De-anonymising data in Wi-Fi positioning systems

Michael C. Fink Amores
University of Cambridge
mcf61@cam.ac.uk

Alastair R. Beresford
University of Cambridge
arb33@cam.ac.uk

Location-based services (LBSs) like Apple Maps rely on crowdsourced location data to provide their functionality, such as routing, real-time traffic updates, or recommendations. Individual users have to disclose their location for these personalised services to work, but their information may also be aggregated to provide services for other users. Unfortunately, GPS requires high energy consumption from a device and has availability issues. Mobile devices therefore use, whenever possible, *Wi-Fi Positioning Systems* (WPSs) to trilaterate their location based on the location of Wi-Fi *Access Points* (APs). These systems work in multiple stages. In the first stage, the device measures location information, e.g., by scanning and detecting *Basic Service Set Identifiers* (BSSIDs) of nearby APs. The locations of these APs have previously been shared with the service by other users. In the second stage, the device infers its location by querying the location database provided by the service.

With its billions of users, Apple offers one of the world's largest and most precise crowdsourced WPSs. Devices exchange data with Apple servers through serialised structured data encoded as protocol buffer messages. When analytics are enabled in the *Location privacy settings*, GPS-enabled iOS devices continuously share location data with Apple's endpoint (`gsp10-ssl.apple.com`) whenever LBSs are used. [2] This includes a device's current geodetic coordinates (e.g., `[52.123, 1.123]`), the BSSIDs of nearby APs (e.g., `22:33:44:55:66:55`), and additional metadata such as its *motion activity type*, indicating whether a user is, for example, stationary, running, or walking. Only the locations of devices which have remained stationary over multiple days are added to the database. To further prevent abuse, Apple filters out mobile devices, such as smartphones or smartwatches, based on *Organizationally Unique Identifiers* that identify the manufacturer of a device. A mobile device can query the database by sharing the BSSIDs of nearby APs, its own model (e.g., `iPhone14,5`), and its operating system (e.g., `iPhone OS16.7.820H90`) with the main API. Apple's location database (`gs-loc.apple.com`) then responds with the BSSIDs and corresponding geodetic coordinates of up to 400 more nearby APs, which are cached for further use. The approximate location is consequently computed on the device itself. Surprisingly, this API is public, can be accessed without a key, does not require a physical Apple device, and allows a user to request the locations of APs which are not physically close to the querying device.

Apple makes strong privacy claims for their system: "The crowd-sourced location data gathered by Apple is stored with encryption and doesn't personally identify you". [1] We investigate if such claims hold. Prior research already shows how an uninformed adversary is able to abuse the system and track semi-stationary devices such as travel routers over a considerable timespan by continuously querying the database with a script. [3, 4] We assume a server-side adversary without background knowledge, for example a malicious employee or an individual who has illicit access to the location database and can observe both query-data of geolocation attempts and updates to the locations of APs over a fixed period of time, e.g., a week. We will investigate if such an adversary can perform a passive attack to extract a device's trajectory. More specifically, given a collection of HTTPS exchanges between the server and a set of smartphones over a considerable timespan, both queries and database updates, we want to see if we are able to link together successive geo-location queries and group them by the underlying user. This time-series data would then allow us to reconstruct individual user trajectories. Grouping is a challenge as a user's identity is not included in the query. We explore whether we can track individuals by reference to the user's (possibly changing) IP address, model identifier and iOS version data, combined with trajectory recovery attacks based on the data submitted and received. Trajectory recovery attacks have previously been shown to be successful for de-aggregating aggregated mobility data. [5]

[1] Apple. About privacy and Location Services on iOS, iPadOS and watchOS. `https://support.apple.com/en-gb/102515`, 2024.

[2] Apple. Location Services & Privacy. `https://www.apple.com/legal/privacy/data/en/location-services/`, 2025.

[3] Erik Rye and Robert Beverly. IPvSeeYou: Exploiting Leaked Identifiers in IPv6 for Street-Level Geolocation, 2022.

[4] Erik Rye and Dave Levin. Surveilling the Masses with Wi-Fi-Based Positioning Systems, 2024.

[5] Fengli Xu, Zhen Tu, Yong Li, Pengyu Zhang, Xiaoming Fu, and Depeng Jin. Trajectory Recovery From Ash: User Privacy Is NOT Preserved in Aggregated Mobility Data. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 1241–1250, 2017.