Precision Leads Recalling You! Improved Location Privacy for Shared Mobility Services

Debasree Das
University of Bamberg
Bamberg, Germany
debasree.das@uni-bamberg.de

Daniela Nicklas
University of Bamberg
Bamberg, Germany
daniela.nicklas@uni-bamberg.de

Abstract

The rapid growth of shared micromobility, such as e-scooters and e-bikes, has transformed urban transportation, bridging the gap between public transit and first/last-mile mobility. As users and cities need information on the status and usage of such micromobility vehicles, operators publish the data using General Bikeshare Feed Specification (GBFS)-compliant APIs. These feeds are extremely useful for operational transparency and enabling third-party integration into navigation apps. However, it has raised significant privacy concerns, particularly around the fine-grained spatiotemporal data sharing, which can reveal potentially sensitive information about travel patterns, even without explicit personal identifiers. For instance, by leveraging high-precision GPS coordinates, battery levels, and timestamps, malicious actors can infer trip origins and destinations, posing a risk of membership inference attacks. Despite efforts to mitigate such risks through dynamic vehicle IDs and GBFS guidelines, the potential for privacy leakage remains.

In this paper, we investigate these privacy risks in the context of micromobility data, addressing four key research questions: (1) identifying vulnerable fields in GBFS data that can leak trip trajectories; (2) validating trip origin-destination inference attacks without access to ground truth data from operators; (3) assessing the generalizability of such attacks across different cities, operators and GBFS version; and (4) proposing effective mitigation strategies. We propose a heuristic method for reconstructing trip origins and destinations using only publicly available GBFS data, without relying on vehicle identifiers or auxiliary quasi-identifiers. Our empirical analysis, conducted on data from two cities with varying sizes, shows that a significant proportion of trips can be accurately *recalled*, with over 80% of trip source and destination pairs identified across both cities. Furthermore, our proposed anonymization techniques, such as data generalization and removal of quasi-identifiers, can prevent up to 97% of successful attacks, ensuring privacy without sacrificing data utility.

Keywords

micromobility, location privacy, origin-destination inference attack, mitigation

Proceedings on Privacy Enhancing Technologies 2026(4), 1041–1054

https://doi.org/10.56553/popets-2026-0158

1 Introduction

The growing human population worldwide has led to a surge in private vehicle ownership, further exacerbating traffic congestion, poor air quality, higher noise levels [1], social inequity [20], to name a few global concerns. Despite the adversarial effects, people often prefer private vehicles over public transport or shared mobility due to various reasons, such as inadequate public transit systems, insufficient infrastructure, delays, and the limited availability of shared mobility options. Amidst such challenging situation, shared "micromobility" appears as a global boon to boost the preference towards public transportation as it often serves the first/last mile gap of public transportation. The term "micromobility" refers to bicycles, e-bikes, kick scooters, or more recently, the *e-scooters* ridden by individual users. It often comes in either dockless or station-based mode. As this lightweight mode of transportation is gaining growing popularity and becoming the focal point of urban mobility, various automobile companies, such as Lime, Dott, Pony, ZEUS, and many more, are promoting dockless infrastructure where vehicles can be accessed using a mobile app and parked within a guided perimeter upon completion of the journey. To enable interoperability among operators, cities, and trip-planning applications, there should be a common way of sharing information on the availability of such micromobility data. Primarily, the operators provide an API, as depicted in Fig. 1, to a city or other third-party stakeholders (e.g., Deutsche Bahn, a national railway company in Germany) to access the availability of micromobility, which they can include in their navigation app to recommend a transport mode. Additionally, these shared mobility operators offer a mobile application that displays location-wise available micromobility options for travelers. However, as any new technology which evolves quickly due to enhancing environmental sustainability, more transit opportunities, etc., also comes up with challenges such as privacy leakage in sharing information [4, 28], security threats due to denial of service attack, spoofing [15], tracking users [5], usage of improper protocols in communicating the data [3].

To alleviate such concerns, a de facto regulatory standard called **Mobility Data Specification** (MDS)[1] states that information such as vehicles that are not in use, their type, location, etc., can be shared. Aligning with this, the General Bikeshare Feed Specification (GBFS)[2] is a common data standard used by micromobility data operators to share their data, focusing on availability, location, and operational status. Note that, using the GBFS compliant data sharing API, people can view the current free vehicles and their status

[1] https://raw.githubusercontent.com/openmobilityfoundation/governance/main/documents/OMF-MDS-Privacy-Guide-for-Cities.pdf (Accessed Online: November 23, 2025)
[2] https://github.com/MobilityData/gbfs (Accessed Online: November 23, 2025)

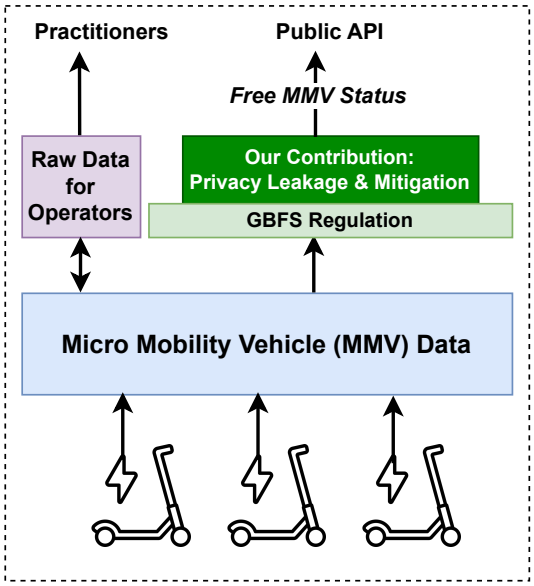

Figure 1: A Glimpse of the Data Sharing Pipeline

in real-time, but not access historical data. This open data ecosystem powers navigation apps, facilitates research on mobility equity and fleet distribution, and enhances public accountability. However, more transparency brings more privacy concerns; although the feeds contain no explicit personal identifiers, they continuously broadcast fine-grained spatiotemporal data about shared vehicles, often updated every few minutes. Using such repeated feed of information, it is possible to pose a membership inference attack [3, 9]. For instance, when a vehicle disappears from the feed (signaling that it has been rented) and later reappears at a new location, the implicit trajectory can reveal the origin and destination of an individual trip. Repeating this process over time enables reconstruction of detailed travel patterns, potentially exposing where users live, work, or frequently visit. To mitigate this, GBFS assigns dynamic IDs to vehicles, which must be changed to a random string after each trip, and information about vehicles in use should not be publicly disclosed. Yet such privacy regulations lack empirical evaluation or any formal privacy guarantees. Operators follow these guidelines and apply such policies with varying degrees of rigor, but still suffer from privacy leakage issues. For example, Elzer et al. [9] show they can infer more than 80% of trips using static quasi-identifiers like license plate, unique vehicle ID, timestamps, etc., from DOTT[3] (previously known as Tier), a leading rental micromobility operator functioning in Europe and Middle East. Moreover, they found that the data policy is GBFS non-compliant, as the IDs remain static over time. Similarly, Xu et al. [29] propose to infer trip origin and destinations when the IDs are static and dynamic. Baltra et al. [4] strongly claim that existing GBFS standards are not enough to ensure formal privacy guarantees, which puts individual riders' information at risk after analyzing GBFS-compliant data for Los Angeles. However, existing works focus on either static IDs or a combination of ID and quasi-identifiers (such as license plates), concentrating on a single city. However, current GBFS regulations, starting from version 1.0 and higher, already mitigate such leakage attacks by using dynamic IDs and not releasing other quasi-identifiers, such as license plate numbers, time upon state change, and, of course, user and payment information. Still, a set of fields with high precision in the feed remains dormant yet vulnerable, providing an opportunity to *recall* trips.

Challenges: GBFS-compliant feeds publish primary information on vehicles such as a dynamic vehicle ID, reservation status, operational status (disabled or not), GPS coordinates up to 6 decimal places, last reported timestamp in seconds, battery level expressed as remaining range in meters, and the vehicle type. Among these, the high-precision coordinates (up to 0.11m), battery level, and timestamps in seconds are particularly vulnerable: when combined intelligently across repeated API queries, they act as unique identifiers and can reveal trip segments without relying on the official vehicle IDs. Consequently, we ask the first **Research Question (RQ1):** *What is the minimal subset of fields and how much information can they leak?* To illustrate the risk, consider an attacker who knows from social media that a particular individual is likely to attend a morning event. Even without any direct personal data in the GBFS feed, the attacker can repeatedly query the API, link fine-grained coordinates with battery-range changes, and reconstruct whether a specific scooter completed a plausible route from the individual's residence to the event location. In this way, external auxiliary knowledge combined with vulnerable GBFS fields enables *membership inference*, revealing whether someone is likely to have participated in an activity or visited a location. A second challenge lies in validating such inference attacks. Prior works either rely on operator-provided ground truth [4], or collect their own datasets using quasi-identifiers [9, 29]. However, obtaining such ground truth at scale is impractical for large metropolitan cities with thousands of micromobility vehicles. Thus, we ask **Second Research Question (RQ2):** *How to construct reliable ground truth using only GBFS-compliant data?* Furthermore, inference pipelines designed for one city often lack generalizability, as they exploit city-specific patterns or implementation quirks of one GBFS version. Here, we ask **Third Research Question (RQ3):** *Whether an inference attack achieves location neutrality by remaining effective across locations and operators?* Overall, these challenges raise a central concern for cities and operators, forming as **Fourth Research Question (RQ4):** *How can shared mobility data remain open and helpful without endangering rider privacy?* Addressing these questions motivates the need to design and evaluate effective attack mitigation and prevention techniques.

Our Contributions: In light of these issues, we propose a comprehensive heuristic method for micromobility vehicles, in short **MMV**, that identifies a set of vulnerable fields from GBFS-based shared mobility data and infers trip origin and destination without relying on vehicle identifiers. We propose a set of preventive anonymization methods that can be applied to the existing GBFS standard, ensuring data utility remains while blocking privacy leakages. To validate our approach and demonstrate location invariance, we conduct the study in two cities of varying sizes: Brussels, the large metropolitan capital of Belgium with a population of 1.25

[3]https://ridedott.com/ (Accessed Online: November 23, 2025)

million, and Bamberg, a mid-scale German university town with a population of 80k. We anonymize the second city name and its associated location names throughout the paper to adhere to the anonymization policy. In summary, we present our work, *Precision Leads Recalling You*, which, to the best of our knowledge, identifies trip trajectories without using MMV identifiers and auxiliary quasi-identifiers, and validates the approach across cities, ensuring generalizability. Fig 1 depicts our contributions, which build upon existing GBFS regulations, as follows.

(1) Data Collection and Empirical Analysis: We utilize an API provided by the Brussels Mobility Twin project [24] to collect long-term data for MMVs. The data collection period lasted for a month and consisted of one month of data for the whole of September $1-30$, 2024 in a regular interval of 6 minutes, comprising 99 million entries. For Bamberg, we collected data for 7 days, April $1-7$, 2025, using an API provided by ZEUS[4], a project partner and the only MMV operator in the city. The dataset was curated every 10 minutes and comprises 371k entries. Following this, we identify the gaps in existing research works (§2) and GBFS policy (refer to Table 2). Consequently, answering **RQ1**, we identified a subset of fields as {is_reserved, is_disabled, current_range_meters, vehicle_type, last_reported, coordinates} through which the plausible trip inference can be done.

(2) Reconstruction Attack and Its Validation: We identified the different status of MMVs as startoff, drop-off, disabled, residential, and passby by using a heuristic approach using the above fields. Later, we propose a spatio-temporal correlation Algorithm 1 to infer trip origins and destinations and subsequently construct plausible trajectories using OpenStreetMaps [19]. While answering **RQ2**, we find 12% of the MMV entries do not change their vehicle ID and remain static throughout the trip for Brussels. Therefore, this satisfies our ground truth requirement and only serves the purpose of our validation.

(3) Evaluation: We identified 82.91% unique MMV trip source and destination and their corresponding trajectories for Brussels, and for Bamberg it is 80%. We also identified the temporal pattern of trips for specific zones in the two cities, and the potential membership inference with access to a secondary data source, such as a known target group, is discussed in detail in §4.4.3. The evaluation across two cities and different GBFS versions further confirms the generalizability of our approach as posed in **RQ3**.

(4) Ensuring Mitigation: Finally, we propose a set of anonymization approaches, including generalization, removal of quasi-identifiers, and reporting on the required precision to achieve sufficient utility, building on existing work in §5. Our approach shows that reconstruction attacks can be prevented by 97%, answering **RQ4**.

The remainder of the paper is structured as follows: section 2 provides a detailed literature review relevant to our study, including two attack models. In section 3, we introduce the public datasets which have been used in this paper throughout, and state the formal problem definition. Following this, we propose our attack model, focusing on inferring origin-destination pair and plausible trip trajectories in section 4, and evaluate the approach in section 4.4. Next, we discuss potential mitigations and their evaluation, focusing on the privacy-utility tradeoff in section 5. Finally, we conclude our work by stating the major goals achieved, clearly mentioning limitations, and discussing possible future work in this direction in section 6. We also present the ethical consideration of the study in section 7.

[4]https://zeusscooters.com/ (Accessed Online: November 23, 2025)

2 Related Work and Background

Ensuring privacy in mobility data is a well-researched area, as location data is inherently sensitive. Mobility data varies across different transport modes, such as walking, cycling, biking, driving, or public transport. Therefore, the diverse data generated from each of these modes has a different amount of privacy risks and subsequent mitigation techniques [2]. As our work focuses on the privacy risks associated with GBFS data from micromobility vehicles (MMVs), it is crucial to conduct a targeted investigation of the privacy implications specific to GBFS data from MMVs. Consequently, we concentrate on three key subareas: (1) inference attacks on mobility and GBFS-like data, (2) privacy analyses of micromobility applications, and (3) anonymization and privacy-preserving mechanisms for location data in this rapidly evolving domain as follows.

2.1 Inference Attacks on Mobility and GBFS Data

As emerging mobility services, such as the first- and last-mile ride options provided by shared bike platforms, are one of the essential pillars for transforming a city's transportation network, cities want them to be safe, accessible, and efficient. To help in exchanging the data between the city and mobility operators (e.g., e-scooters, bicycles, mopeds, car share, taxis, rideshare, personal delivery devices), **Mobility Data Specification** (MDS)[1] provides a standard on sharing information such as type of vehicle, operation status, location, etc. But location data is marked as "personal data" according to General Data Protection Regulation (GDPR) [22] of the European Union (GDPR Art. 4) as using a trail of locations over a period of time, private and sensitive information [23, 28, 30] such as individual's stay points (e.g., home location), way points (office, ATM, etc.,), religious practices, medical information, and many more can be revealed. In the MDS specification itself, it directly reports that, "Although MDS does not contain any specific information about who uses a shared vehicle, data on how vehicles and devices move through space over time, such as *MDS trip data, can potentially be linked with other datasets to identify people*. Notably, the main purpose of sharing such data is to enrich the services provided by the transportation services, such as predicting the availability vs. usability of bikes in a city-center area. Montjoye et al. [7] show 95% of human traces are possible to infer using the unicity test by using just 4 spatiotemporal points. Freudiger et al. [12] inferred home and work locations of individuals using GSM, WiFi, and GPS data. More closely related to our setting, using GBFS data specific to micromobility, recent works [4, 9, 29] analyze privacy risks in micromobility and GBFS-like datasets. These studies demonstrate that it is possible to infer trips or partial trajectories of vehicles. However, they typically rely on stronger linkage signals, such as static or persistent vehicle identifiers, or infer destinations and subsequent origins independently when identifiers reset. In contrast,

our work considers a more constrained and realistic setting with frequently rotating identifiers and shows how origin–destination (OD) pairs can be inferred using only GBFS-compliant attributes without relying on identifiers.

2.2 System-Level Privacy and Security Risks in Micromobility

Beyond inference attacks on published mobility data, prior work has examined privacy and security risks in micromobility systems at the application and infrastructure level. For instance, Vinayaga-Sureshkanth et al. [27] analyze data collection and sharing practices in mobile e-scooter rental applications, highlighting risks related to third-party tracking, user profiling, and policy-level privacy violations. Other works focus on system-level threats such as denial-of-service attacks [15], user tracking through application vulnerabilities [5], and improper use of communication protocols in ride-sharing platforms [3]. These studies emphasize risks arising from application design, network communication, and system misuse. In contrast, our work focuses on a different but complementary attack surface: inference over continuously published, publicly accessible GBFS feeds. Unlike app-level analyses that rely on user-device interactions or privileged data access, we demonstrate that sensitive mobility patterns can be inferred solely from open data streams using algorithmic linkage. Despite the breadth of prior work, a systematic study of trajectory or origin–destination inference risks in GBFS data under realistic deployment constraints remains limited. In particular, existing efforts do not fully address how exposed attributes—such as spatial coordinates, timestamps, or battery-related information—can be jointly exploited when identifiers are dynamic or unavailable. This gap highlights the need for privacy-enhancing mechanisms that are both context-sensitive and utility-preserving for real-time micromobility data.

2.3 Privacy-Preserving Mechanisms for Mobility Data

A large body of research has proposed techniques to mitigate privacy risks in location data. Also, several works [11, 16] have already proposed measures for protecting and publishing the location data more privately, by removing the user details according to MDS standard, removing ID, randomizing ID over a certain period of time [21], spatial cloaking and location generalization [17], as well as formal approaches such as k-anonymity [25], l-diversity [18], and differential privacy [8]. While these methods reduce re-identification risks, they introduce inherent privacy-utility trade-offs. For example, coarse spatial granularity or heavy noise injection can significantly degrade the usability of mobility data for applications such as vehicle availability prediction or demand estimation. Recent works explore context-aware and utility-preserving mechanisms [6, 31], but these are not specifically tailored to real-time GBFS feeds.

2.4 Positioning and Limitations of Existing Work

Table 1 presents the existing literature that focuses on ensuring privacy for MMV data. We present a feature-wise comparison of what the existing study already covers and where our contributions stand in relation to it. Existing studies largely fall into two categories: (i) general anonymization techniques for mobility datasets, and (ii) inference-based analyses for micromobility systems. Generic anonymization approaches (e.g., identifier removal, spatial cloaking, differential privacy) are not tailored to micromobility settings, where sparse vehicle distributions can still enable linkage and inference. Moreover, stronger privacy protections often reduce data utility, affecting applications such as availability prediction and fleet management. More closely related works [4, 9] quantify reconstruction risks and propose mitigation strategies. However, they typically rely on stronger linkage signals, such as static identifiers or additional auxiliary attributes (e.g., speed limits or detailed state-change timestamps). Even with dynamic identifiers, these assumptions may not hold in GBFS-compliant deployments. These limitations highlight a gap in understanding privacy risks under realistic constraints, where identifiers are frequently rotated and only standard GBFS fields are available.

Our work addresses this gap by inferring origin–destination (OD) pairs using only GBFS-compliant attributes under dynamic identifiers. We further derive plausible, map-constrained routes to illustrate feasible movement, without claiming recovery of true trajectories. Finally, we propose lightweight mitigation strategies that balance privacy protection with operational utility for real-world GBFS deployments.

Table 1: Comparison of Existing Works Based on Mobility Data: Covers (✓), Does Not Cover (×), Partially Covers (~)

Feature/Criteria	[9]	[29]	[4]	[28]	Our Work
GBFS Compliant	✓	✓	✓	×	✓
Infers Origin-Destination	✓	✓	✓	✓	✓
Infers Full Trajectory	×	×	×	✓	~
Vehicle ID Static	✓	✓	✓	×	×
Vehicle ID Dynamic	~	~	×	×	✓
Uses Vehicle ID or Auxiliary Info	✓	✓	✓	×	×
Uses Battery Level	✓	×	×	×	✓
Uses Last State Change Timestamp	✓	×	×	×	×
Evaluated on Multiple Cities	×	×	×	✓	✓
Preventive Mechanisms Suggestions	✓	×	✓	×	✓

2.5 Related Attacks

We discuss two attacks, relevant to our contribution as inference and re-identification attacks, as they represent common privacy risks in the context of mobility data. Their focal point is to infer sensitive information about individuals, either by linking mobility data with external knowledge or by reidentifying individuals based on shared

data patterns. Understanding them is necessary to contribute to our reconstruction attack without relying on direct identifiers in the later stage.

2.5.1 Inference Attack. In this attack, the adversary attempts to infer a user's unknown location within a trajectory, assuming that all other locations in that trajectory are already known. The adversary leverages the sanitized dataset to exploit preserved correlations among locations. Specifically, for a given target trajectory, the adversary constructs a query containing all known locations and retrieves all sanitized trajectories that include these locations. From the retrieved trajectories, the adversary identifies additional locations that frequently co-occur with the known set but are not part of it. The most frequently observed location among these candidates is inferred as the user's hidden location. An attack instance is considered successful if the inferred location corresponds to the user's actual hidden location. This process captures whether the sanitization mechanism has effectively disrupted the co-occurrence relationships between locations within trajectories. If such correlations remain intact, the adversary can still reconstruct missing locations with high confidence. Consequently, users with few, recurrently visited locations that also appear in other users' trajectories are more susceptible to this attack.

2.5.2 Re-identification Attack. In a re-identification attack, the adversary seeks to link a sanitized trajectory to a known user by exploiting a small set of location observations. Starting from a target user's trajectory, the attacker selects a subset of known locations and queries the sanitized database for trajectories that contain those locations. If a query returns exactly one sanitized trajectory and that trajectory corresponds to the target user, the query is deemed successful. By varying the size and composition of the known subset, the attacker can determine the minimum number of locations required to uniquely re-identify the user; this per-user minimum quantifies how easily an individual can be re-identified from the sanitized collection.

This attack measures dataset uniqueness: users whose movement patterns are distinctive relative to the population (even after sanitization) are more vulnerable. Re-identification differs from the inference attack in that it attempts to recover identity rather than a missing location; consequently, users who visit few locations that are also common across others may be less at risk of re-identification. The effectiveness of the attack depends on the distortion introduced by the privacy mechanism; non-randomized sanitizers can enable stronger attacks because adversaries may apply the same transformation to their queries.

3 Dataset and Problem Definition

According to the GBFS specification[5], the vendors need to publish the data in JSON format and to make it accessible by the common people a restful API is recommended. The widely adopted specifications for shared MMVs (bikes, e-scooters, etc.) embed certain anonymization and generalization practices that help guard user privacy, while still enabling real-time data for trip planning and system monitoring. The Key features include the mandatory rotation of vehicle identifiers after each rental, the exclusion of vehicles currently in use from the public feed, and explicit recognition that the feed does not contain personal data or full trip histories, as summarized in Table 2.

These features support several privacy-protection goals: (i) they reduce the linkage of a specific vehicle (and by extension potentially a user) across time and locations; they prevent public exposure of ongoing trips (which could reveal a user's start and end points); and they encourage data minimization by warning against overcollection. However, despite these protections, there are limitations and trade-offs. For example, if vehicle locations are published very frequently and/or once a trip ends, the vehicle appears again almost immediately (with a new ID) in the same approximate area, one might still infer the trip origin and destination via disappearance/reappearance algorithms. Additionally, the precision of location data (six decimal places) may or may not be suitable from a privacy perspective in all contexts, and some operators/publishers may opt for coarser precision or further anonymization. In summary, GBFS provides a solid foundation for generalizing privacy-aware data publication for MMVs; however, depending on the exact use case (especially for research that infers trips or links vehicles across time), additional anonymization steps may be warranted.

As we concentrate specifically on the MMV data, we give a brief description of two datasets acquired from two different European cities, large metropolitan capital city Brussels, Belgium, and a mid-scale city Bamberg, Germany which is also a university town. For the city of Brussels, due to an overwhelming surge in the micromobility vehicles, Sakr et al. [24] have created a restful API[6] to better understand the overall distribution of the fleet at different times of the day under the project Brussels Mobility Twin (BMT). The data can be pulled every 6 minutes with any number of requests for a given API key. They use the GBFS specification 1.0, where the main focus of releasing the data is to check or count the number of free bikes as "free_bike_status". We are particularly interested in the attributes of "free_bike_status" field which gives us the following information listed in Table 3.

According to GBFS, the location data should not be shared when the bikes are already in use, or the free status is false, but from the acquired dataset, we get a 1% of data where location data is still available from "in use" bikes. For Bamberg, we utilize an API acquired for city research on transportation, provided by ZEUS[4], the only operator in this city and it uses GBFS version 2.2. From this acquired dataset, we get no MMV mobility data while in use. We assume that either the latest specification uses a tighter approach towards privacy compared to the lower version, i.e., 1.0, or a leaky API might be the reason.

The data feeds are refreshed based on their time-to-live (TTL) values. As the GBFS specification does not mandate a fixed TTL, the interval can differ across providers. Common TTL values include 4, 300, and 600 seconds. If a feed has a TTL of less than 60 seconds, we can collect data at one-minute intervals. For feeds with higher TTLs, we need to align our scraping frequency with the TTL value itself. However, as both the APIs publish information on the available bikes, the TTL is always 0 in our case. Therefore, for Brussels our

[5]https://gbfs.org/documentation/reference/ (Accessed Online: November 23, 2025)

[6]https://data.lime.bike/api/partners/v2/gbfs/brussels/gbfs.json (for instance, LIME bike and Accessed Online: November 23, 2025)

Table 2: Key Practices in GBFS

Practice	Required?	Description	Source	Privacy Rationale
Rotate vehicle IDs after each trip	Yes	The specification states that the field vehicle_id "MUST NOT be persistent for privacy reasons" and that it "MUST be rotated to a random string after each trip".	v3.0 reference for vehicle_status.json (vehicle_id)	Prevents linking a vehicle's location before and after rental across time, which reduces risk of reconstructing trip origin/destination patterns for a given user (especially in dockless systems)
Do not publish vehicles while in active rental	Conditionally Required	For the free-floating (dockless) vehicle feed (formerly free_bike_status.json, now vehicle_status.json), GBFS states: "Vehicles that are part of an active rental MUST NOT appear in this feed."	v3.0 reference, and policy guide	If vehicles in use (with a rider) were published, one could infer the user's whereabouts and journey endpoints – omitting active rentals reduces that risk.
Limit coordinate precision guidance	Conditionally Required	The guidance on coordinate precision, e.g., 6 decimal places are supported allowing up to 0.11 m accuracy; there is a recognition that publishing exact locations has privacy implications.	While not strictly "anonymization", this helps thinking about granularity.	Reducing precision or choosing not to publish extremely fine-grained location helps avoid identifying home/office locations of users.
State that feed does not contain personal data	Yes	In the FAQ the spec acknowledges: "GBFS does not contain personal data."	FAQ page.	Explicitly clarifies the intent of the specification: it is designed to serve users/trip planners rather than for tracing individuals/trips.
Operators/publishers responsibility: collect only required data	Conditionally Required	The Data Policy guidance warns: "Extreme care should be taken when requiring data from operators that is outside the scope of the specification."	Data Policy guide.	Encourages limiting data exposure to only what is necessary, reducing privacy risk by design (data minimization).

Table 3: Important Attributes for free_bike_status

Attribute	Description	Type
bike_id	randomized bike id	string
is_reserved	is the bike currently reserved for someone else	boolean
is_disabled	is the bike currently broken or damaged	boolean
current_range_meters	how much the bike can travel in meters	integer
vehicle_type_id	a unique number assigned to type of vehicle, e.g., 2 for scooter	string
last_reported	Timestamp in seconds since epoch. If not specified, the latest value will be returned. Otherwise, the closest value to the specified timestamp will be returned	integer
coordinates	latitude and longitude reported for the current bike	number
vehicle_type	whether it is a scooter or bike	string

scraping frequency is every 6 minutes and for Bamberg, it is every 10 minutes.

Problem Definition Given a shared bike system database $\mathcal{D}$ with randomized bike ids, say i for an individual bike and stationary information $\mathcal{Z}$, identify the vulnerable fields $\mathcal{Z}' \subset \mathcal{Z}$ through which a plausible trajectory $\mathcal{T}_i$ corresponding to bike i can be derived through inferring the origin-destination pair (u_i, v_i). We identify a possible set of anonymization methods, denoted as $\mathcal{M}$, that can prevent such attacks. Furthermore, we evaluate $\mathcal{M}$ to determine its extent in preventing such attacks.

4 Origin-Destination Inference Attack

We use the GBFS feed of MMVs, which is updated every Δt units of time. Although individual MMVs cannot be directly tracked due to dynamic identifiers, the exposed attributes enable indirect linkage across snapshots. Intuitively, when an MMV disappears from the feed (e.g., due to being rented) and later reappears, an adversary can attempt to associate these events to infer its trip. Our approach leverages such disappearance and reappearance patterns, combined with multi-attribute consistency checks, to infer origin-destination (OD) pairs of MMVs. Importantly, we do not directly observe or reconstruct the full trajectory taken by a vehicle. Instead, we derive a plausible route between the inferred origin and destination using map constraints (e.g., the underlying bike network). Thus, the primary privacy risk arises from accurately inferring OD pairs, while the route reconstruction step serves only to illustrate a feasible path rather than the exact trajectory. Importantly, we do not rely on any single attribute as a unique identifier. Instead, inference is achieved through the joint consistency of spatial (coordinates), temporal (`last_reported`), and battery-related (`current_range_meters`) signals. To achieve this, we first introduce our threat model, followed by identifying MMV status transitions, and then inferring OD pairs using Algorithm 1. Finally, we evaluate the validity of our approach (see Section 4.4) in terms of accuracy and robustness across two cities.

4.1 Threat Model

We consider a passive, external adversary interacting with a GBFS-compliant system. The adversary does not compromise the system or modify its behavior, but can repeatedly query publicly accessible APIs and collect data over time. Through these queries, the adversary observes all exposed fields, including spatial coordinates, timestamps (e.g., `last_reported`), `current_range_meters`, and MMV state indicators (e.g., `is_reserved`, `is_disabled`). We assume that the adversary has moderate auxiliary knowledge, such as typical MMV speeds, operational constraints, and general characteristics of urban mobility, but does not have access to persistent identifiers or privileged internal data. An attacker assumes that an MMV can travel at an average speed of 10 $\mathrm{km\,h^{-1}}$, given its maximum speed and minimum moving speed as 25 $\mathrm{km\,h^{-1}}$ and 5 $\mathrm{km\,h^{-1}}$, respectively. In particular, MMV identifiers are dynamic or reset across queries, preventing direct linkage. The adversary's objective is to reconstruct origin-destination pairs and derive plausible trajectories of individual MMVs by linking observations across successive API snapshots. This is achieved by exploiting consistency across

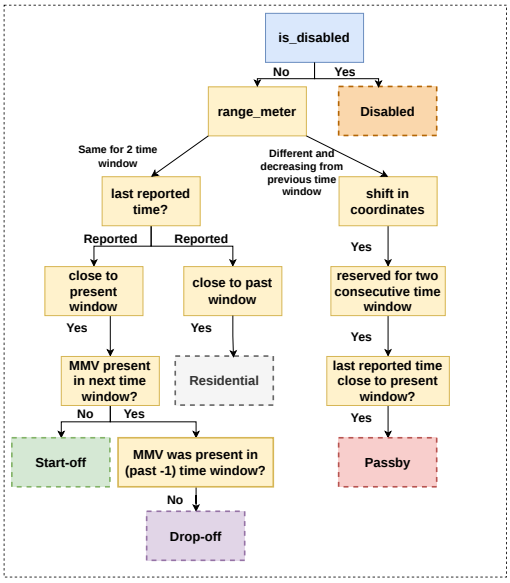

Figure 2: MMV Status Identification

multiple attributes, including spatial proximity, temporal continuity, and physically feasible movement patterns. The reconstruction attack described in the following subsections operationalizes this threat model.

4.2 Identifying MMV Status

As previously discussed, the GBFS feed is updated for currently available MMVs. MMVs that are in use are not reported. Although a lower version of GBFS still updates information of MMVs until a certain time, and then it disappears. Therefore, we have 5 types of status for the MMV based upon its usage type as follows.

- **Residential:** The MMVs that do not change their location for a long duration of time and do not disappear from the system are labeled as Residential MMVs.
- **Startoff:** In this type, the MMVs disappear from the updated feed after Δt units of time and basically come into use by enabling "is_reserved = true", which are called Startoff MMVs.
- **Drop-off:** Here, while updating the feed, if new MMVs appear after a Δt unit of time and disable "is_reserved = false", they are called Drop-off MMVs.
- **Passby:** This type of MMV generally rides through some zones where its locations are sometimes reported despite being in use.
- **Disabled:** Finally, we look for the MMVs that can not be ridden either due to low fuel/battery or because they are defective. We call them Disabled MMVs.

For identifying the above-mentioned MMV status, we use the fields from GBFS feeds as {"current_range_meters", "coordinates", "is_disabled", "is_reserved", "last_reported"} as follows. As depicted in Fig. 2, we first look for the "Disabled MMVs" by checking the field if "is_disabled = true" or not. If it is false, we proceed to identify the next 4 statuses. We check for the "current_range_meters" field, which is precise, unique, and not noisy or anonymized, and is not

repeated for a large number of MMVs. This gives the opportunity to identify the MMVs uniquely. We check this field for subsequent time windows after each Δt unit of time, and if it is different and decreasing, we check for the coordinates corresponding to that MMV. In case we obtain the moving coordinates of that MMV and the "is_reserved" field is also *true* for two consecutive time windows, we label its status as **Passby**.

Next, again we reiterate the same search for the field "current_range_meters" (see Fig. 2, left branch) to look for the next 3 status. In case "current_range_meters" field is same for the MMV for 2 subsequent time windows and the last reported time "last_reported" is close to past time window, we label the status as **Residential**. Now to identify whether an MMV is dropping off or starting off, we check one more field, whether the MMV is appearing in the next window or not. If the MMV is not appearing in next time window, then we label it as **Startoff**. On the contrary, if the MMV is appearing in next time window, we check whether it appeared in past by checking past to past time window. In case, it was not there, we call it **Drop-off**.

Note that, in our proposed algorithm, we do not consider the **Passby** or **Disabled** MMVs as **Passby** MMVs have a little amount of contribution in the data and evaporate shortly. Whereas **Disabled** MMVs are simply moved away by stakeholders or bike companies for recharging, this has very little to do with the inference attack.

4.3 Inference of Origin-Destination Pair

Importantly, we emphasize that `current_range_meters` is not used as a unique identifier. Instead, it serves as an auxiliary signal to reduce the candidate search space. Trajectory derivation relies on the joint consistency of multiple attributes, including spatial proximity (coordinates), temporal continuity (`last_reported`), state variables (e.g., `is_reserved`, `is_disabled`), and physical feasibility constraints. Algorithm 1 shows how we derive trajectories under dynamically changing IDs. We first process the MMV stream at each timestamp within every 6-minute window. Because `current_range_meters` (`crm`) values tend to be locally distinctive within a short time window (but are not globally unique), we use them as a filtering signal to identify small candidate subsets of MMVs. We select only those MMVs whose `crm` frequency < 10 in the feed. In line no. $4-5$ of Algorithm 1, for each such rare `crm` value, we obtain the corresponding subset of MMVs, ideally one but sometimes several. When the subset contains exactly one MMV (refer to line no. 6), we initiate tracking by logging its initial coordinate as $\mathcal{G}_i$, crm as crm_i, and last updated time as t_i as performed in line no. 7, and continue to record its position until the MMV disappears from subsequent data pulls.

Once disappearance is detected as **Drop-off** (line no. $8-9$), we continue to fetch the next available feed at a regular time interval. A candidate match is accepted only when multiple signals jointly agree, including consistency in spatial displacement, temporal continuity, and feasible battery consumption. Thus, `crm` alone is insufficient for identification and only contributes as one of the several validation signals. Consecutively, in line no. $10-12$, we filter out the MMVs whose `crm` has decreased from crm_i. We check the change in `crm` value as Δcrm is consistent with actual travel. Consequently, we fetch the corresponding coordinates $\mathcal{G}_j$ and last reported time t_j for candidate MMV j. If the Harvesine distance between $\mathcal{G}_i$ and $\mathcal{G}_j$ with an average MMV speed of $10km/hr$ is compliant within the time of t_i and t_j, then MMV j is the possible candidate as computed in line no. 12. We also check the drop in `crm` value for MMV i and j to check whether it is consistent with the

Algorithm 1: Plausible Trajectory Derivation with Dynamic IDs

Input: MMV feed $\mathcal{D}$; window $W = 6$ min; `crm` freq threshold f_{crm}
Output: Derived trajectories $\mathcal{T}$

1 Partition $\mathcal{D}$ into windows of length W.
2 **foreach** *window* w **do**
3 Compute `crm` frequencies; let $\mathcal{R}_w = \{\text{crm} \mid \text{freq} < f_{crm}\}$.
4 **foreach** $crm \in \mathcal{R}_w$ **do**
5 $S \leftarrow$ MMVs in w with this `crm`
6 **if** $|S| = 1$ **then**
 `// One-to-One mapping`
7 Initialize track for MMV $i \in S$ with $(\mathcal{G}_i, \text{crm}_i, t_i)$
 Store $u_i \leftarrow \mathcal{G}_i$ as start location
8 **while** i *observable* **do**
9 Pull next feed; detect drop-off of i
10 $C \leftarrow$ MMVs in next feed with $\text{crm}_j \le \text{crm}_i$
11 log MMV j in C with $(\mathcal{G}_j, \text{crm}_j, t_j)$
12 Filter C by: (a) Haversine distance between $(\mathcal{G}_i, \mathcal{G}_j)$ feasible for time $\Delta t = t_j - t_i$, (b) CRM drop consistent with traveled distance
13 Let C = plausible successors.
14 **if** $|C| = 1$ **then**
 `// One-to-One`
15 $i \leftarrow C[1]$ (update state).
16 **else if** $|C| > 1$ **then**
 `// One-to-Many`
17 Keep top $k = 3$ candidates by successive time stamp.
18 **else**
19 Set $v_i \leftarrow$ last known coord of i; break;
 `// Destination found for` i
20 **end**
 `// No successor found`
21 **end**
 `// Finalization: log destination and derive route`
22 If v_i not assigned, set $v_i \leftarrow \mathcal{G}_i$ (final state). Use OSMnx to compute bike-path route between (u_i, v_i) and store resulting geometry as trajectory $\mathcal{T}_i$.
23 **end**
24 **if** $|S| > 1$ **then**
 `// Many-to-One mapping`
25 Cluster S spatially into groups $G_1, \ldots$.
26 For each group G, run the above successor search in parallel. If all MMVs in G point to the same successor, map them jointly.
27 For each MMV in G, record (u_i, v_i) and derive route using OSMnx as above.
28 **end**
29 **end**
30 **end**
31 **return** $\mathcal{T}$.

distance traveled. Now, while running this algorithm, three cases arise as follows.

(1) **One-to-One:** Referring to line no. 14 – 15 in Algorithm 1, there is only one MMV i with unique crm_i and there is only one possible candidate j. Here, the above steps are run until we converge to one possible candidate (j) for MMV i using subsequent time windows. To illustrate, we provide a working example in Fig. 3 from our dataset where we filter the MMVs with `crm` = 10219 and got 14 MMVs at 10 am (See Fig. 3(a)). After 20 minutes, some MMVs disappeared as "startoff" and some reappeared as "drop-off" as shown in Fig. 3(b). We infer the origin and destination of two MMVs where both have a one-to-one mapping by eliminating other successor candidates by infeasible traveling distance as depicted in Fig. 3(c).
(2) **One-to-Many:** As written in line no. 16 – 20 in Algorithm 1, if multiple candidates $\{j, k, l \dots\}$ satisfy the constraints, all are retained as possible successors of i. For example, 2 MMVs with similar distance traveled and similar drop in `crm`. Note that we keep the top three candidates by running the above algorithm for subsequent time windows, as each represents a feasible continuation of the trajectory.
(3) **Many-to-One:** In Algorithm 1, line no. 24 – 26 depicts an interesting case - multiple MMVs can be in very close spatial proximity (i.e., coordinates differ only marginally), in the same time window and with the same `crm` value. But during inference, one successor MMV satisfies all the constraints. In that case, all MMVs in the group are mapped to the same successor, as they are indistinguishable within the GPS resolution.

We do not inspect Many-to-Many case as it is not relevant in trajectory inference.

Finally, we obtain the inferred origin and destination of individual MMVs, which are then used to derive a plausible route between these points. We use OSMnx[7] and NetworkX [14] libraries to compute a shortest feasible path over the OpenStreetMap [19] (OSM) street network. Given a origin and destination coordinate pair, we first calculate the great-circle distance to estimate their spatial separation. Using the origin coordinate as the center, we download the relevant OSM bike network within a 3 km radius, ensuring coverage of cycleways, paths, and bicycle-permissible streets. The origin and destination positions are then matched to their nearest OSM network nodes, after which the shortest path between these nodes is computed using edge length as the optimization weight. The resulting path represents a realistic, map-constrained route that follows the street network rather than a direct straight-line approximation. We emphasize that this route is a plausible estimate of the MMV's movement and *may* not be the exact trajectory taken in reality. Its purpose is to illustrate the feasibility of travel between inferred origin-destination pairs, while the primary privacy risk stems from the accurate inference of these OD points.

4.4 Evaluation

Here, we analyze the performance of our inference approach. To validate the results, we found that in the dataset, out of 98 million MMV entries, 12% of the MMV's IDs did not change dynamically and remained static for up to 1.5 hours, even after completing trips. At the same time, it reveals that the BMT dataset is not completely GBFS compliant. This set of MMVs with static ID represents our concrete ground truth.

4.4.1 Inference of MMV Status. Using our approach in §4.2, we identify a mean of 29.5% startoff, 53.9% residential, 28.4% drop-off, 13.61% disabled, and 1% passby MMVs per hour for a single day as shown in Fig. 4. To further validate our detection accuracy, we run the Xu et al. [29] approach based on static ID to infer startoff and drop-off MMVs. We achieve an accuracy of 95% for startoff and 92% for drop-off. We did not achieve a 100% accuracy due to a number of MMVs sharing the same `crm` and nearly spatial coordinates leading to ambiguity. Further, the accuracy for drop-off is lower as the last reported time is not updated promptly for some MMVs. From Fig. 4, we observe that although a huge number of MMVs are stationary (residential) from midnight to morning hours, start-of-day MMVs increase as the day progresses. An adversary capable of monitoring a remote zone with limited MMV availability can extract substantial actionable intelligence by inferring these MMV status transitions such as startoff, drop-off. Even though these states appear operational and harmless, their temporal patterns reveal predictable behavioral routines within the region. This temporal leakage allows an attacker to infer local mobility patterns and identify when MMVs (and thus riders) are most exposed. In sparsely populated zones, such as near a detached house, these status patterns may even enable clustering over spatial (a street) and temporal (hour of day) dimensions over several days to infer the mobility patterns of an individual or a group of members. Furthermore, such OD pair information are sufficient to infer their home, office, or university, highlighting the privacy risks associated with MMV status.

4.4.2 Origin and Destination Inference for MMVs. Based on our inferred status of MMVs, we detect an average of 427 trip origins for weekdays (Monday to Friday) per hour, as shown in Fig. 5, denoted as start-off MMVs. Similarly, we infer an average of 509 destination for the drop-off MMVs as shown in Fig. 6. The temporal distribution of these inferred origin and destinations exhibits a broadly consistent pattern across the day. In particular, weekdays show two distinct activity surges for startoff MMVs: a morning rise between 06 : 00 and 12 : 00, followed by a pronounced late-afternoon to late-evening rise from 16 : 00 to 21 : 00. In contrast, for the destination inference, the number of drop-offs significantly increases in late morning hours, from 09 : 00 and goes until 21 : 00 hours, for the first weekdays (Monday-Wednesday), but for Friday, sudden peaks after 21 : 00 hours signify more leisurely travels. For the weekends, we achieve an average of 380 origin per hour and 493 destination per hour, showing a trend of more availability of MMVs during the weekend. As the patterns of startoff and drop-off MMVs remain consistent across the week, spatio-temporal clustering over zones with sparse availability is sufficient for reconstructing a meaningful origin–destination pattern. Next, we demonstrate how our approach, utilizing the origin and destination, derives the plausible trajectory of individual MMVs.

4.4.3 Performance on Derivation of Route. We derive 85575 (82.91%) trajectories as shown in Fig. 7 using Algorithm 1 which

[7] https://osmnx.readthedocs.io/en/stable/ (Accessed Online: November 22, 2025)

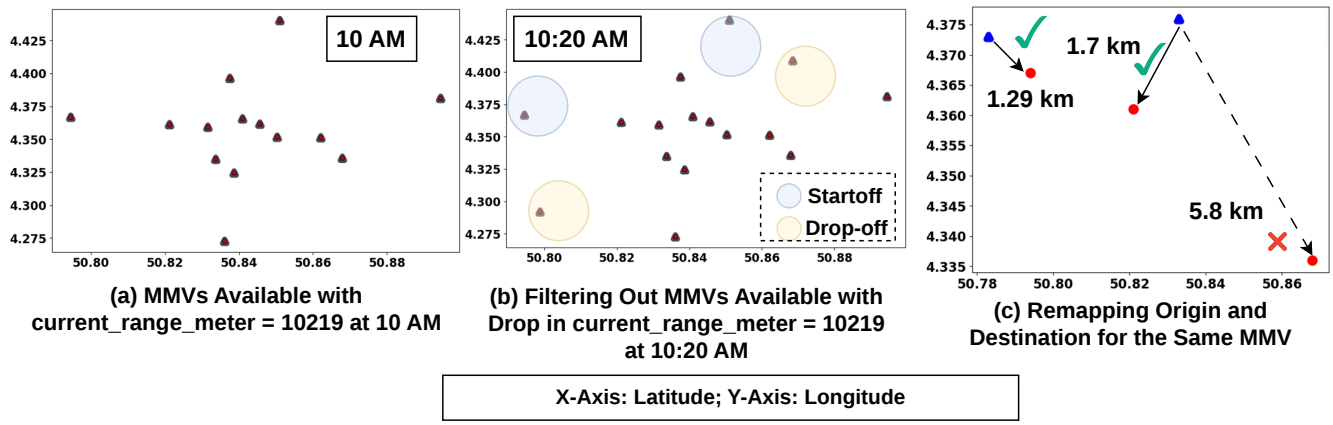

Figure 3: Snapshot of a Inference Attack using Algorithm 1 on data from 2nd September, 2024. (a) Filter out available MMVs with `crm` = 10219 at 10 AM, (b) log the coordinates and last_reported timestamp for each MMV for every 6 minutes, at 10 : 20 AM, 2 MMVs disappeared as *startoff* and 2 new MMVs appeared as *drop-off* with the same `crm` value, (c) in later time windows, look for a drop in `crm` value feasible with the difference in timestamp. Record their coordinates and calculate the Haversine distance. For the MMV in the upper left corner, we found the destination as the traveled distance (1.29 km) is feasible with the drop in `crm` value. For the second MMV, two candidates are found, one with 1.7 km distance traveled and the other in the lower right corner, traveled 5.8 km, which is infeasible with the drop in `crm` value. Therefore, the former becomes its destination, and finally, two trajectories could be derived.

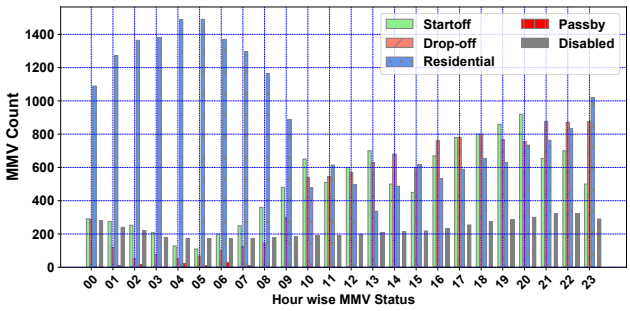

Figure 4: MMV Status Inference

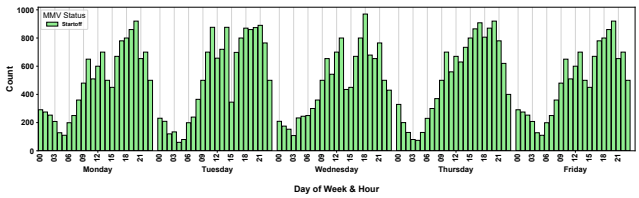

Figure 5: MMV Trip Origin Inference

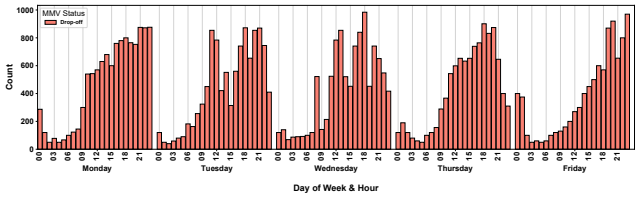

Figure 6: MMV Trip Destination Inference

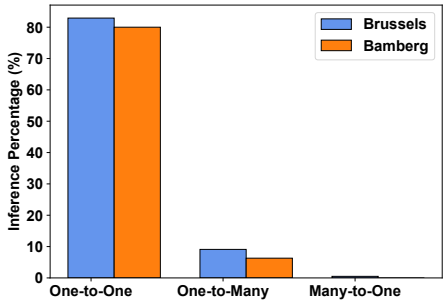

Figure 7: Percentage of Trips Derived for 2 Cities

have one-to-one mapping. For one-to-many, we obtain 12347 (9.1%) and many-to-one mapping, we get 683 (0.5%) trajectories. The analysis is reported for the whole month of September (1 – 30), 2024 for Brussels. To verify the validity of our approach, we obtained 135697 MMV entries for September, which were used to complete trips without changing their IDs, i.e., static. Therefore, we achieve an accuracy of 82.91% for one-to-one mapping. For the one-to-many mapping, we cross-validated whether the actual MMV is among the top 3 successors that we inferred. For 98% of the cases, we achieve the above-mentioned validation. The remaining 2% are failed because our algorithm works in a greedy manner using timestamps in non-increasing order and storing the consecutive best matched successors. Consequently, it failed to capture the correct successor, which appeared in a later timestamp. For the many-to-one mapping,

we only get at most 2 MMVs sharing the same location (in terms of coordinates), timestamp, and `crm` values. We suspect the issue is due to GPS noise, which reflects in the dataset as showing the same coordinate readings for two MMVs. But, for all the 683 MMVs, it correctly inferred the successor where the successor's ID matched with one of the two MMVs.

Spatial Hotspots: While performing the inference, we observed repetitive trajectory patterns for some zones. First, we perform density-based clustering using DBSCAN [10] for the spatial trajectories only. We report 5 top spots where most of the trajectories are discovered, such as in the city center, {Rue du Sceptre, Marollen}, and outskirts {Aveneu Leopold III, Eugène Ghijsstraat, Rue Bollinckx}. The regions are well populated with a high density of MMVs, e.g., $\approx$ 300 in an hour. Still, a subset of GBFS fields with precise values makes the mobility data vulnerable. Furthermore, we performed temporal grouping for every hour of the day throughout the month for these zones based on the derived trajectories and their respective timestamps. We receive trajectories that repeat during morning hours on weekdays from the origin Rue du Sceptre to Stefania and also to Rue Vautier. These places are well-crowded and have significant points of interest, such as metro stations, hotels, and university libraries. As we do not have an external source of information, we could not verify further whether these trips are done by the same set of people or an individual. However, access to such information also gives the attacker the ability to infer their profession (e.g., a student, hotel employee) or even stay points, such as office locations. This becomes more dangerous for the outskirts location Aveneu Leopold III, where the number of MMV availability is also very low, i.e., $3-5$ per 10 minutes. Here, we discovered trips are more repetitive during weekends and end at the central station during evening hours. We refrain from reporting exact hours for ethical considerations and to respect the privacy aspect.

Bamberg Statistics: Now, we report on deriving trajectories for another mid-scale city, Bamberg, where the rental operator is ZEUS[4] and the API data is logged for 1 week (April 1–7, 2025) in a 10 minute interval. As depicted in Fig. 7, we could derive 80% trajectories for 7 days for one-to-one mapping. For one-to-many, we achieved 6.8% and 0.5% for many-to-one mapping. But as the API follows GBFS version 2.2, it does not suffer from static ID. This also prevents us from creating the ground truth for further validation and verify the accuracy of the derived trajectories. But using our above approach of spatial clustering, we identified the top zones as ZoB, Feldkirchen Str., An d. Spinnerei, Gutenberg Str. and Laubanger. Using temporal grouping by an hour for the trajectories, we infer repeating trips from An d. Spinnerei to Gutenberg Str. during morning hours and the reverse trip during evening hours in weekdays. Similarly, on weekends, major trips from Feldkirchen Str. and Laubanger are ending at ZoB during afternoon hours. As these Feldkirchen Str. and Gutenberg Str. have fewer MMVs ($<$ 5 and sometimes even 1) during peak hours, it becomes very easy to infer other member sensitive information by observing their mobility pattern, or if already the target group is known.

5 Mitigation

Here, we discuss a set of defensive methods designed to prevent the privacy attacks highlighted earlier in the paper and evaluate their implications on data utility. Given the unique challenges posed by GBFS-based micromobility data, effective privacy preservation is essential without sacrificing the utility of the data for urban mobility analysis. To evaluate the mitigation strategies, we use a real GBFS dataset collected from the BMT API on June 1, 2025 polled at a regular interval of every 6 minutes for 24 hours. This dataset was selected as a representative daily snapshot to analyze how different mitigation techniques affect attribute uniqueness and inference feasibility, independent of the September 2024 dataset used for OD inference evaluation. We begin by presenting our proposed mitigation approach, followed by secondary risks that may get introduced through the proposed strategies, specifically tailored for GBFS feeds.

5.1 API Limit

Currently, BMT Data allows querying the API every 6 minutes for a user API key. A user can create multiple APIs and thus can even create a historical dataset every minute. One possible prevention is to impose an API key creation limit, i.e., a single key per user. Secondly, limiting the number of queries per API and increasing the duration of subsequent queries to 6 minutes. As we have analyzed, the maximum duration of a trip is an hour or so; thus, limiting the duration to $15-30$ minutes prevents the inference of the trajectory to a great extent. We have tested this time duration limit with 15 minutes and 30 minutes. As either the origin or destination falls within a single time window for most short-duration trips of less than 20 minutes, we cannot infer their trajectory. Therefore, only 24% of the trips of longer duration can be inferred using a 30 minute API query duration limit. For a 15-minute duration, inference increases by up to 41%; therefore, we suggest keeping the duration at 30 minutes. However, this approach primarily increases the operational cost for an adversary. A deterministic attacker could still evade the API rate limit, e.g., by deploying a bot farm or acquiring multiple API keys, and therefore the attack cannot be fully prevented.

5.2 Coarsen Vulnerable Fields

As identified, the most vulnerable fields are "last_reported", "current_range_meters", and "coordinates". To limit the adversary's ability to derive a plausible individual MMV trajectory, we adopt a coarsening-based defense applied to the most sensitive fields exposed through GBFS. To mitigate this, GPS coordinates are binned to 3 decimal places ($\approx$ 110 m spatial granularity, compared to the current 0.11 m), effectively grouping nearby MMVs into common spatial cells and removing precise movement traces. Additionally, temporal information is generalized from exact Unix timestamps to broad time-of-day categories (morning, afternoon, evening, night), thereby suppressing short-interval movement cues. Finally, current_range_meters is coarsened into the next available floor value to break the correlation between small battery drops and micro-movements.

Together, these transformations substantially reduce the adversary's ability to link sequential GBFS snapshots to the same MMV by collapsing multiple vehicles into indistinguishable spatio-temporal groups. To quantify this effect, we analyze the distribution of MMVs grouped by their joint attributes ("last_reported", "current_range_meters", and "coordinates"). Fig. 8 shows the cumulative

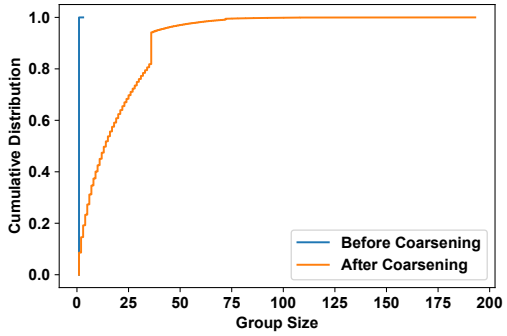

Figure 8: Exemplary Cumulative Distribution Function of Singleton Groups Before and After Suggested Coarsening ("last_reported", "current_range_meters" and "coordinates")

distribution of group sizes before and after coarsening. Before coarsening, nearly all MMVs belong to singleton groups, i.e., individually distinguishable MMVs, within each time window ($\approx$ 99.98%), indicating that they are uniquely identifiable based on their attributes and can be reliably tracked across snapshots. After applying coarsening, however, multiple MMVs fall into the same spatial, temporal, and battery-level bins, significantly increasing group sizes. As a result, the proportion of singleton groups drops sharply to $\approx$ 8.5%, meaning that the vast majority of MMVs become indistinguishable from at least one other MMV within the same spatio-temporal cell. This loss of distinguishability introduces ambiguity in matching observations across time. While singleton prevalence does not directly correspond to the success of trajectory derivation, it is a strong indicator of the adversary's ability to uniquely identify candidates for linkage.

Consistent with this observation, the coarsening analysis was repeated across multiple runs on the same BMT dataset collected throughout the day, and consistent trends in both singleton prevalence and inference reduction were observed. Trajectory derivation using raw GBFS data achieved approximately 83% recoverability in our setting, whereas applying the proposed coarsening reduced derivation success to 3%. Importantly, we apply the same inference pipeline before and after coarsening to isolate the effect of information reduction; therefore, we do not claim optimality against adaptive inference strategies specifically designed for coarsened data. Targeted coarsening of vulnerable fields offers an effective privacy mitigation strategy but can impact data utility. For example, publishing GPS coordinates with a granularity of $\approx$ 110 m may confuse users in densely populated areas. However, this might not be an issue in sparser regions, as it will be easy to locate. In contrast, coarsening the current_range_meters to its floor value always provides a buffer against travel distance, preserving the user experience. Therefore, the degree of coarsening should be adjusted based on other contextual factors (e.g., the city's density) and data utility, so that it does not degrade the user experience.

5.3 $K-$anonymity

An alternative approach to protect MMV users against inferring origin-destination and plausible trajectory reconstruction is $k-$anonymity [25], which ensures that each released record is indistinguishable from at least $k-1$ other records in the dataset. In the context of GBFS, $k-$anonymity can be applied by aggregating MMVs reports spatially and temporally. For example, releasing only the presence of at least k MMVs within the same geospatial zone during a given time interval. This prevents an adversary from uniquely linking an MMV's movement over consecutive snapshots, thereby reducing the risk of reconstructing individual trajectories. We tried this approach with $k = 3$, but a huge amount of data (38%) got dropped to ensure 3$-$anonymity. As GBFS data targeted for MMV is highly important for both operators and citizens, losing a substantial amount of data and publishing the rest will falsify the actual availability and may lead to operational loss. Thus, we do not suggest this approach even at the cost of ensuring high privacy.

5.4 Discussion: Trade-offs, Secondary Risks, and Implications for GBFS

While the proposed mitigation strategies significantly reduce reconstruction risks, they introduce secondary effects and highlight important implications for the evolution of the GBFS standard.

Secondary Risks and System-level Effects: Spatial and temporal coarsening reduce the identifiability of individual MMVs, but may still enable group-level inference, such as identifying demand hotspots, commuting patterns, or fleet imbalances. For instance, coarsened data can reveal demand hotspots and deadspots, peak usage periods, or recurring commuting patterns across zones. While such insights do not allow linking origin-destination to specific users, they may expose commercially sensitive information, such as fleet utilization trends or strategic deployment gaps. Competing operators or third parties could exploit such aggregate patterns for strategic advantage; for instance, a competitor might identify temporarily undersupplied hotspots from another operator's data and proactively deploy a large number of MMVs to those areas to capture demand and increase revenue. This highlights a shift from individual privacy risks to aggregate information leakage, which is inherent to any data publishing mechanism that retains utility. Reduced temporal resolution and spatial precision also decrease the observability of the system, which may enable adversarial manipulation. For example, due to delayed or coarse updates, an attacker could temporarily create artificial scarcity or availability in a region that persists in the published data, thereby misleading users or influencing demand and rebalancing decisions. For instance, consider a zone where coarsened data reports approximately two available vehicles with updates every 30 minutes. An attacker could quickly reserve or physically relocate these vehicles, while the system continues to display their availability until the next update cycle. This mismatch between the reported and actual state can mislead users, creating artificial perceptions of availability or scarcity, and may further influence demand patterns or operator rebalancing decisions. These risks arise from reduced observability rather than increased identifiability. Finally, coarsening may impact downstream analytics (e.g., rebalancing or maintenance planning) and disproportionately affect low-density regions, where each data point carries higher significance. In such areas, spatial and

temporal aggregation can obscure the presence or movement of a small number of MMVs, leading to inaccurate demand estimation and potentially reduced service allocation. This may introduce fairness concerns, as users in already sparse regions could experience systematically lower service quality compared to high-density areas.

Implications for GBFS Standardization: Our findings suggest that these risks stem not only from individual deployment choices, but also from the lack of standardized privacy mechanisms in current GBFS implementations. To address this, the GBFS standard should incorporate configurable privacy-aware controls rather than relying on ad hoc operator policies. **First**, the standard should define API governance mechanisms that distinguish between low-frequency user queries and high-frequency automated access, enabling rate-limiting or delayed/coarsened responses for bulk data collection. **Second**, it should support configurable coarsening of sensitive fields, such as spatial granularity, temporal resolution, and discretization of operational attributes, allowing deployments to adapt privacy levels based on contextual factors like urban density. **Third**, the standard may include optional support for aggregation-based reporting (e.g., k-anonymity-inspired mechanisms) for non-interactive use cases, despite their known utility trade-offs. **Finally**, We believe that studying how different coarsening configurations perform across dense and sparse mobility regions would provide valuable guidance for deploying context-aware privacy protections in GBFS feeds.

Balancing Privacy, Utility, and Deployability: Importantly, we do not advocate a single mandatory configuration. Instead, privacy mechanisms should be exposed as standardized, interoperable controls that operators can tune based on their operational and regulatory requirements. Our empirical results (e.g., reduction from 83% to 3% reconstruction success through coarsening, and the limitations of k-anonymity due to data loss) highlight that lightweight, configurable defenses can provide strong privacy benefits while preserving usability. Overall, embedding such mechanisms into the GBFS standard would enable more consistent privacy guarantees across deployments while maintaining the utility of real-time mobility data.

6 Conclusion

Our work demonstrates that the GBFS-based data shared by micromobility operators, while essential for improving urban mobility and enabling integration in third-party applications, exposes significant privacy vulnerabilities. By leveraging the high precision of GPS coordinates, battery levels, and timestamps, we primarily demonstrate that it is possible to derive detailed trip origin-destination pair and plausible map-constrained trajectories which may contain potentially sensitive user information, such as home or work locations, even in the absence of explicit identifiers like vehicle IDs. Using data collected from two distinct cities, Brussels and Bamberg, we were able to reconstruct the origins and destinations of over 80% of trips, highlighting the unintended privacy risks posed by different versions of GBFS feeds.

While the dynamic vehicle IDs in GBFS are designed to protect user anonymity, we found that GBFS version 1.0 has potential privacy issues, as it does not change the MMV ID for 12% of the data and reveals 1% of MMVs' locations while in use, analyzed over a month. Despite these privacy concerns, our study provides a practical solution by proposing a set of anonymization techniques that can be applied to the existing GBFS standard. Our results show that data generalization, removal of certain quasi-identifiers, and other mitigation strategies can reduce the success of inference attacks by up to 97%, while maintaining sufficient data utility for urban mobility applications. This represents a significant step forward in ensuring the privacy of micromobility users without compromising the operational value of the data.

However, our work is not without limitations. First, our study is based on data from only two cities. While we demonstrated the generalizability of our approach, further work is needed to validate our findings across a broader range of cities and micromobility operators, particularly in non-European contexts. Another limitation of our approach is that back-to-back rentals of the same MMV may occasionally be merged if the vehicle's unavailable interval is not captured between successive API queries. Given the 6-minute polling rate, such cases are expected to be infrequent but may lead to slight overestimation of trip continuity.

Additionally, while our anonymization techniques are promising, we do not claim that our algorithm is optimal. There remains the potential for novel inference attacks that could circumvent these defenses. Thus, in the future direction, we would focus on developing more robust privacy-preserving techniques, potentially incorporating emerging approaches like differential privacy, which could offer formal privacy guarantees while allowing for the continued utility of micromobility data. In a nutshell, our work highlights the ongoing challenge of balancing privacy with data utility in shared mobility systems. As micromobility continues to grow, it will be crucial to refine existing standards and develop new privacy-preserving mechanisms that can protect individuals' data while still enabling the benefits of open, real-time mobility information for end-users and city managers. Nonetheless, to the best of our knowledge, this paper contributed to designing a reconstruction attack-based algorithm and subsequently mitigating it, applied over 2 different cities (Brussels and Bamberg), operators (LIME and ZEUS), and GBFS versions (1 and 2.2).

7 Ethical Consideration

The research in this paper was conducted in full compliance with ethical standards. Before data collection, we ensured no personal or sensitive information was used without consent. Data from shared mobility operators were accessed via publicly available APIs, in strict accordance with their terms of service and privacy policies. We did not access private or restricted data and followed all guidelines for API usage. Furthermore, the study did not involve direct interaction with individuals or collection of personally identifiable information (PII). It did not target any specific group or make claims that could identify or profile users. The focus was on analyzing shared mobility data privacy risks and the impact of spatiotemporal data. In addition, we carried out responsible disclosure to the dataset maintainers 14 weeks prior to the revised manuscript submission (as of April 29, 2026), informing them that approximately 12% of MMVs remained static for durations of up to 1.5 hours.

Acknowledgments

This work was funded by the German Federal Ministry of Research, Technology and Space and the European Union (NextGenerationEU). During the preparation of the manuscript, the authors used Grammarly [13] to check for grammatical error, flow of writing and spelling checks. While using this tool, the authors reviewed and edited the resulted text as and when necessary. To check for the originality in writing, authors have used Turnitin [26] software on text similarity to ensure the originality in writing. We also want to thank the PoPETs reviewers and the editor for their constructive comments, which have improved our paper substantially.

References

[1] Layth Riyadh Abdulrazzaq, Mohammed Naeem Abdulkareem, Muhamad Razuhanafi Mat Yazid, Muhamad Nazri Borhan, and Mina Salah Mahdi. 2020. Traffic congestion: Shift from private car to public transportation. *Civil Engineering Journal* 6, 8 (2020), 1547–1554.

[2] Leonie Ackermann, Michael Mühlhauser, Alexandru Burdusel, Michael Federlin, Dominik Herrmann, Steffen Holly, Daniela Nicklas, and Daniel Wolpert. 2023. Towards Anonymizing Intermodal Mobility Data for Smart Cities. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geo-Privacy and Data Utility for Smart Societies*. 24–27.

[3] Marco Balossini, Michele Carminati, Stefano Zanero, and Stefano Longari. 2025. Micro-Mobility Security: A Holistic Approach via Mobile App Analysis. (2025).

[4] Guillermo Baltra, Basileal Imana, Wuxuan Jiang, and Aleksandra Korolova. 2020. On the data fight between cities and mobility providers. *arXiv preprint arXiv:2004.09072* (2020).

[5] Marco Casagrande, Riccardo Cestaro, Eleonora Losiouk, Mauro Conti, and Daniele Antonioli. 2024. E-Trojans: Ransomware, Tracking, DoS, and Data Leaks on Battery-powered Embedded Systems. *arXiv preprint arXiv:2411.17184* (2024).

[6] Debasree Das, Rahat Rafiq, and Daniela Nicklas. 2025. Does One Noise Fit All? Analyzing Utility in Transportation Mode-Based Anonymization. In *2025 IEEE International Conference on Smart Computing (SMARTCOMP)*. 216–218. https://doi.org/10.1109/SMARTCOMP65954.2025.00042

[7] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3, 1 (2013), 1376.

[8] Cynthia Dwork. 2006. Differential Privacy. In *Automata, Languages and Programming*, Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener (Eds.). Springer, Berlin, Heidelberg, 1–12. https://doi.org/10.1007/11787006_1

[9] Karina Elzer, Eric Jedermann, Stefanie Roos, and Jens Schmitt. 2025. They See Me Scooting—A Long-Term Real-World Data Analysis of Shared Micro-Mobility Services and their Privacy Leakage. In *2025 IEEE 10th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 78–92.

[10] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (Portland, Oregon) *(KDD'96)*. AAAI Press, 226–231.

[11] Liyue Fan and Ishan Gote. 2021. A Closer Look: Evaluating Location Privacy Empirically. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 488–499.

[12] Julien Freudiger, Reza Shokri, and Jean-Pierre Hubaux. 2011. Evaluating the privacy risk of location-based services. In *International conference on financial cryptography and data security*. Springer, 31–46.

[13] Grammarly. 2026. Grammarly. Computer software. https://www.grammarly.com Accessed Online: May 7, 2026.

[14] Aric Hagberg, Pieter J Swart, and Daniel A Schult. 2008. *Exploring network structure, dynamics, and function using NetworkX*. Technical Report. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).

[15] Gizay Kisa Isik, Theo Tryfonas, and George Oikonomou. 2023. E-scooter Sharing Platforms: Understanding Their Architecture and Cybersecurity Threats. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. 5909–5916. https://doi.org/10.1109/ITSC57777.2023.10421849

[16] Alexandra Kapp. 2024. Collection, usage and privacy of mobility data in the enterprise and public administrations. *arXiv preprint arXiv:2407.03732* (2024).

[17] John Krumm. 2007. Inference Attacks on Location Tracks. In *Pervasive Computing*, Anthony LaMarca, Marc Langheinrich, and Khai N. Truong (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 127–143.

[18] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. l-diversity: Privacy beyond k-anonymity. *Acm transactions on knowledge discovery from data (tkdd)* 1, 1 (2007), 3–es.

[19] OpenStreetMap contributors. 2017. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

[20] Daniel Oviedo. 2021. Making the links between accessibility, social and spatial inequality, and social exclusion: A framework for cities in Latin America. In *Advances in transport policy and planning*. Vol. 8. Elsevier, 135–172.

[21] Daniele Quercia, Ilias Leontiadis, Liam McNamara, Cecilia Mascolo, and Jon Crowcroft. 2011. Spotme if you can: Randomized responses for location obfuscation on mobile phones. In *2011 31st International Conference on Distributed Computing Systems*. IEEE, 363–372.

[22] Protection Regulation. 2018. General data protection regulation. *Intouch* 25 (2018), 1–5.

[23] Chiara Renso, Miriam Baglioni, Jose Antonio F de Macedo, Roberto Trasarti, and Monica Wachowicz. 2013. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and information systems* 37 (2013), 331–362.

[24] Mahmoud Sakr and Gaspard Merten. 2023. Brussels Mobility Twin *(SIGSPATIAL '23)*. Association for Computing Machinery, New York, NY, USA, Article 84, 4 pages. https://doi.org/10.1145/3589132.3625634

[25] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems* 10, 05 (2002), 557–570.

[26] Turnitin. 2026. Turnitin. Computer software. https://www.turnitin.com/ Accessed Online: May 7, 2026.

[27] Nisha Vinayaga-Sureshkanth, Raveen Wijewickrama, Anindya Maiti, and Murtuza Jadliwala. 2022. An investigative study on the privacy implications of mobile e-scooter rental apps. In *Proceedings of the 15th ACM conference on Security and Privacy in Wireless and Mobile Networks*. 125–139.

[28] Fengli Xu, Zhen Tu, Yong Li, Pengyu Zhang, Xiaoming Fu, and Depeng Jin. 2017. Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In *Proceedings of the 26th international conference on world wide web*. 1241–1250.

[29] Yiming Xu, Xiang Yan, Virginia P Sisiopiku, Louis A Merlin, Fangzhou Xing, and Xilei Zhao. 2022. Micromobility trip origin and destination inference using general bikeshare feed specification data. *Transportation Research Record* 2676, 11 (2022), 223–238.

[30] Hui Zang and Jean Bolot. 2011. Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th annual international conference MobiCom*. 145–156.

[31] Yongping Zhang. 2025. Infer Trip Purposes of Shared Micromobility Using a LDA-Based Analytical Framework. In *Understanding Shared Micromobility in China Using Massive User-Generated Trip Data*. Springer, 49–79.