluate the effectiveness of the privacy mechanism presented in previous section. We first establish a baseline by quantifying how easily devices can be identified without any protection. We then examine how the privacy parameter ϵ affects attack performance, and measure how perturbation impacts the fidelity of mobility traces.

A. Baseline linkability

First, we evaluate the effectiveness of the identification attacks in the absence of any private mechanism. This baseline scenario reflects the current operational setup of the network, where devices associate with APs according to their natural mobility patterns.

Table II shows the identification accuracy of the two attacks described earlier: the n -gram similarity attack and the Markov likelihood attack. We evaluate both methods using snippets of length $L \in \{20, 30, 50\}$ association events. As outlined in Section III, for each snippet the attacker ranks all candidate devices based on the likelihood or similarity that they generated the observed sequence. Performance is then measured using Top- k accuracy.

Attack	Top1 (%)	Top5 (%)	Top10 (%)
Snippet length $L = 20$			
n -gram	3.55	8.34	11.44
Markov	4.12	10.35	14.27
Random	0.01	0.03	0.06
Snippet length $L = 30$			
n -gram	4.58	10.41	13.99
Markov	5.23	13.48	18.33
Random	0.01	0.03	0.06
Snippet length $L = 50$			
n -gram	7.24	14.69	19.02
Markov	8.41	19.93	26.16
Random	0.01	0.03	0.06

TABLE II: Device re-identification accuracy without anonymization.

The results show that both attacks can re-identify devices from relatively short mobility snippets, with accuracy increasing as the snippet length grows. This trend is expected, as longer snippets provide more information about device mobility patterns. For instance, at $L = 30$,

the Top-10 accuracy reaches approximately 14% for the n -gram attack and 18% for the Markov attack, while Top-1 accuracy remains below 6%.

Although these Top-1 rates may appear low, they must be interpreted relative to the size of the candidate space. With $N = 16,801$ devices, random guessing yields only $1/N \approx 0.006\%$ Top-1 accuracy. In contrast, even the best observed Top-1 performance exceeds 8% (Markov, $L = 50$), representing a gain of several orders of magnitude over random identification. Similarly, Top-10 accuracy reaches up to 26%, compared to a random baseline of $10/N \approx 0.06\%$.

These results indicate that WLAN association traces encode distinctive behavioral patterns that enable effective device re-identification from short observation windows. Notably, the Markov attack consistently outperforms the n -gram approach, suggesting that first-order transition dynamics already capture a substantial fraction of the identifying information of the device.

B. Impact of the proposed mechanism

We now evaluate the impact of the privacy parameter ϵ on re-identification attacks. To this end, we focus on the Markov-based attack, as it consistently achieves the highest accuracy across all snippet lengths.

Figure 5 shows Top-1, Top-5, and Top-10 re-identification accuracy as a function of ϵ for $L \in \{20, 30, 50\}$. The dashed horizontal lines indicate the baseline (no protection), while the solid curves show the performance under the proposed mechanism for different values of ϵ .

A consistent trend is observed across all evaluations: accuracy decreases as ϵ decreases. This reflects stronger privacy, since lower ϵ increases the level of perturbation applied by the mechanism. In the limit $\epsilon \rightarrow 0$, the transition distribution approaches a uniform distribution over candidate APs, effectively removing distinguishable sequential patterns. As ϵ increases, the distribution becomes more concentrated around the intended transitions, restoring structure in the mobility traces and enabling higher re-identification accuracy.

Accuracy increases monotonically with ϵ , but the dominant factor is its interaction with snippet length. For small L , accuracy remains relatively limited even at higher ϵ . For larger L , it increases sharply (e.g., Top-10 $\approx 24\%$ for $L = 50$ at $\epsilon = 4$), indicating that longer sequences provide enough evidence to reliably distinguish users.

This shows that privacy risk scales with sequence length: even moderate ϵ can lead to significant leakage when observations are long enough. Lower ϵ mitigates this by flattening transition probabilities and limiting how much information accumulates across steps, whereas higher ϵ restores structure and enables effective matching.

C. Privacy-Utility tradeoff

The studied mechanism perturbs AP transitions at the client side, directly affecting the sequence of associations generated by the device. Accordingly, utility is defined in terms of how closely the resulting perturbed traces follow the intended mobility behavior, i.e., the sequence of associations that would arise under the default connectivity-driven policy. While the previous subsection characterized

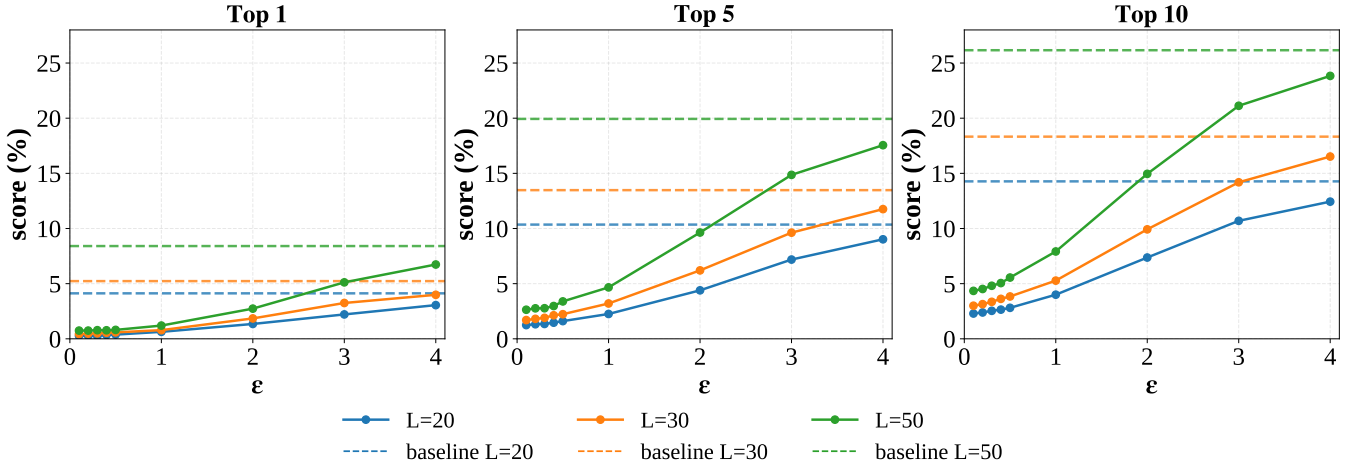


Fig. 5: Re-identification accuracy as a function of the privacy parameter ε .

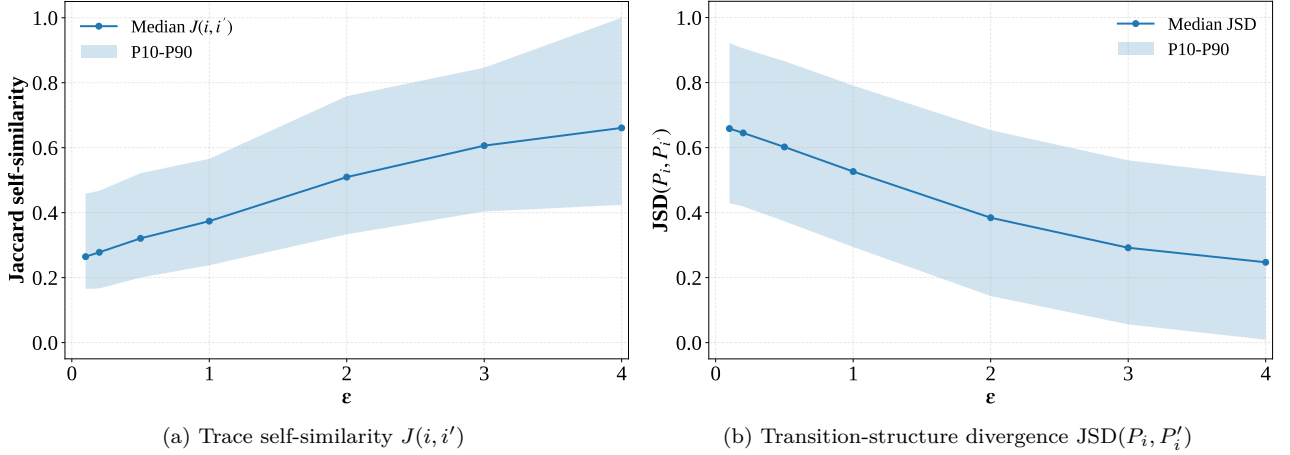


Fig. 6: Utility analysis of the proposed mechanism as a function of the perturbation parameter ε . a) self-similarity between original and perturbed traces. b) Jensen–Shannon divergence between the corresponding transition distributions.

the privacy implications of this perturbation, it is equally important to assess how much this intended behavior is preserved after randomization. To this end, we evaluate utility along two complementary dimensions: instance-level trace fidelity and preservation of transition dynamics.

Figure 6a shows the self-similarity $J(i, i')$ between each original trace and its perturbed counterpart as a function of the perturbation parameter ε . The similarity is computed using Jaccard Similarity over the set of STPs. We see that the similarity increases monotonically with ε , from values below 0.3 under strong perturbation to approximately 0.7 at $\varepsilon = 4$, indicating a progressive recovery of the original spatiotemporal structure.

Figure 6b complements this analysis by measuring the Jensen–Shannon divergence between the transition distributions of original and perturbed traces, $JSD(P_i, P'_i)$, a standard measure for comparing probability distributions and detecting changes in time series dynamics [24]. The JSD takes values close to zero when the two distributions are similar, and increases as they diverge, indicating larger differences in the underlying transition dynamics. Consistent with the previous results, the divergence decreases monotonically with ε , indicating that

the transition dynamics are increasingly preserved as perturbation weakens. Together, these results show that the mechanism enables a smooth and controllable adjustment of both trace-level fidelity and underlying mobility structure.

Taken together with the privacy results presented earlier, these findings highlight a clear privacy–utility trade-off governed by ε in the considered mechanism. Lower values provide stronger privacy by distorting both trace content and transition dynamics, leading to greater deviation from the intended connectivity-driven behavior. Conversely, higher values better preserve the original mobility structure and its underlying decision patterns, but increase the risk of re-identification. This highlights ε as an effective tuning parameter for balancing privacy and fidelity to the intended mobility behavior in downstream applications.

VII. CONCLUSION

In this work, we analyzed the privacy enhancements of a stochastic association mechanism that introduces controlled randomness into AP selection to mitigate mobility-based re-identification attacks in wireless networks. By perturbing transition patterns between APs,

such mechanisms reduce the distinctiveness of mobility traces while preserving realistic user behavior. This directly targets the information leakage embedded in association sequences, which remains even when explicit identifiers are removed.

Our evaluation on real mobility traces shows that increasing the perturbation parameter ε significantly degrades the effectiveness of sequence-based identification attacks, while maintaining plausible connectivity patterns. These results highlight the potential of shaping the association process itself as a complementary privacy mechanism, enabling a tunable privacy-utility trade-off without relying solely on identifier obfuscation.

ACKNOWLEDGMENT

This work is partially supported by the TUCAN6-CM project (TEC-2024/COM-460), funded by CM (ORDEN 5696/2024).

REFERENCES

- [1] J. Shen, J. Cao, and X. Liu, "Bag: Behavior-aware group detection in crowded urban spaces using wifi probes," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1669–1678. [Online]. Available: <https://doi.org/10.1145/3308558.3313590>
- [2] L. Huang, F. Xia, H. Chen, B. Hu, X. Zhou, C. Li, Y. Jin, and Y. Xu, "Reconstructing human activities via coupling mobile phone data with location-based social networks," *Travel Behaviour and Society*, vol. 33, p. 100606, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2214367X23000571>
- [3] F. Xu, Q. Wang, E. Moro *et al.*, "Using human mobility data to quantify experienced urban inequalities," *Nature Human Behaviour*, vol. 9, pp. 654–664, 2025. [Online]. Available: <https://doi.org/10.1038/s41562-024-02079-0>
- [4] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports*, vol. 3, p. 1376, 2013. [Online]. Available: <https://doi.org/10.1038/srep01376>
- [5] M. Cunha, R. Mendes, Y.-A. de Montjoye, and J. P. Vilela, "Wifi-based location tracking: A still open door on laptops," *IEEE Open Journal of the Computer Society*, 2025.
- [6] M. Cunche, M. A. Kaafar, and R. Boreli, "I know who you will meet this evening! linking wireless devices using wi-fi probe requests," in *2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2012, pp. 1–9.
- [7] J.-C. Zúñiga, C. J. Bernardos, and A. Andersdotter, "State of Affairs for Randomized and Changing Media Access Control (MAC) Addresses," RFC 9724, Mar. 2025. [Online]. Available: <https://www.rfc-editor.org/info/rfc9724>
- [8] J. M. Montes-Lopez, P. Serrano, M. Gramaglia, and A. P. Bianzino, "Xd-rcm: explainable deanonymization of 802.11 randomized and changing mac address," *2nd Winter School on AI for 6G Communications*, p. 23, 2026.
- [9] H. Hong, C. Luo, and M. C. Chan, "Socialprobe: Understanding social interaction through passive wifi monitoring," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MOBIQUITOUS 2016. New York, NY, USA: Association for Computing Machinery, 2016, p. 94–103. [Online]. Available: <https://doi.org/10.1145/2994374.2994387>
- [10] F. Xu, C. C. Tan, Q. Li, G. Yan, and J. Wu, "Designing a practical access point association protocol," in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1–9.
- [11] W. Wu, Y. Liu, J. Yao, X. Fang, F. Shan, M. Yang, Z. Ling, and J. Luo, "Learning-aided client association control for high-density wlangs," *Computer Networks*, vol. 212, p. 109043, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S138912862200192X>
- [12] A. Rubens, C. Rigney, S. Willens, and W. A. Simpson, "Remote Authentication Dial In User Service (RADIUS)," RFC 2865, Jun. 2000. [Online]. Available: <https://www.rfc-editor.org/info/rfc2865>
- [13] I. Palamà, A. Amici, G. Bellicini, F. Gringoli, F. Pedretti, and G. Bianchi, "Attacks and vulnerabilities of wi-fi enterprise networks: User security awareness assessment through credential stealing attack experiments," *Computer Communications*, vol. 212, pp. 129–140, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S014036642300347X>
- [14] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.
- [15] Y. Li, J. Barthélemy, S. Sun, P. Perez, and B. Moran, "A case study of wifi sniffing performance evaluation," *IEEE Access*, vol. 8, pp. 129 224–129 235, 2020.
- [16] Android Open Source Project, "Mac randomization behavior," <https://source.android.com/docs/core/connect/wifi-mac-randomization-behavior>, 2024, accessed: 2026-03-27.
- [17] X. Lu, E. Wetter, N. Bharti *et al.*, "Approaching the limit of predictability in human mobility," *Scientific Reports*, vol. 3, p. 2923, 2013. [Online]. Available: <https://doi.org/10.1038/srep02923>
- [18] H. Shi, Y. Li, H. Cao, X. Zhou, C. Zhang, and V. Kostakos, "Semantics-aware hidden markov model for human mobility," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 3, pp. 1183–1194, 2021.
- [19] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1116–1124.
- [20] B. Gómez, E. Coronado, J. Villalón, R. Riggio, and A. Garrido, "User association in software-defined wi-fi networks for enhanced resource allocation," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–7.
- [21] F. Fioretto, P. Van Hentenryck, and J. Ziani, "Differential privacy overview and fundamental techniques," *arXiv preprint arXiv:2411.04710*, 2024.
- [22] B. Bebersee, "Local differential privacy: a tutorial," *arXiv preprint arXiv:1907.11908*, 2019.
- [23] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [24] D. M. Mateos, L. E. Riveaud, and P. W. Lambert, "Detecting dynamical changes in time series by using the jensen shannon divergence," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 8, 2017.