

The SBLE data is recorded in two different files:

- SBLE_data
 - Contains timestamps, user names, and all other measurements.
 - Major/minor: These indicate the bus vehicle (Major) and the location of the beacon (Minor; 1:front, 2:back)
 - Each entry has a measurement from either beacon (if data from that beacon was received) along with data from other sensors in the phone
 - Relevant measurements for us are:
 - latitude/longitude
 - rssi (power of the signal from beacon)
 - We can neglect the other measurements for now
 - Measurements from the two beacons (two minors, same major) are normally taken at the same time. When this is the case, we have two rows (different minors) where the rssi is different (because it comes from different beacons) but all other measurements (including latitude and longitude) are identical
- SBLE_notification
 - Contains timestamps, user names, message_types, and messages
 - Timestamps are synchronized with those of SBLE_data. In most cases, for each timestamp in SBLE_notification you can find the same timestamp in SBLE_data. In some cases, this is not true
 - When the user left the bus and walked away, and there is no signal from the beacons thus no data recorded
 - Message_types:
 - Collecting_data
 - TRUE: The app sensed the presence of a beacons and started collecting data
 - This is recorded only the first time that data is received. The system only “resets” when:
 - The app crashes or is restarted
 - The user notified the system that they left the bus (see below)
 - Until either of these events happens, the system will never record a new collecting_data=TRUE
 - FALSE: User replies ‘yes’ to the prompt “Are you off the bus?”
 - This prompt is only generated if the user earlier confirmed that they are on the bus (replying yes to the prompt “Are you in the bus?”; see below) and the system does not received data for a while
 - ‘answered no’: User replies ‘no’ the prompt “Are you off the bus?”
 - **Note:** This option was only enabled in September 2023. Before that, the system would not record anything if the user replied ‘no’ to the prompt “Are you off the bus?”
 - Sitting_on_bus

- TRUE: The user responded 'yes' to the prompt "Are you in the bus?"
 - This prompt is generated when the system senses that the users has been moving by a certain speed for a certain time
- FALSE: The user responded 'no'
 - **Note:** The 'FALSE' option was only enabled in September 2023. Before that, the system would only record TRUE messages
- Users may decided not to respond to the prompt, in which case nothing is recorded
- seat_location
 - 'front', 'middle' or 'back', as selected by the user when prompted by the "Which part of the bus are you on?" prompt
 - This prompt is generated after the user replied 'yes' to the prompt "Are you in the bus?"
 - Note that the user may request a new "Which part of the bus are you on?" prompt anytime while on the bus
 - Users may decide not to reply to the prompt, in which case nothing is recorded

Our goals are:

- Reconstruct each individual trip, from the time the user confirms that they are in the bus (sitting_on_bus=TRUE) to the time they confirm that they left the bus (sitting_on_bus=FALSE). These are shown in blue in the Matlab visualization.
 - Individual trip = one trip by one user on one bus (bus=Major)
 - Note that in same trip, there may be occasional recording from beacons in other nearby busses (different majors)
 - Find beginning/end timestamps
 - Read timestamped rssi from both beacons, along with latitude/longitude
 - Also read rssi from different majors, if any (see above)
 - Read seat location, including any changes if multiple seat_location messages during trip
- Reconstruct each individual "pre-trip" (shown in red in Matlab visualization)
 - Pre-trip = time from when data from a major was first recorded, to the time user confirmed they are in bus (sitting_on_bus=TRUE) while receiving data from the same major
 - Same measurements as for trips

Challenges:

- We need to consider the following situations:
 - A collecting_data = TRUE followed a very long time later by sitting_on_bus=TRUE

- For example: the user may have been at a bus stop when a shuttle was passing by (recorded collecting_data = TRUE), then took a Metro bus, then later on took the shuttle. [There is at least one such case]
 - In this case we would have an unrealistically long pre-trip
 - We may need to add a threshold on the length of a pre-trip
- Data from the same bus being recorded long after a trip ended.
 - This may happen if the user did not respond to the "leaving the bus" notification, and then takes the same bus later on.
 - In this case we would have an unrealistically long trip
 - Typically this would come with a long gap of measurements. We may need to set a threshold for these "gaps"
- Unexplained situation: a seat_location = * followed by a collecting_data = TRUE, with the same major [this occasionally occurs]
 - This can only occur if the app crashed or the user closed it.

Observations:

- For Major=2, it seems that we only recorded data from Minor=1 [need to double check]
 - This would mean that the beacon for that bus with Minor=2 was malfunctioning
- For Major=5, we only recorded data from Minor=1
- For Major=8, we only recorded data from Minor=1
- No data from Major=13
- For Major 18, there are extended periods where we only record data from Minor=1
- For Major 20, there are extended periods where we only record data from Minor=1