

Proposal

To implement and evaluate my proposed method for enhancing Geo-Indistinguishability (Geo-I) for trajectory privacy in urban environments, specifically in the context of Ho Chi Minh City, I plan to utilize the following tools, frameworks, and datasets:

1. Geographic Data and Spatial Analysis Tools

- **OpenStreetMap (OSM):** An open-source geographic database providing detailed maps of urban environments, including road networks, building footprints, water bodies, and points of interest (POIs). OSM data for Ho Chi Minh City will form the primary geographic reference to ensure the generation of semantically plausible fake trajectories.
- **OSMnx and GeoPandas:** Python libraries designed for processing and analyzing geographic and spatial data. OSMnx will be utilized to extract and manipulate road networks, while GeoPandas will assist in spatial queries and filtering, ensuring generated points avoid unrealistic locations such as rivers, building interiors, or restricted areas.

2. Map-Matching and Route Validation Tools

- **Open Source Routing Machine (OSRM):** A routing engine capable of performing map-matching, which aligns noisy trajectory points to real-world road networks. OSRM will enable post-processing of generated trajectories to maintain route realism and adherence to physical road constraints.
- **Hidden Markov Model (HMM)-based Map Matching Algorithms:** Algorithms using probabilistic models for trajectory correction and optimization. These algorithms will ensure generated fake trajectories correspond to realistic paths users might actually follow, thereby enhancing semantic plausibility and service quality.

3. Privacy-Preserving Frameworks

- **CORGI Framework (User Customizable Robust Geo-Indistinguishability):** A customizable obfuscation framework designed for Geo-I, allowing users to exclude or constrain locations where fake points should not appear. I plan to adapt and extend the CORGI framework to integrate semantic-aware constraints and road network filters specific to Ho Chi Minh City.
- **Custom Geo-I Implementations:** Leveraging literature-based algorithms, including semantic-aware indistinguishability and optimization-based selection methods (as described by Yan et al. (2022)), I will develop a tailored Geo-I algorithm to balance privacy with trajectory realism effectively.

4. Evaluation Datasets

- **Microsoft GeoLife Dataset:** A publicly available dataset comprising extensive real-world human trajectories from various urban settings, primarily in Beijing. This dataset will provide realistic trajectory patterns and movements for initial testing and validation of the privacy mechanisms.
- **Cabspotting Dataset (San Francisco Taxis):** A widely-used urban vehicle trajectory dataset providing realistic driving patterns within a dense urban environment. While this dataset originates from San Francisco, it serves as a useful benchmark for evaluating how effectively my method maintains utility while preserving privacy.
- **Simulated Trajectory Dataset for Ho Chi Minh City:** Due to the possible lack of publicly available trajectory datasets for Ho Chi Minh City specifically, I propose to generate simulated user trajectories based on realistic urban mobility patterns using OSM data. This dataset will enable direct testing and validation of my privacy-preserving mechanisms within the precise geographic and infrastructural context of Ho Chi Minh City.