Improving the Quality and Cost Effectiveness of Multimodal Travel Behavior Data Collection

CUTR

Follow this and additional works at: https://digitalcommons.usf.edu/cutr_nctr

Recommended Citation


DOI: https://doi.org/10.5038/CUTR-NCTR-RR-2018-10

Available at: https://scholarcommons.usf.edu/cutr_nctr/13
Improving the Quality and Cost Effectiveness of Multimodal Travel Behavior Data Collection

January 2020
Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.
Multimodal transportation such as transit, bike, walk, transportation network companies (TNCs) (e.g., Uber, Lyft), car share, and bike share are vital to supporting livable communities. However, current data collection techniques for multimodal travel behavior, including apps built specifically for travel behavior surveys, have limitations (e.g., significant negative impact on battery life, user acquisition) which prevent a better understanding of significant real-world challenges (e.g., multimodal traveler choices, relationships between travel behavior and health). In this project, the research team developed and deployed a proof-of-concept system to collect multimodal travel behavior data on an ongoing basis directly from users of a popular open-source mobile app for multi-modal information, OneBusAway (OBA). To overcome battery life challenges, the research team used the Android Activity Transition API, which leverages hardware advancements in modern mobile phones. An update to the OBA app was released to 676 beta testing users. Over 10 weeks, 74 users opted into the study without any incentive and contributed 65,582 trips. Key concerns for data collection when conserving battery life are the timeliness and accuracy of data. Location data was collected for 86% of all origins and destinations. Most delays in location acquisition when starting or ending an activity were under a few minutes (e.g., 90th percentile of delay at origins was 3.2 minutes and the 68th percentile was 14 seconds). The locations for trip origins and destinations were accurate approximately to a building-level or better – the 95th percentile of estimated accuracy was approximately 48 meters. The primary cause of some low activity classification confidence values seems to be uncertainty as to when the user is walking or standing still, although further evaluation is required. The software deployed in this project is a promising new tool with a tradeoff of reduced data density for the ability to collect data from many users for longitudinal studies with little to any incentives required.
Executive Summary

Multimodal transportation options such as transit, bike, walk, transportation network companies (TNCs) (e.g., Uber, Lyft), car share, and bike share are vital to supporting livable communities. To build safe and effective multimodal infrastructure, Departments of Transportation (DOTs), Metropolitan Planning Organizations (MPOs), and transit agencies need quality data about how the public is currently traveling via these modes. However, current data collection techniques for multimodal travel behavior have limitations that restrict the ability to solve significant real-world multimodal problems. The primary method of data capture for public transit, on-board surveys, completely misses travelers that opted to use a different mode such as a transportation network company (TNC) instead of transit for a trip. In addition, any information that is captured only covers a day or two of behavior – on-board surveys do not capture longitudinal behavior. As a result, practitioners and researchers have yet to understand the precise relationship between transit and TNCs. A 2014 USDOT-UTC Pedestrian/Bicycle Workshop determined that a lack of data on when and where bicyclists travel, as well as their interactions with vehicular traffic, is one of the greatest limitations to better understanding Florida’s extremely high bicyclist and pedestrian fatality rates. The Florida Pedestrian & Bicycle Strategic Safety Plan (PBSSP) emphasizes the need for more reliable and effective data collection methods.

There have been past efforts to collect transit and bicyclist behavior data via smartphones. Apps like Cycle Tracks, Cycle Atlanta, and Strava collect bike path data from bicyclists’ smartphones. However, these systems do not collect information for transit or any other modes of transportation, including connectivity to transit. Additionally, users must start the app just for recording their trip, which is burdensome to the user and can result in user fatigue and reduced data contributions. In addition, Strava, the app with the largest number of users, does not provide trip origin-destination (O-D) data at an individual level to analysts. Other mobile apps have been designed specifically to replace travel behavior surveys, including TRAC-IT, Future Mobility Survey, Quantified Traveler, and SmarTrAC, and Florida Trip Tracker. However, these apps also suffer from user fatigue when manually recorded trips and have only been deployed in small research settings. They also do not provide an ongoing incentive or immediate value to the user for continuing to use the app. Because of user fatigue, as well as a negative impact on battery life, these apps are also typically only deployed for a short time period of several days to a week, resulting in a limited longitudinal view of travel behavior and multimodal interactions.

In this project, the research team designed, developed, and deployed a proof-of-concept system to collect multimodal travel behavior data over extended periods (e.g., weeks, months) directly from users of a popular open-source mobile app for multi-modal information, OneBusAway (OBA). OBA, maintained by the non-profit Open Transit Software Foundation, is currently deployed in over 10 cities around the world, including at Tampa, Florida, Seattle, Washington, San Diego, California, and Washington, D.C. with approximately 350,000 active users of the Android and iOS apps. Experts within the field of multimodal
travel behavior data collection and analysis were consulted to assist in the design of the data collection
process and prioritize data to be collected. This new embedded data collection tool can immediately
leverage the huge existing user base of OneBusAway that opt-in to contributing data instead of trying to
create a new user base specifically for a dedicated travel behavior data collection app.

This project also aimed to overcome some of the past battery life challenges cited by other travel behavior
data collection apps by prioritizing extremely energy-efficient data collection. This design allows the
software to collect data from users that opt-in to longitudinal studies without requiring an incentive to
compensate for the impact on their mobile device. To this end, the research team used the Android
Activity Transition Application Programming Interface (API), launched in March 2018, to conserve battery
life while still detecting when user activity changes. The Activity Transition API takes advantage of
hardware advancements in the last few generations of mobile phones where sensor co-processors have
been added to assist with energy-efficient activity recognition. As of December 2019, the Activity
Transition API supports the detection between the following activities: IN_VEHICLE, ON_BICYCLE,
RUNNING, STILL, and WALKING.

To provide additional context to a user’s travel behavior, the application also saves information about the
user’s interactions with the application (viewing arrival times, planning a trip, and setting a destination
reminder) that are attached to an activity transition when it occurs. Google’s Firebase Cloud Firestore was
chosen for storing activity transition, location, and app usage data on the local device, which also fulfills
the role of sending the data to a server for analysis. The user’s email address and mapping to their travel
behavior universally unique identifier from the Firebase data is stored separately in a secure Google Sheet
to allow follow-up online surveys while protecting user privacy. The travel behavior data collection
software, which consisted of approximately 3,200 new lines of code, was implemented by the research
team and merged into the publicly available GitHub repository for the OneBusAway Android app.

To process and visualize the collected data, the research team created a methodology to convert from
an activity transition data model, where each data record consists of a transition from one mode to
another at a single location, to a more traditional origin-destination (OD) model, where each record
consists of both an origin and destination location with the mode used to travel between them. This
work also included developing an algorithm to select the “best” location associated with the activity
transition. The team created a desktop application that retrieves the data from Firestore, processes it
according to the defined conversion rules, and exports to comma-separated value (CSV) for a tabular
data format and Keyhole Markup Language (KML) for a spatial data format. The Java application is
approximately 5,200 lines of code and is publicly available GitHub.

The data collection software developed in this project was released as an update to the OneBusAway
app to a beta testing group of 676 users on July 11, 2019. No incentives were provided to users for
opting into the study (i.e., contributing their travel behavior data). No users withdrew from the study.
during this period, and no users reported any negative consequences (e.g., reduced battery life) from participating in the study.

Data collected from 74 users from the beta testing group over about 10 weeks was examined during this project. This data consisted of 65,582 trips representing detected user activity (WALKING, RUNNING, IN_VEHICLE, ON_BICYCLE, or WALKING/RUNNING) from an origin to a destination. A very large percentage (about 73%) of detected trips are walking trips, likely because the Activity Transition API can detect very short walking trips within a building (e.g., from an office to a conference room). When filtering out any trips that had a duration less than 5 minutes or had a distance between an origin and destination of less than 50 meters, 13,698 trips remained, with a lower share of walking trips (a drop of 73% to 43%) but a higher share of vehicle trips (an increase from 21% to 48% of trips).

Activity classification confidence values, or Google’s measure for the likelihood a user is performing the detected activity on a scale from zero to 100, collected from the Android Activity Recognition API should be further analyzed. When filtering out trips less than 5 minutes in duration or 50 meters in distance, 50% of activities following transitions had a confidence greater than 92, 68% of transitions have confidence greater than 69, and 75% of activities following transitions have a confidence greater than 31. The role of false positives in detecting WALKING activity while stationary, which seems to be a cause of some of the low confidence values, was examined. The bias of the Transition API towards triggering false positives rather than false negatives (i.e., missed trips) should be a benefit to the research. Further analysis and filtering of false positives can be implemented — however, if false negatives were to occur, the research team would have a much harder time reconstructing the travel behavior information because no data during that time period would be recorded. The only false negatives observed by the research team occurred when power savings mode was activated on an older device (Samsung Galaxy S6, SAMSUNG-SM-G920A, with Android 7.0). Older devices that do not have newer sensor co-processor hardware to assist in saving power while monitoring activity transitions may especially be prone to stopping all data collection when battery saver mode is active. Research based on this data will need to determine what impact the lack of these older devices may have on the study population. Additional analytics capabilities were added to the app to track if users have enabled battery saver mode or have otherwise changed settings to ignore battery optimizations.

The location data for each trip origin and destination was also examined. Location data was collected for 86% of all origins and destinations of trips—the remaining 14% of locations could not be acquired due to user settings turning off location or restricting location permissions. The algorithm the research team developed to choose the “best” location for each trip origin and destination selected the GPS provider location for around 26% of transitions, the fused location provider location for around 39% of transitions, and a network-based location for around 21% of transitions. The delays between detecting an activity transition and acquiring a user position were reasonable, especially given the low-energy tradeoffs of the data collection software used in this research. While there are some outliers with large acquisition times (likely due to the phone being in deep sleep and not actively monitoring location
sensors to save energy), the majority of time delays are under a few minutes. The 90th percentile of delay at origins was 3.2 minutes and the 68th percentile is even lower at 14 seconds, with destinations having even less delay, likely because the phone was recently woken up by the user during travel. The locations for each trip origin and destination were accurate approximately to a building-level or better, with the 95th percentile of estimated accuracy at both origins and destinations of trips being about 48 meters. This level of accuracy should be enough for most general origin-destination analysis.

Travel behavior data collection within the OneBusAway app will continue beyond the end of this research project, with the anticipation of adding in-app and on-line survey in future research to collect additional information and validate passively collected data. Users continue to enroll at a rapid pace, especially given that no incentive is being offered. As of January 23, 2020, 105 users from the beta testing group of 740 users had enrolled in the study because of a single in-app prompt to “learn more” about the research – over 14% of the beta user base. No users have withdrawn from the study, and no users have reported any negative consequences (e.g., reduced battery life) from being enrolled in the study. Clearly, users are willing to contribute travel behavior as part of a longitudinal research study with little to no incentives if there is not a noticeable negative impact on their device. Given that the data collection software has proven to be successful in the deployment to the beta test user group in this study, the software update will be rolled out to all OneBusAway Android users.

In summary, the proof-of-concept travel behavior data collection software deployed in this project is a promising technology with a fundamental tradeoff of reduced data density (i.e., collecting origin and destination locations instead of a breadcrumb travel path) for a reduced impact on battery life and the ability to collect data from a large number of users for longitudinal studies with few incentives required. The data acquired during the study appears to be viable for various types of travel behavior research (e.g., origin/destination, health, route choice), although additional data processing will be needed to filter data for specific types of research and to validate the data at a larger scale than was possible in this study (e.g., via user feedback). It should be noted that additional location data collection (e.g., breadcrumb travel paths) could be selectively added to the app, but it comes at a cost to device battery life and therefore should be only added if vital to a study.
1 Introduction

Multimodal transportation options such as transit, bike, walk, transportation network companies (TNCs) (e.g., Uber, Lyft), car share, and bike share are vital to supporting livable communities. To build safe and effective multimodal infrastructure, Departments of Transportation (DOTs), Metropolitan Planning Organizations (MPOs), and transit agencies need quality data about how the public is currently traveling via these modes. However, current data collection techniques for multimodal travel behavior have limitations that restrict the ability to solve significant real-world multimodal problems.

One example area lacking robust multimodal data is the relationship between TNCs and public transit. Some see TNCs as a competition to public transit that is primarily responsible for trends of dropping transit ridership. Others see TNCs as vital first/last mile service that is a complement to public transportation. Some agencies, such as Pinellas Suncoast Transit Authority (PSTA) in Tampa, FL, have formed partnerships with a TNC (e.g., Uber) in order to help fill the first/last mile gap. Other agencies such as Hillsborough Area Regional Transit (HART) have operated their own TNC-like service (in the cases of HART, via a contractor Transdev). However, it can be extremely difficult to capture travel behavior data that includes information about the interactions between TNC use and public transit – as of today, there is little hard data that includes origin/destination trip data for transit and TNCs, especially when one mode is used in place of another. The primary method of capture, on-board surveys, would completely miss travelers that opted to use a TNC instead of transit for a trip. In addition, any information that is captured only covers a day or two of behavior – on-board surveys do not capture longitudinal behavior that could reveal different choices under different circumstances. As a result, practitioners and researchers have yet to understand the precise relationship between the two modes as well as other modes such as walking or bicycling.

Transit rider personal safety and travel patterns also are not adequately captured using existing data collection tools such as on-board surveys. A 2014 report funded by the Florida Department of Transportation in collaboration with Florida’s Transit Safety Network (FTSN) determined that while bus operator assaults are significant in Florida, assaults on transit vehicle riders were even more significant both in terms of the number of riders injured and the rate at which these assaults were occurring [1]. The report also stated that more data is needed to better understand this issue. Current public transportation on-board surveys may occur face-to-face with riders at a single location during the day and may not adequately capture how a rider feels about safety at other locations that they visit at different times of day (e.g., night). As a result, a transit agency may not have adequate data to respond proactively to concerns about safety on their system based on quantitative or qualitative information linked to specific bus stops or routes.

Collecting travel behavior data from bicyclists is also challenging. For example, a 2014 USDOT-UTC Pedestrian/Bicycle Workshop [2] determined that a lack of data on when and where bicyclists travel, as well as their interactions with vehicular traffic, is one of the greatest limitations to better understanding
Florida’s high bicyclist and pedestrian fatality rates. As a result, one of the greatest research needs identified within this workshop was to develop better tools to collect data from multimodal travelers. The Florida Pedestrian & Bicycle Strategic Safety Plan (PBSSP) also emphasizes the need for more reliable and effective data collection methods [3].

There have been past efforts to collect bicyclist behavior data via smartphones. San Francisco created an open-source project Cycle Tracks [4], a mobile app which was used to collect bike path data from bicyclists’ smartphones. A similar project, Cycle Atlanta [5], was implemented in Atlanta, GA, and was based on the Cycle Tracks open-source code. Strava, a company that offers a recreational bicycling activity-recording app, offers bike path data collected by its users to FDOT and city planners for a fixed period for a fee. However, all these systems only collect bike path data from bicyclists – they do not collect information for transit or any other modes of transportation, including connectivity to transit. Additionally, users must start the app just for the purpose of recording their trip, which is burdensome to the user and can result in fatigued and reduced data contributions [6]. In addition, Strava, the app with the largest number of users, does not provide trip origin-destination (O-D) data at an individual level to analysts.

Other mobile apps have been designed specifically to replace travel behavior surveys, including TRAC-IT [6], Future Mobility Survey [7], Quantified Traveler [8], and SmarTrAC [9], and Florida Trip Tracker [10]. However, these apps also suffer from user fatigue from manually recorded trips and have only been deployed in small research settings. They also do not provide an ongoing incentive or immediate value to the user for continuing to use the app. Because of user fatigue, as well as a negative impact on battery life, these apps are also typically only deployed for a short time period of several days to a week, resulting in a limited view of travel behavior that longitudinal data might reveal.

Some naturalistic bike studies, such as those conducted by Virginia Tech [11] and USF [12], focus on outfitting bicycles with additional equipment such as on-board computers and cameras to collect more data about a bicyclist’s first-hand experience. However, these systems can be very costly to deploy on a large scale, and logistically difficult to manage, as the equipment must be installed on and later retrieved from participant bikes. While mobile apps yield less data per-user than the equipment used in Naturalistic Bike Studies (e.g., no recorded video), apps are very cost-effective to deploy on a large scale as they take advantage of devices already owned by many multimodal travelers: mobile phones. Apps can also be deployed indefinitely, while other equipment must eventually be collected.

1.1 Purpose and Benefits of Research

In this project the research team designed, developed, and deployed a proof-of-concept system to collect multimodal travel behavior data on an ongoing basis directly from users of a popular open-source mobile app for multi-modal information, OneBusAway (OBA) to address many of the aforementioned limitations. OBA, maintained by the non-profit Open Transit Software Foundation, is currently deployed in over 10 cities around the world, including at Tampa, Florida, Seattle, Washington, San Diego, California, and
Washington, D.C. with approximately 350,000 active users of the Android and iOS apps. The purpose of this research is to increase quality and cost-effectiveness of multimodal travel behavior data collection. To address the perceived user value and overcome user fatigue, this data collection tool can immediately leverage the huge existing user base of OneBusAway instead of trying to create a new user base specifically for a dedicated travel behavior data collection app. Better data will assist planners in understanding how and where people are traveling via non-single occupancy vehicle (SOV) modes, which will enable DOTs, MPOs, and transit agencies to better prioritize infrastructure investments and make operational improvements.

In this project, open-source software was integrated into the OneBusAway Android to enable multimodal activity data, including trip origin and destination, transfers, trip travel path, and travel time for users that opt-in. Data shows that the OneBusAway mobile app is very “sticky” and retains approximately 67% of users that download the app [13], compared to an average retention rate of 20% for other mobile apps [14]. Users have an incentive (real-time transit information) to continue to use the mobile app, and therefore data collection can be ongoing for months or even years. Additionally, as new regions continue to deploy the open-source OneBusAway mobile app, travelers in those regions will become potential data contributors. For example, during the project Spokane (WA) Transit Authority launched OneBusAway, providing an opportunity to learn more about travelers in Spokane.

The versatility of this tool and enormous potential audience of data contributors will enable new research opportunities and many different types of travel behavior studies well beyond the end of this project. Travel behavior data is passively collected from travelers via various technologies in the mobile device (e.g., GPS, Wi-Fi, cellular, accelerometers). This data can yield information about trip origins and destinations, dwell and travel times, and even mode of transportation and purpose information via machine learning and data mining techniques.

This research will assist regions in collecting multimodal transportation data to better understand how and where people are traveling via public transportation, bike, and pedestrian modes, and will help inform DOTs, MPOs, and transit agencies in multimodal infrastructure and program investments, including improvements to bike and pedestrian facilities to access transit and transit rider safety. This research will contribute towards meeting the goals and objectives related to Data Analysis and Evaluation portion of the Florida Pedestrian & Bicycle Strategic Safety Plan [3].
2 System Design and Implementation

2.1 Key challenges

The system to implement travel behavior data collection must address two primary challenges:

1. How to capture activity data (e.g., start and end time, position) on a mobile device
2. How to transfer this data to a server where it can be examined

Both tasks must be implemented in a way to avoid any potential negative side effects on the user’s device. Any negative consequence could result in the user leaving the study.

2.1.1 Challenges detecting user activities

Past travel behavior data collection apps have cited battery life as a key limitation in deployments [6, 9]. Data collection apps typically use global navigation satellite system (GNSS) hardware of mobile devices to continuously calculate travel path information. While GNSS position information can be very accurate while traveling outdoors, it can have difficulty acquiring a position fix in highly obstructed environments such as indoors. Other positioning technologies such as Wi-Fi and cell network technologies can help locate users indoors with varying precision depending on the density of the network access points.

However, all these positioning technologies, especially GNSS hardware, negatively affect battery life [15]. For example, if GNSS hardware is constantly active a mobile device battery will be depleted in a matter of hours. And, battery life has continuously been cited as the top priority of mobile phone users [16]. Because users care about battery life so much, any noticeable impact by an app on battery life must have a justifiable benefit to the user. Therefore, it is critical to manage the tradeoff between the frequency of position updates and the impact on end user battery life. If users notice a difference in battery life when using an application that collects travel behavior data, they are likely to uninstall it.

2.1.2 Challenges sending user data to a server

After a user activity is detected on a mobile device, this data must be sent to a server so researchers and analysts can examine it. The transmission of this data can also negatively impact users if not appropriately managed [15]. For example, data must be sent using the cellular data or Wi-Fi hardware of the mobile device, which consumes battery energy. Additionally, if the data is sent over a cellular connection and the user does not have an unlimited data plan and exceeds their monthly quota, the user may incur data fees. Finally, security of the user’s data is critical – travel behavior could potentially contain data for sensitive locations that the user would not want others to know about. Therefore, an essential design consideration is the intelligent management of user data transmissions to a server in a secure manner.
2.2 Data collection

This project aimed to overcome some of the past battery life challenges cited by other travel behavior data collection apps by prioritizing extremely energy-efficient data collection over other considerations. This design allows the software to collect data from users that opt-into longitudinal studies without requiring an incentive to compensate for the impact on their mobile device. The following two sections discuss the design and implementation of the software to detect user activities and send this data to a server so the research team can access it. Experts within the field of multimodal travel behavior data collection and analysis were consulted to assist in the design of the data collection process and prioritize data to be collected, including:

- Dr. Candace Brakewood, Assistant Professor, University of Tennessee-Knoxville
- Dr. Achille Fonzone, Associate Professor, Edinburgh Napier University
- Dr. Jan-Dirk Schmoecker, Associate Professor, Kyoto University
- Dr. Kari Watkins, Associate Professor, Georgia Institute of Technology

2.2.1 Detecting user activities and related metadata

2.2.1.1 Activity Transition API

Google launched the Android Activity Transition Application Programming Interface (API) in March 2018 to help developers conserve battery life while still detecting when the user activity changes [17]https://android-developers.googleblog.com/2018/03/activity-recognitions-new-transition.html. This new activity transition technology takes advance of hardware advancements in modern mobile phones where sensor co-processors have been added to assist with energy-efficient activity recognition.

In previous hardware designs, the main application processor was periodically turned on to process sensor data or calculate the user’s position to detect a change in activity, which resulted in a large energy cost (Figure 1).
In the new hardware design, a sensor co-processor, which consumes a fraction of the energy required by the main application processor, stays active to monitor and analyze input from on-board sensors such as accelerometers and a gyroscope. When the sensor co-processor detects motion that seems to represent a new activity (on devices that have a Sensor.TYPE_SIGNIFICANT_MOTION sensor [18]), it wakes up the main application processor to perform an action such as calculating the user’s position and saving the data (Figure 2).
Because the application processor lays dormant most of the time and is only activity when an actual activity transition occurs, this design saves a significant amount of battery energy. This also means that the OneBusAway app receives transitions even when the user is not actively using the app without the app constantly running in the background. For these reasons, the research team chose to use the Android Activity Transition API to detect user activity transitions.

One limitation of this approach is that the application monitoring user activity relies entirely on the sensor co-processor for activity detection and classification, and as a result is limited to detecting only the activities supported by the sensor co-processor. As of December 2019, the Android Activity Transition API supports detecting the transition between the following activities [19]:

- IN_VEHICLE
- ON_BICYCLE
- RUNNING
- STILL
- WALKING

Activity Transition API also doesn’t directly provide confidence values for detected activity types, so the research team used the Activity Recognition API immediately after the Activity Transition API to collect the confidence value for the detected activity type [20]. Confidence is defined by Google as “a value from 0 to 100 indicating the likelihood that the user is performing this activity.”

To further save energy, the sensor subsystem may not immediately wake up the application process exactly when the activity transition occurs. Additionally, a location may not be immediately determined by the application processor as soon as it wakes up (e.g., if it needs to wait for a GPS location to be acquired). As a result, there may be a delay between the time at which the activity transition happens and the timestamp of the calculated location. Analysis later in this report will focus on determining the length of this potential delay based on user data.

2.2.1.2 Epoch time calculation

Time on computers is traditionally represented as the difference, measured in milliseconds, between the current time and midnight, January 1, 1970 UTC (coordinated universal time), which is also referred to as the “epoch”. Because the activity transition event is calculated in low-level hardware, the timestamp captured by the hardware is not an absolute time, but instead is captured as “nanoseconds since boot” of the device. This value is a simple counter that may start at one when the device is powered on and increment by one as each nanosecond passes.

Therefore, the research team had to calculate the exact time at which an activity transition occurred, or transition\_mills\_since\_epoch, by capturing the time at which the application is woken up by the transition event.
in both nanoseconds since boot as well as milliseconds since the epoch, or \( \text{wakeup}_{\text{nanos since boot}} \) and \( \text{wakeup}_{\text{millis since epoch}} \), respectively. Figure 3 illustrates these timestamp values in context of a timeline.

\[
\text{transition}_{\text{nanos since boot}} = \? \\
\text{wakeup}_{\text{nanos since boot}} \\
\text{transition}_{\text{millis since epoch}} = \? \\
\text{wakeup}_{\text{millis since epoch}}
\]

*Figure 3 – Several timestamps must be captured to calculate the exact activity transition time (\( \text{transition}_{\text{millis since epoch}} \)).*

The \( \text{transition}_{\text{millis since epoch}} \) can then be calculated using the following equation:

\[
\text{diff}_{\text{nanos}} = \text{wakeup}_{\text{nanos since boot}} - \text{transition}_{\text{nanos since boot}}
\]

\[
\text{transition}_{\text{millis since epoch}} = \text{wakeup}_{\text{millis since epoch}} - \text{diff}_{\text{nanos}}
\]

### 2.2.1.3 Device location

When the application is woken up and provided the activity transition event, it can request the device’s position from the Android Location API. Android currently supports three different location providers from which a location can be determined:

1. **GNSS** – Uses global navigation satellite systems (GNSS) such as the U.S. global positioning system (GPS) (Android uses the label “GPS” for this type of data, but it includes positions calculated via multiple GNSS). GNSS typically does not work well indoors and can have a delay while trying to calculate a position but is the most accurate positioning technology outdoors.

2. **Network** – Uses the network to which the device is connected, either from a database of Wi-Fi access point locations or cellular tower positions. Wi-Fi network positions can be very accurate (e.g., within the building) and fast to compute when the user is indoors and the Wi-Fi access point location database is of high quality, but not available when the user is outside. Positions based on the cellular network have a far less accuracy, typically in the hundreds of meters.
3. Fused – Uses a hybrid of GNSS, network, and on-board sensors to determine the users position [21]. Attempts to providing a fast, accurate location whether the device is indoors or outdoors.

After each event transition, the research team chose to request new position calculations from all three of these location providers and store all this information. This allows for further post-processing and analysis to determine which of the positions is “best” for each given activity transition event. Android PendingIntents, which allows the application to stop running while the locations are computed, were used for energy-efficiency.

2.2.1.4 Caching app usage data

To provide additional context to a user’s travel behavior, the application also saves information about the user’s interactions with the application that are later attached to an activity transition when it occurs. Knowing what arrival information the user saw clarifies why they did or did not use public transportation for a trip. Similarly, knowing the trip that the user planned prior to traveling gives an indication of what information they had on a variety of modes at the time (currently walking, transit and bike share) as well as their intended destination. Knowing that the user has chosen to be reminded at a destination stop gives an explicit indication of which transit route, trip, and vehicle they boarded, as well as the exact stop that they exited public transportation.

Figure 4 shows screenshots from the application when viewing arrival times, planning a trip, and setting a destination reminder. The raw data behind each of these screens, along with the user’s location at the time, is saved to a temporary private cache on-board the device as JavaScript Object Notation (JSON) files when the user interacts with these screens. These cached JSON files are then read when an activity transition occurs, as discussed in the next section.
2.2.1.5 Summary of process

Google’s Firebase Cloud Firestore [22] was chosen as the method of persisting activity transition and location data on the local device, which is discussed in detail in the following section.

After the user enrolls in the study (which is discussed in a later section), the application registers for activity transition events, which occur whether or not the application is actively running. When the activity transition occurs, the application retrieves the locally cached app usage metadata and if the data has been cached within the last 30 minutes it is saved along with the activity transition information in Firebase. The locally cached usage data is then permanently deleted. Device information such as the application version, device model, Android version, and if accessibility features are enabled are captured via Firebase whenever these properties change.

Much of the data collection process takes place in parallel upon the app receiving the initial activity transition data in an asynchronous manner so that each task can finish on its own timeline without requiring the application to continuously run. Figure 5 illustrates the data collection process in its entirety, starting with the initial registration for activity transition the first time the app runs, followed by the sequence of events that occur when the user performs each activity transition.
The travel behavior data collection software following this design, which consisted of approximately 3,200 new lines of code, was implemented by the research team and merged into the publicly available GitHub repository for the OneBusAway Android app [23].

### 2.2.2 Sending user data to a server

Google’s Firebase Cloud Firestore [22] was chosen for storing activity transition, location, and app usage data on the local device, which also fulfills the role of sending the data to a server for analysis. Firestore is a NoSQL data storage solution that provides a secure, object-oriented data store for “documents”, or a set of nested fields, and “collections”, or groups of documents. Firestore data is also automatically synchronized with Google’s cloud services, and therefore uploads data when it is efficient for the device to do so (e.g., when the device is plugged in at night, when on Wi-Fi instead of a cellular network, when batched with other network communication). As a result, Firestore assumes the overhead of managing network communications in an energy efficient way and reduces this burden on the application. Because Firestore is No-SQL, it adapts well to the various object-oriented data structures being stored, especially because some fields may not be available under certain conditions. Figure 6 shows the view of activity transition data in the Firebase web console.
Each user is assigned an anonymous universally unique identifier (UUID) and a Firestore document is created with this UUID as its name. Within this document, a Firestore collection is then created for each of the four data types collected for this project (activity transitions, arrivals and departures, device information, trip plans). The individual data records (e.g., each activity transition) is then stored as a document within each collection, respectively. This nested structure can be seen in Figure 6. Each activity transition is also given its own UUID, which is also prefixed with an integer counter to make it easier to order records for debugging purposes. The activity transition UUID is then used to name the other related data documents (e.g., viewed arrivals and departures, trip plans, destination reminders) for that activity transition so each data type can be cross-referenced later. Device information documents are named with the timestamp at which that device information is recorded. If any of the device information changes (e.g., an Android platform version update, a new OneBusAway app version), a new device information document is created with the timestamp at which the new information was detected. No user-identifying information is stored in Firestore.

The research team also wanted to be able to survey users about their travel behavior and be able to relate this survey information to the activity transition data collected within OneBusAway. The research team decided to ask for the user’s email address during enrollment (discussed in the next section) to enable outreach for on-line surveys. A Google Sheet was used to securely store the mapping between the user’s email address and UUID. This separation of storage, shown in Figure 7, ensures that the user’s email is not directly associated with their activity transition or app usage data. As a result, the user’s email is not stored within the OneBusAway app on their device or in Firestore.
The software that synchronizes Firestore with the server and sends the email-to-UUID mapping to a separate data store is part of the previously mentioned approximately 3,200 new lines of code publicly available GitHub repository for the OneBusAway Android app [23].

### 2.2.3 User interface for participant enrollment

After the methods to collect user data and transfer it to a server were identified, a user interface needed to be designed and implemented to allow users to opt-into the study. The research team drafted a design for the user enrollment process, as well as a description of the project and an informed consent document, to the USF Institutional Review Board for review. This IRB application is included in Appendix A and the informed consent document is included in Appendix B. The design of the experiment was also presented at a meeting of the Open Transit Software Foundation Board of Directors for feedback on the research. The diagram describing the user enrollment that was submitted to the USF IRB is shown in Figure 8.
When the user first installs the OneBusAway app, or after the update is installed that includes the software developed during this project, the user is asked if they would like to help improve public transportation. If they answer in the affirmative, they are asked if they are over 18. If they again answer in the affirmative, they are shown informed consent text for the research study. If they agree to the informed consent, they are asked for their email address for follow-up surveys, and then the activity transition data collection software registers with the Android platform to receive activity transition events. If the user declines any of these dialog options, the user is returned to the main screen of the application and no data is collected.
USF IRB determined that this research was exempt from IRB review. As a result, the research team followed the Accreditation of Human Research Protection Programs (AAHRPP) guidelines and included a shorter version of the informed consent document within the OneBusAway mobile app, which is included in Appendix C.

2.3 Data processing and visualization

While Firestore serves as a convenient way to store data locally on a device and synchronize this information with a server, Google does not provide an efficient way to view or query this data for analysis. The Firebase dashboard shown in Figure 6 is effective for browsing data for debugging purposes, but to organize user data in a more traditional “origin-destination” format to represent their trip from one location to another requires further processing.

2.3.1 Activity transition vs trip origin-destination models

As part of this data export process, trips must be constructed from activity transitions. In other words, we know where and when a user stopped one activity and started another – we must convert this into discrete trips from an origin to a destination.

Figure 9 shows the activity transition data representation stored in Firestore, where each circle is one data record that includes the activity the user exited and the activity the user entered at a location.

![Activity transition data representation stored in Firestore](Figure 9)

Figure 10 shows this same information, but represented in a trip “origin-destination” (OD) format, where each data record is a combination of two circles (the origin and destination locations) as well as the information about the activity used to travel between them (i.e., the mode of transportation).
To transfer the activity transition model to the O-D format, two activity transition records must be linked together to obtain the location and time information for the origin and destination.

Following are some considerations when translating between the two data representations:

- **Fault tolerance** - Because activity transition events are not guaranteed to be delivered to the application for each user transition (e.g., if the device battery dies), some error tolerance is required when processing the data. For example, the application could receive an activity transition event indicating that a user stopped being still and started walking, but if they phone battery dies while they are walking the application will never get an event that signals that the user stopped walking.

- **Location selection** - As mentioned earlier, for each activity transition we can have up to three recorded locations from the GNSS provider, network provider, and fused provider. Each of these locations’ latitudes and longitudes, timestamps, and accuracy uncertainty values may differ. Therefore, the “best” location for each activity transition must be determined.

- **Event consolidation** – The activity transition data can be noisy, with multiple transitions between events happening in relatively short order. This can happen when the user is sitting relatively still but slightly moving, either within a vehicle or in a chair. Activity transition data is also received separately for walking and running events. If a person partially runs and partially walks from an origin to a destination, this should not be considered multiple trips.
The following rules were developed to assist in converting from the activity transition model to the O-D model:

- To avoid errors propagating across long time periods (e.g., accidentally grouping an ENTER event from earlier in the day with an EXIT event the next day), activity transitions are split into groups of one “day” of travel behavior. Days are split at 3am by default, with each 3am to 3am set of events processed separately.
- Ordering activity transitions - Activity transitions are ordered by the activity transition event epoch time. If that time cannot be obtained, then the “best” location time is used.
- Trip start and end times - Activity ENTER event timestamps define the start time of a trip at the origin location, and activity EXIT event timestamps define the end time of a trip at the destination location.
- Choosing the best location for each activity transition event – Sort all the locations in temporal order of the location calculation time. Use the earliest location with an estimated accuracy within an accuracy threshold (default = 50 meters, lower numbers are better). If no locations are within the accuracy threshold, then use the one with the best accuracy. Accuracy is determined by the mobile device when the location is acquired and is defined as “the radius of 68% confidence” by the Android specification [24].
- Merging events - Every time a transition record is added to the same day record list, we look if the previous record is a still event and the two previous event is the same event as this event and if the still event’s duration is less than the threshold (default = 2 minutes) we remove the middle still event and merge two same (e.g., IN_VEHICLE, IN_VEHICLE) or similar events (e.g., WALKING, WALKING/RUNNING) events into a single event.
- Tours – To further group sets of trips, we define a “tour” as a set of trips starting at one location and ending at the same location. For example, the trips “home to work” and “work to home” would be considered a tour, with home being the anchor location. A threshold of 50 meters is used to determine if two locations are the “same”.

2.3.2 Data output formats

The research team wanted to be able to visualize trip O-D information in both tabular and spatial data formats.

The comma-separated value (CSV) format was selected as the tabular format because of its support and ease of use in many data processing tools. Figure 11 shows example trip origin-destination data that has been formatted in CSV and then opened in Microsoft Excel.
### Example 1 - A transit journey (e.g., home to work) that consists of (walking > transit > walking). User waited for the bus about 8 minutes (duration between end of trip 1 and start of trip 2).

<table>
<thead>
<tr>
<th>User ID</th>
<th>Trip ID</th>
<th>Google Activity</th>
<th>Google Activity Confidence</th>
<th>Start Date</th>
<th>Start Time</th>
<th>Start Latitude</th>
<th>Start Longitude</th>
<th>End Date</th>
<th>End Time</th>
<th>End Latitude</th>
<th>End Longitude</th>
<th>Activity Duration (minutes)</th>
<th>Tour ID (these “tours” define a group of trips that starts and ends at the same location)</th>
<th>Tour Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ss233</td>
<td>1</td>
<td>WALKING/RUNNING</td>
<td>0.9</td>
<td>7/12/2019</td>
<td>7:56:39 AM</td>
<td>26.393348</td>
<td>-97.2373743</td>
<td>7/12/2019</td>
<td>8:10:12 AM</td>
<td>-97.237983</td>
<td>13.56</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ss283</td>
<td>2</td>
<td>IN_VEHICLE</td>
<td>0.8</td>
<td>7/12/2019</td>
<td>8:18:56 AM</td>
<td>26.3888385</td>
<td>-97.237985</td>
<td>7/12/2019</td>
<td>9:02:22 AM</td>
<td>-87.237831</td>
<td>42.7</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Ss283</td>
<td>3</td>
<td>WALKING</td>
<td>0.9</td>
<td>7/12/2019</td>
<td>9:03:01 AM</td>
<td>26.389954</td>
<td>-97.237382</td>
<td>7/12/2019</td>
<td>9:05:01 AM</td>
<td>-87.237838</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

### Example 2 - An in-car journey (e.g., work to home) that consists of (walking > car > walking).

<table>
<thead>
<tr>
<th>User ID</th>
<th>Trip ID</th>
<th>Google Activity</th>
<th>Google Activity Confidence</th>
<th>Start Date</th>
<th>Start Time</th>
<th>Start Latitude</th>
<th>Start Longitude</th>
<th>End Date</th>
<th>End Time</th>
<th>End Latitude</th>
<th>End Longitude</th>
<th>Activity Duration (minutes)</th>
<th>Tour ID (these “tours” define a group of trips that starts and ends at the same location)</th>
<th>Tour Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ss233</td>
<td>4</td>
<td>WALKING</td>
<td>0.9</td>
<td>7/12/2019</td>
<td>6:00:01 PM</td>
<td>26.389954</td>
<td>-97.237383</td>
<td>7/12/2019</td>
<td>6:01:01 PM</td>
<td>-97.237332</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Ss283</td>
<td>5</td>
<td>IN_VEHICLE</td>
<td>0.9</td>
<td>7/12/2019</td>
<td>6:01:03 PM</td>
<td>26.3888368</td>
<td>-97.237831</td>
<td>7/12/2019</td>
<td>6:46:03 PM</td>
<td>-87.237986</td>
<td>45</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Ss283</td>
<td>6</td>
<td>RUNNING</td>
<td>0.9</td>
<td>7/12/2019</td>
<td>6:45:03 PM</td>
<td>26.389369</td>
<td>-97.237563</td>
<td>7/12/2019</td>
<td>6:46:33 PM</td>
<td>-87.2373743</td>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 11 - Example trip origin-destination data from a CSV file export in a tabular format*
To visualize the data spatially, the research team wanted to avoid re-creating visualization tools and instead focus on exporting the data in standardized formats that could be used in existing tools. The research team chose the Keyhole Markup Language (KML) [25] as the data export format. KML can be visualized in tools such as Google Earth [26]. Figure 12 shows a short walking trip from a parking lot to a building that was exported to the KML file format and then loaded into Google Earth.

![Figure 12 - A short walking trip exported to KML and visualized in Google Earth, with the actual start and end locations overlaid](image)

The green triangle represents the origin for the trip where the location for the activity transition from vehicle to walking was calculated, a purple line connects the origin and destination, and a red circle shows the calculated location for the activity transition from walking to still at the end of the trip.

Clicking on the origin or destination marker shows more information about that location, including the starting time, confidence level of activity recognition (“97” in Figure 12), the difference in time between the detection of the activity transition and the acquisition of the location data, the activity duration, the location provider (GPS/GNSS, network, or fused), and the estimated horizontal accuracy of the location data.
In Figure 12, the actual user start and end locations for the trip have also been overlaid on the Google Earth screenshot – these icons do not appear in the output from the tool (as they are unknown for study participants), but are shown here to demonstrate how the calculated and actual locations for trip origins and destinations can relate to one another. The blue solid triangle represents the actual location where the user parked their car (i.e., started the walking trip), and the blue circle represents the actual walking trip end location when the user sat down.

The information for the calculated locations in the white boxes indicates that the origin location was calculating using GPS/GNSS and has a slightly better estimated accuracy of 9.65 meters when compared to the destination location, which was calculated using the Wi-Fi network with an estimated accuracy of 21.89 meters. However, it is also important to note the potential delay between the activity transition occurring and the location being calculated. In this case, at the start of the walking trip even though the GPS location for the green triangle start location is estimated to have better accuracy, the slight delay between when the user started the activity and the position was acquired (0.22 minutes, or 12 seconds) places the calculated location further from the actual location, when compared to the end location. The distance between the actual start location and the calculated start location is the distance the user could walk in 12 seconds. The calculated end location using Wi-Fi was estimated to be slightly less accurate than the calculated origin location, but because the user stopped moving and was indoors, it ends up being closer to the actual end location than the calculated start location was to the actual start location.

In summary, the Google Earth visualization demonstrates that the distance between the actual and calculated locations at the origin and destination of the trip is a function of both the accuracy of the position as well as the delay between the activity transition and the location acquisition. The issue of delays between activity transition and location acquisition, as well as estimated location calculation accuracies, is examined in detail later in this report.

### 2.3.3 Implementation

The desktop application that retrieves the data from Firestore, processing it according to the defined rules, and exports to the KML and CSV file formats was implemented for this project. The Java application is approximately 5,200 lines of code and is publicly available GitHub [27].
3 Deployment and Analysis

The data collection software developed in this project was released as an update to the OneBusAway app to a beta testing group of 676 users on July 11, 2019. No incentives were provided to users for opting into the study (i.e., contributing their travel behavior data) as part of this deployment. As of September 19, 2019, 74 users had enrolled in the study. No users withdrew from the study during this period, and no users reported any negative consequences (e.g., reduced battery life) from being enrolled in the study.

The analysis of deployment data in the sections that follow is for the period of July 11, 2019 to September 19, 2019 (about 10 weeks) from these 74 users. The Google Cloud Platform console indicated that 1.71 GB of data had been collected and stored in Firestore during this period. The travel behavior data analysis tool described in previous sections was used to export the data from Firestore to KML and CSV files on September 19, 2019. The following default settings were used:

- Still event merge threshold – 2 minutes
- Walking/running event merge threshold – 2 minutes
- Day start time – 3am

3.1 Trip data

The resulting CSV file contained 113,903 trip records. Figure 14 shows the breakdown of the contributed activities by region, while Figure 13 shows the breakdown of participating users based on their region.

![Region participation by number of contributed activities](image)
Both graphs are very similar and mirror the general breakdown of all users by region.

For this initial data export, “STILL” records with the same origin and destination (e.g., generated in between WALKING->STILL and STILL->WALKING transitions) were included to assist with debugging and analysis (i.e., to confirm that the activity to trip OD model transition was working correctly). In traditional trip OD analysis, these STILL records would be excluded, as they are not traditional trips from one location to another.

48,321 “STILL” trip records were included in the data – therefore, the remaining 65,582 trips represented detected user activity (WALKING, RUNNING, WALKING/RUNNING, IN_VEHICLE, ON_BICYCLE) from an origin to a destination. Figure 15 shows the breakdown of trips by activity types.
A large percentage (about 73%) of detected trips are walking trips. This is likely because the Activity Transition API can detect very short walking trips within a building (e.g., from an office to a conference room).

To take a closer look at longer trips that would be examined in traditional origin and destination studies, the research team filtered out any trips that had a duration less than 5 minutes or had a distance between an origin and destination of less than 50 meters. Other studies may want to use other filters (e.g., health studies may be interested in very fine-grained activity time duration). The resulting breakdown in trips by activity type for the remaining 13,698 trips are shown in Figure 16.
When looking at these longer trips the percentage of vehicle vs walking trips increases significantly – from 21% of trips to 48% of trips. Conversely, the share of walking trips drops from 73% to 43%.

Additional thresholds could be added to the data analysis tool to filter out other trip profiles depending on the type of travel behavior that is being studied.

Note that Android does not specifically detect a “public transportation” activity type – trips by bus or train would presumably fall under the “IN_VEHICLE” activity type. Future work can examine how to determine if a vehicle trip was a transit trip, private vehicle, or TNC service.
3.2 Activity confidence values

Each time an activity transition is detected, the device attempts to classify the new activity type. The research team examined the confidence value calculated by the device for each classified activity. Confidence is defined by Google as “a value from 0 to 100 indicating the likelihood that the user is performing this activity” [8]. Table 3.1 and shows the breakdown of confidence values collected shortly after an activity transition is detected (e.g., when a user just left their location to start a new trip or just after they arrived at a destination). These numbers include the confidence in classifying STILL activities (e.g., the activity type detected when a person transitions from a moving activity to being still).

Table 3.1 - The confidence values calculated for all activities when leaving origins or arriving at destinations

<table>
<thead>
<tr>
<th>Activity transition confidence at origin or destination</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>95% of transitions have confidence greater than 10</td>
<td></td>
</tr>
<tr>
<td>75% of transitions has confidence greater than 15</td>
<td></td>
</tr>
<tr>
<td>68% of transitions has confidence greater than 29</td>
<td></td>
</tr>
<tr>
<td>50% of transitions has confidence greater than 68</td>
<td></td>
</tr>
<tr>
<td>25% of transitions has confidence greater than 97</td>
<td></td>
</tr>
<tr>
<td>20% of transitions has confidence greater than 99</td>
<td></td>
</tr>
</tbody>
</table>

![Distribution of activity transition confidence values](image)

Figure 17 - When looking at all activity transitions, there is a long tail of relatively low confidence values
While 25% of activities following transitions have a confidence greater than 97, when looking at a larger percentage of activities that number drops. 50% of activities following transitions have a confidence greater than 68 (and therefore 50% of activities have a confidence less than 68), and 75% of activities following transitions have a confidence greater than 15 (and therefore 25% of activities have a confidence less than 15).

One possible explanation for lower confidence values is that we are sampling the confidence value immediately when the new activity is being started, and therefore there is limited data on the new activity for classification. It is possible that if the activity was classified again after it had been in progress for a longer period that the confidence level would be higher.

Another potentially source of lower confidence values may be related to some activity transitions being triggered and then corrected when the device obtains better estimates (e.g., a STILL to WALKING transition, immediately followed by a WALKING to STILL transition with a higher confidence values). This matches the anecdotal evidence from the research team when testing the application – the device often correctly detected walking events indoors but did trigger some false positives for larger movements that did not consist of walking. It makes sense for the Transition API to err on the side of false positives rather than false negatives (i.e., missing a transition). False positions could be further filtered out with additional processing, while a false negative would result in the application not being alerted that an activity could have occurred. It should also be noted that false positives with a low confidence value would effectively be counted twice (when entering and exiting the WALKING activity) with the methodology used in this section, so future analysis should attempt to only count low confidence values for false positives once.

To test if these short duration and distance trips had an impact on the confidence values, the research team created a subset of trips that only included trips with motion activities (i.e., no STILL trips), with all trips being greater than 5 minutes in duration and having an origin-to-destination distance of greater than 50 meters. The results for these 13,698 trips are shown in Table 3.2 and Figure 18.
Table 3.2 - The confidence values for motion activities for trips greater than 5 minutes in duration and 50 meters in distance

<table>
<thead>
<tr>
<th>Activity recognition confidence - No STILL events, trips greater than 5-min duration and 50-meter distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% of transitions have confidence greater than 10</td>
</tr>
<tr>
<td>75% of transitions have confidence greater than 31</td>
</tr>
<tr>
<td>68% of transitions have confidence greater than 69</td>
</tr>
<tr>
<td>50% of transitions have confidence greater than 92</td>
</tr>
<tr>
<td>25% of transitions have confidence greater than 97</td>
</tr>
<tr>
<td>20% of transitions have confidence greater than 97</td>
</tr>
</tbody>
</table>

The confidence values are greatly improved in this subset of data. 25% of activities following transitions still have a confidence greater than 97, but when looking at a larger percentage of activities the numbers are greatly improved. 50% of activities following transitions have a confidence greater than 92 (versus a confidence of 68 previously), 68% of transitions have confidence greater than 69 (versus a confidence of 29 previously), and 75% of activities following transitions have a confidence greater than 31 (versus a confidence of 15 previously). There is still a spike in the number of activity transitions with a confidence value of 10, which may be an artifact of the underlying activity classification algorithm.
The device is estimating these activity transitions in real-time and therefore doesn’t have the advantage of knowing that another activity transition will occur immediately in the future with a higher confidence. However, in our post-processing analysis higher confidence events could help filter out preceding lower confidence events – the results shown in Table 3.2 indicate that this approach is promising. Future data processing could focus on accurately filtering out false positive activity transitions while keeping high confidence short duration or distance trips (e.g., short walking trips).

Google does not disclose the exact methodology behind how it calculates the classified activity confidence value. As a result, further benchmarking should be performed to determine how this value correlates to real-world performance. Past research has investigated the accuracy of Google’s activity recognition algorithm under certain conditions, but has not evaluated the accuracy of the confidence value associated with a recognized activity [28].

Additionally, it should be noted that the confidence value is not directly provided by Android for the activity transition itself – as mentioned in an earlier section, the research team is obtaining the confidence value by calling the Activity Recognition API when the Activity Transition API is triggered. As a result, activity confidence may not directly correlate to the exact instant when the activity transition is calculated, and this value always represents the confidence of the activity that is being started (i.e., not the one that just concluded). Therefore, it’s not immediately clear if this confidence reflects only the classification of the activity being performed, or if it also reflects the accuracy of detecting a transition between two activities – although it’s likely that the two values are related. The accuracy of confidence on various Android version and device models should also be examined. Performance of activity recognition is expected to be better on newer devices with more modern hardware. Anecdotally, the research team’s experience collecting activity transitions on two Samsung Galaxy S8+ devices with two different users was quite accurate – modes were typically classified correctly and accurately represented where the user started and stopped moving. These devices were released in March 2017 and therefore a little over two years old when tested by the research team. It should also be noted that Google’s algorithm for activity classification might change over time – potentially bringing along improvements in accuracy.

3.3 Location data providers

The location data for each trip origin and destination was examined next. Using the “best” location selection rule presented earlier, Figure 19 shows the breakdown of the location providers selected for each trip origin location. Similarly, Figure 20 shows the breakdown of the location providers selected for each trip destination. These charts include the locations calculated for STILL activities.
Further analysis shows that of the 65,582 trips (excluding STILL events), only a little under half (32,556) have the same origin and destination location provider (Figure 21).
The origins and destinations with “no location” shown in Figure 19 and Figure 20 indicates that the user likely changed a device setting preventing access to location information – likely either turning location technology off on their phone or turning off the permission for the OneBusAway app to access location information for this trip. The app prompts the user to turn on location technology and permissions when the user first installs the app, and when the user enrolls in the study, although the user can select not to be prompted again in the future. The research team intentionally did not repeatedly prompt or require the user to enable these location settings to allow the user privacy if they chose to turn off these settings for trips. For these users, the Activity Transition API will still be triggered, and activity transitions are recorded, although no location information is collected when these settings are off. The lack of location information obvious limits the ability to use this data for transportation planning purposes requiring spatial data, although the data could still be useful for studies involving health and information when it’s important to know how active a user is (e.g., duration of activities) but spatial information isn’t important. If future studies required location information for all users and provided an incentive that required users to leave these settings on, the app could be configured to require the user turn these location settings on.
3.4 Time delay between activity transition and location acquisition

The time delay between when an activity transition occurs (e.g., a trip starts or ends) and a location is calculated is also important to better understand how close a calculated location may be to the position where the activity transition occurred. As mentioned earlier, there can potentially be a delay between when the device hardware captures activity transition information and that information is delivered to the application, as well as between when the application receives the activity transition information and new position information is acquired. Figure 22 shows a typical timeline for user travel behavior – in this example, a walking trip.

Before the trip starts, at time \( t_{-1} \), the user is sitting still at their desk. At \( t_0 \), the Activity Transition API fires when the user transitions from being still to walking. At this time, the OneBusAway app requests a location update, which is delivered to the app at \( t_1 \). The walking trip duration is from \( t_0 \) to \( t_n \), when the Activity Transition API is fired again when the user stops walking. The OneBusAway app again requests a location update, which is delivered to the app at \( t_{n+1} \). As shown in Figure 23, we can determine the time delay at the origin, or \( \Delta t_{\text{origin}} \), by subtracting \( t_0 \) from \( t_1 \). Similarly, we can determine the time delay at the destination, or \( \Delta t_{\text{destination}} \), by subtracting \( t_n \) from \( t_{n+1} \).

\[
\Delta t_{\text{origin}} = t_1 - t_0 \\
\Delta t_{\text{destination}} = t_{n+1} - t_n
\]

Table 3.3 shows the time delays, \( \Delta t_{\text{origin}} \) and \( \Delta t_{\text{destination}} \), between when an activity transition occurs, and a new location is calculated.
Table 3.3 - The time delay (Δt) between activity transitions and location acquisition

<table>
<thead>
<tr>
<th></th>
<th>Origin</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>minutes</td>
<td>seconds</td>
<td></td>
</tr>
<tr>
<td>95th percentile</td>
<td>7.57</td>
<td>454.05</td>
<td></td>
</tr>
<tr>
<td>90th percentile</td>
<td>3.20</td>
<td>192.00</td>
<td></td>
</tr>
<tr>
<td>68th percentile</td>
<td>0.23</td>
<td>14.00</td>
<td></td>
</tr>
<tr>
<td>50th percentile</td>
<td>0.10</td>
<td>6.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Destination</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>minutes</td>
<td>seconds</td>
<td></td>
</tr>
<tr>
<td>95th percentile</td>
<td>6.25</td>
<td>375.00</td>
<td></td>
</tr>
<tr>
<td>90th percentile</td>
<td>2.93</td>
<td>175.50</td>
<td></td>
</tr>
<tr>
<td>68th percentile</td>
<td>0.35</td>
<td>21.00</td>
<td></td>
</tr>
<tr>
<td>50th percentile</td>
<td>0.10</td>
<td>6.00</td>
<td></td>
</tr>
</tbody>
</table>

The 95th percentile of \(\Delta t_{\text{origin}}\) is 7.57 minutes, although the 90th percentile brings it down to 3.20 minutes and the 68th percentile is even lower at 14 seconds. Therefore, while there are some outliers with large acquisition times (likely due to the phone being in deep sleep and not actively monitoring location sensors to save energy), most time delays are under a few minutes.

When the user is actively traveling the phone has already been woken up from a deep sleep state and may have recently calculated a location, and as a result, \(\Delta t_{\text{destination}}\) seems to have fewer outliers of large delays than \(\Delta t_{\text{origin}}\). The 95th percentile of \(\Delta t_{\text{destination}}\) is 6.25 minutes (over a minute less than \(\Delta t_{\text{origin}}\)), with the 90th percentile being 2.93 minutes, and the 68th percentile being 21 seconds.

While there are some large outliers that could affect how close the calculated location is to where the activity transition occurred, this is an acceptable tradeoff for constantly monitoring the user’s travel behavior without negatively affecting battery life. Further research could examine trips that have a large time lag when leaving the origin to determine if the previous trip’s destination could be used as a proxy for the origin location of the next trip. Additional research could also look at having the user label locations such as “home” or “work” to assist in identifying the correct trip start or end positions for frequently visited locations, as well as averaging the positions for frequently visited locations over time to try to remove some of the delay outliers.

3.5 Estimated accuracy of locations

The estimated horizontal accuracy of the calculated locations for each trip origin and destination are also important. Estimated accuracy of each location is calculated by the device and is defined as “the radius of 68% confidence” by the Android specification [24]. A visual representation of estimated accuracy is shown in Figure 24.
Table 3.1 shows that the “best” location selected using the algorithm outlined earlier in this report were accurate approximately to a building-level or better, with the 95th percentile of accuracy at both origins and destinations of trips being approximately 48 meters.

Table 3.4 - Estimated accuracy of locations following activity transitions

<table>
<thead>
<tr>
<th>Estimated accuracy of locations following activity transitions</th>
<th>meters</th>
<th>miles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Origin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95th percentile</td>
<td>47.96</td>
<td>0.03</td>
</tr>
<tr>
<td>90th percentile</td>
<td>38.17</td>
<td>0.02</td>
</tr>
<tr>
<td>68th percentile</td>
<td>22.99</td>
<td>0.01</td>
</tr>
<tr>
<td>50th percentile</td>
<td>20.00</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Destination</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95th percentile</td>
<td>48.00</td>
<td>0.03</td>
</tr>
<tr>
<td>90th percentile</td>
<td>38.46</td>
<td>0.02</td>
</tr>
<tr>
<td>68th percentile</td>
<td>23.00</td>
<td>0.01</td>
</tr>
<tr>
<td>50th percentile</td>
<td>20.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

This level of accuracy should be sufficient for most general origin-destination analysis.

Additional analysis could be performed with changes to the “best” location selection algorithm to determine if additional accuracy could be achieved without sacrificing further delay between when the activity transition occurs, and the location is calculated. Locations from the previous trip destination could also be leveraged when attempting to determine the origin location of a trip and vice versa as well, if one of the two locations has a worse accuracy value.
Additional analysis of the estimated accuracy value could also be useful to confirm that the devices are correctly reporting the radius of 68% confidence. One potential method is to crowdsourced actual origin and destination locations directly from users, although this would require additional effort from the user and a method of capturing this data. A relatively simple way to start this type of analysis would be to record the users’ true home and work locations, and then compare the accuracy of location data collected over time as they visited those locations.

3.6 Power savings mode

When the research team tested one particular device (a Samsung Galaxy S6, SAMSUNG-SM-G920A, with Android 7.0), it became apparent that when the battery saver mode of Android was active the application did not receive updates from the Activity Transition API unless the OneBusAway application was actively being used. This behavior aligns with the battery saver mode documentation [29], which says that the background execution of many services may stop. Older devices that do not have newer sensor co-processors to assist in saving power while monitoring activity transitions may especially be prone to stopping all data collection when battery saver mode is active.

A typical use case of battery saver mode is users with older devices whose battery capacity has declined over time and may not typically last an entire day when battery saver mode is turned off. Therefore, users turn on this mode to stretch out the lifespan of their device. As a result, deployments of the data collection software should monitor the use of power savings mode by users to determine if the use of this feature affects the ability to collect data from users with older devices.

The user could also potentially affect the ability of the OneBusAway app to collect activity transitions in the background by changing other power management-related Android system settings. Via Android system settings, users could potentially “blacklist” the OneBusAway app from running in the background, or conversely, “whitelist” the OneBusAway app so it was allowed to run more freely in the background (i.e., ignore the “Doze” feature [30] of Android). While Android provides the functionality for developers to actively prompt the user to “whitelist” the app, Google prohibits apps from requesting exemption from power management features unless “the core function of the app is adversely affected” [31]. As a result, the research team chose not to prompt the user to whitelist the OneBusAway app.

Additional analytics capabilities were added to the app to track if users have enabled battery saver mode or have otherwise changed settings to ignore battery optimizations, which are discussed in the following section.
4 Enhancements following initial deployment

After the data collection software was released to the beta testing group on Google Play, the research team monitored the data being collected in real-time and made a variety of enhancements. These enhancements have been merged into the master branch of the OneBusAway Android project [32] and deployed as application updates to beta testers via Google Play.

The enhancements include:

- Additional analytics to capture when a user leaves the study if they choose to turn off travel behavior data collection in the app settings
- Enhancements to the enrollment process:
  - Additional analytics for each enrollment step so the number of people exiting enrollment at different stages can be calculated (e.g., those that dismiss the initial dialog prompt vs. those that decline the informed consent dialog)
  - The informed consent was updated to inform users that they may be asked to answer micro-surveys in the future. Users who had already enrolled were notified via email of this updated informed consent.
  - Added a "Not now" option when the user is first prompted to enroll in the study. This feature allows users to put off enrolling if they do not want to read the informed consent the first time they see the dialog. If they choose this option, they will be prompted again later. This is especially important given that OneBusAway contains time-sensitive information and if users are forced to choose only between “Accept” and “Decline” they may choose “Decline” simply because they don’t have time to read the informed consent in that moment. Therefore, users in this situation can choose to be reminded later of the option to enroll, which will potentially increase the overall number of users enrolled.
  - Added an email confirmation entry to the dialog when enrolling participant in study. This email confirmation field, in addition to the first email entry field, helps detect any user errors when entering their email address.
- Capture if power savings-related settings are enabled as part of device information whenever these settings are changed. As discussed earlier, these settings may affect activity transition detection in the background. The value of the Android PowerManager.isPowerSaveModeEnabled() [33] for device-wide power settings is captured, while the value of Android PowerManager.isIgnoringBatteryOptimizations() [34] is captured for user power settings specific to the OneBusAway app.
5 Conclusions and Future Work

This project successfully implemented and deployed travel behavior data collection software within the OneBusAway Android open-source mobile app that allowed users to opt-in to contribute data about their activities and app usage. The Android Activity Transition API was used to minimize impact on device battery life while still capturing the locations and times at which users changed from one activity to another (IN_VEHICLE, ON_BICYCLE, RUNNING, STILL, WALKING). This design allows for an extended data collection period (e.g., months, years) and collects data even when the OneBusAway app is not running with a minimal impact on the device. User emails were also collected and stored in a separate location from the activity data to allow future follow-up for additional data collection or verification (e.g., on-line or in-app survey questions).

Data was collected from 74 of the 676 users from the OneBusAway Android beta testing group, who opted into the study without any incentive, over about 10 weeks during this project. This data consisted of 65,582 trips representing detected user activity (WALKING, RUNNING, IN_VEHICLE, ON_BICYCLE, or WALKING/RUNNING) from an origin to a destination. A very large percentage (about 73%) of detected trips are walking trips, likely because the Activity Transition API can detect very short walking trips within a building (e.g., from an office to a conference room). When filtering out any trips that had a duration less than 5 minutes or had a distance between an origin and destination of less than 50 meters, 13,698 trips remained, with a lower share of walking trips (a drop of 73% to 43%) but a higher share of vehicle trips (an increase from 21% to 48% of trips).

The confidence values of activity recognition from the Android Activity Transition API deserve further scrutiny and evaluation. When filtering out trips less than 5 minutes in duration or 50 meters in distance, 50% of activities following transitions had a confidence greater than 92, 68% of transitions have confidence greater than 69, and 75% of activities following transitions have a confidence greater than 31 (versus a confidence of 15 previously). The role of false positives in detecting WALKING activity while stationary, which seems to be a cause of some of the low confidence values, should be examined. However, the bias of the Transition API towards triggering false positives rather than false negatives (i.e., missed trips) should be viewed as a benefit to the research. Further analysis and filtering of false positives can always be implemented – however, if false negatives were to occur, the research team would have a much harder time reconstructing the travel behavior information because no data during that time period would be recorded. The only false negatives observed by the research team occurred when power savings mode was activated on an older device (Samsung Galaxy S6, SAMSUNG-SM-G920A, with Android 7.0). Research based on this data will need to determine what impact the lack of these older devices may have on the study population.
Location data was collected for 86% of all origins and destinations of trips—the remaining 14% of locations could not be acquired due to user settings turning off location or restricting location permissions. The delay between detecting an activity transition and acquiring a position were reasonable, especially given the low-energy tradeoffs of the data collection software used in this research. While there are some outliers with large acquisition times (likely due to the phone being in deep sleep and not monitoring location sensors to save energy), the majority of time delays are under a few minutes. The 90th percentile of delay at origins was 3.2 minutes and the 68th percentile is even lower at 14 seconds, with destinations having even less delay, likely because the phone was recently woken up by the user during travel. The “best” location for each trip origin and destination selected using the algorithm outlined earlier in this report were accurate approximately to a building-level or better, with the 95th percentile of estimated accuracy at both origins and destinations of trips being approximately 48 meters. This level of accuracy should be enough for most general origin-destination analysis.

Travel behavior data collection within the OneBusAway app will continue beyond the end of this research project, with the anticipation of adding in-app and on-line survey in future research to collect additional information and validate passively collected data. Users continue to enroll at a rapid pace, especially given that no incentive is being offered. As of January 23, 2020, 105 users from the beta testing group of 740 users had enrolled in the study because of a single in-app prompt to “learn more” about the research – over 14% of the beta user base. No users have withdrawn from the study, and no users have reported any negative consequences (e.g., reduced battery life) from being enrolled in the study. Clearly, users are willing to contribute travel behavior as part of a longitudinal research study if there is not a noticeable negative impact on their device. Given that the data collection software has proven to be successful in the deployment to the beta test user group in this study, the software update will be rolled out to all OneBusAway Android users.

As of January 2020, the hosting cost to store the 105 users’ data in Firebase was approximately $7 per month using the Blaze Plan (https://firebase.google.com/pricing), with a breakdown of the costs from the Google Billing Console shown in Table 5.1.

<table>
<thead>
<tr>
<th>Service description</th>
<th>SKU description</th>
<th>Usage amount</th>
<th>Usage unit</th>
<th>Cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BigQuery</td>
<td>Streaming Insert</td>
<td>49,462.14</td>
<td>mebibyte</td>
<td>2.47</td>
</tr>
<tr>
<td>BigQuery</td>
<td>Long Term Storage</td>
<td>195.129</td>
<td>gibibyte month</td>
<td>1.85</td>
</tr>
<tr>
<td>BigQuery</td>
<td>Active Storage</td>
<td>89.67</td>
<td>gibibyte month</td>
<td>1.59</td>
</tr>
<tr>
<td>App Engine</td>
<td>Cloud Firestore Storage</td>
<td>7.034</td>
<td>gibibyte month</td>
<td>1.09</td>
</tr>
<tr>
<td>App Engine</td>
<td>Cloud Firestore Entity Writes</td>
<td>702,383</td>
<td>count</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>7.23</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 - Data hosting costs for the month of January 2020 for 105 users
In summary, the software deployed in this project is a promising technology with a fundamental tradeoff of reduced data density (i.e., collecting origin and destination locations instead of a breadcrumb travel path) for a reduced impact on battery life and the ability to collect data from a large number of users for longitudinal studies with few incentives required. The OneBusAway app can be downloaded from Google Play at https://play.google.com/store/apps/details?id=com.joulespersecond.seattlebusbot, with the source code available on GitHub at https://github.com/OneBusAway/onebusaway-android. The data acquired during the study appears to be viable for various types of travel behavior research (e.g., origin/destination, health, route choice), although additional data processing will be needed to filter data for specific types of research and to validate the data at a larger scale than was possible in this study (e.g., via user feedback). It should be noted that additional location data collection (e.g., breadcrumb travel paths) could be selectively added to the app, but it comes at a cost to device battery life and therefore should be only added if vital to a study.

The following section describes some future work directions observed by the research team during this project that were not explicitly mentioned above.

5.1 Future work

As previously stated, the Transition API recognizes a general IN_VEHICLE type of activity. Further analysis, perhaps using additional machine learning algorithms as well as the other in-app usage data collected (e.g., arrival information at bus stops viewed, trips planned, destination reminders used) could be used to further classify trips via public transportation, ride-hailing services, private vehicle, and more. Currently, when collecting arrival information viewed, the app does not detect if the user scrolled the screen down to view all arrival times in the list – further data collection could more precisely identify which times in the list were viewed and which were not, which may be important for route-choice studies. Explicitly asking users to identify their home and work locations, either in-app or via an on-line survey, would greatly help in further analyzing the users’ data, and could also assist the user when planning frequent trips to and from home and work locations.

Additional data processing could examine filling in missed trips (e.g., when the user turns on power savings mode). Trip planners could be used to supplement the data collected from users.

Analytics data is being collected for the various screens a user sees when prompted to opt into the study. This data could be analyzed to determine when users decline to join the study to determine if the on-boarding process for users could be improved.

Various incentive mechanisms could be examined to further boost participation, including micro-targeting specific geographic locations (e.g., users that drive on a particular road) or demographics.
While there are some large outliers that could affect how close the calculated location is to where the activity transition occurred, this is an acceptable tradeoff for constantly monitoring the user’s travel behavior without negatively affecting battery life. Further research could examine trips that have a large time lag when leaving the origin to determine if the previous trip destination could be used as a proxy for the origin location of the next trip. User-entered “home” or “work” locations would assist in identifying the correct trip start or end positions for frequently visited locations, as well as averaging the positions for frequently visited locations over time to try to remove some of the delay outliers.

The device is estimating these activity transitions in real-time and therefore doesn’t have the advantage of knowing that another activity transition will occur immediately in the future with a higher confidence. However, in our post-processing analysis higher confidence events could help filter out preceding lower confidence events—the results shown in Table 3.2 indicate that this approach is promising. Future data processing could focus on accurately filtering out false positive activity transitions while keeping high confidence short duration or distance trips (e.g., short walking trips). It should also be noted that false positives with a low confidence value would effectively be counted twice (when entering and exiting the WALKING activity) with the methodology used in this section, so future analysis should attempt to only count low confidence values for false positives once.

Google does not disclose the exact methodology behind how it calculates this confidence value. As a result, further benchmarking should be performed to determine how this value correlates to real-world performance.

The lack of location information from some users due to user permissions and location settings limits the ability to use this data for transportation planning purposes requiring spatial data, although the data could still be useful for studies involving health and information when only time data is important (e.g., duration of activities). If future studies required location information for all users and provided an incentive that required users to leave these settings on, the app could be configured to require the user turn these location settings on, although it would come at a cost to battery life.

Additional analysis could be performed with changes to the “best” location selection algorithm to determine if additional accuracy could be achieved without sacrificing further delay between when the activity transition occurs, and the location is calculated. Locations from the previous trip destination could also be leveraged when attempting to determine the origin location of a trip and vice versa as well, if one of the two locations has a worse accuracy value.

Additional analysis of the estimated accuracy value could also be useful to confirm that the devices are correctly reporting the radius of 68% confidence. One potential method is to crowdsourcer actual origin and destination locations directly from users, although this would require additional effort from the user and a method of capturing this data. A relatively simple way to start this type of analysis would be to
record the users’ true home and work locations, and then compare the accuracy of location data collected over time as they visited those locations.

To obtain qualitative information, very brief targeted micro-surveys (1-2 questions) for high-priority issues such as safety could be delivered to travelers when and where they are likely to have more free time – for example, while they are waiting for the bus at a bus stop. Connections to social media, with the user’s permission, could be used to obtain demographics information without placing a burden on the user.

The software implemented in this project was for the OneBusAway Android app, as it used the Android Transition API. Future work could investigate a similar implementation to collect data from OneBusAway iOS users (e.g., using CMMotionActivity [35]).
6 References


7 Appendix A – USF IRB Application

Study Protocol Title:
Improving the Quality and Cost Effectiveness of Multimodal Travel Behavior Data Collection

List of Abbreviations:
List of commonly used abbreviations and acronym.

- OBA: OneBusAway
- TNC: Transportation Network Company
- DOT: Department of Transportation
- SOV: Non-single Occupancy Vehicle
- MPO: Metropolitan Planning Organization
- GPS: Global Positioning System

Principal Investigator, Research Team, and Study Site:

Principal investigator: Sean J. Barbeau

Sean J. Barbeau, Ph.D.
Principal Mobile Software Architect for R&D
National Center for Transit Research (NCTR) at the
Center for Urban Transportation Research
University of South Florida
4202 E. Fowler Avenue CUT 100
Tampa, FL 33620-5375
barbeau@cutr.usf.edu
813-974-7208

Research team and contact Information:

Cagri Cetin
Graduate Research Assistant
Center for Urban Transportation Research
University of South Florida
4202 E. Fowler Avenue CUT 100
Tampa, FL 33620-5375
cagricetin@mail.usf.edu

Study site:
Research Synopsis:

This study observes the travel behavior of a user of the OneBusAway (OBA) Android mobile app, a popular mobile app for real-time public transportation information available in seven cities [1]. The travel behavior data will be used to better understand when and how transit riders choose to use public transportation, which will in turn help improve multi-modal transportation planning and transit operations. The travel behavior data will consist of the trip origin and destination, activity of the user such as walking and riding, and stop wait times.

Existing users of the OneBusAway app will be given the option to opt into the study. If the user opts in, data travel behavior data will be collected automatically and stored in a secured location in the user’s device and will be collected from the user after he or she agrees to the informed consent. The data will be sent to a secure server (e.g., Google Firebase Storage) through a secure encrypted communication channel. This raw data will only be accessible by the research team. Participants will also be sent online surveys to complement the automated app-based data collection. These surveys will collect demographics information as well as survey as a one-day travel diary for manually recording travel behavior.

Background and Significance:

Multimodal transportation options such as transit, bike, walk, transportation network companies (TNCs e.g., Uber, Lyft), carshare, and bikeshare are vital to supporting livable communities. To build safe and effective multimodal infrastructure, Departments of Transportation (DOTs), Metropolitan Planning Organizations (MPOs), and transit agencies need quality data about how the public is currently traveling via these modes. However, current data collection techniques for multimodal travel behavior have limitations which restrict the ability to solve significant real-world multimodal problems.

One example area lacking robust multimodal data is the relationship between (Transportation Network Companies (TNCs) and public transit. Some see TNCs as a competition to public transit that is primarily responsible for trends of dropping transit ridership. Others see TNCs as vital first/last mile service that is a complement to public transportation. Some agencies, such as Pinellas Suncoast Transit Authority (PSTA) in Tampa, FL, have formed partnerships with a TNC (e.g., Uber) in order to help fill the first/last mile gap. Other agencies such as Hillsborough Area Regional Transit (HART) have started operating their own TNC-like service (in the cases of HART, via a contractor Transdev). However, it can be extremely difficult to capture travel behavior data that includes holistic information about TNCs and public transit
– as of today, there is little hard data that includes origin/destination trip data for transit and TNCs, especially when one mode is used in place of another. The primary method of capture, on-board surveys, would completely miss travelers that opted to use a TNC instead of transit for a trip. And, any information that is captured only covers a day or two of behavior – on-board surveys do not capture longitudinal behavior. As a result, practitioners and researchers have yet to understand the precise relationship between the two modes.

This research will develop and deploy a system that will collect multimodal travel behavior data on an ongoing basis directly from users of a popular mobile app for multi-modal information, OneBusAway (OBA). The OBA mobile app is currently deployed in six cities, including at Florida transit agencies Hillsborough Area Regional Transit (HART) and Pinellas Suncoast Transit Authority (PSTA). The purpose of this research is to increase quality and cost-effectiveness of multimodal travel behavior data collection by automating much of the data collection process and provide the research team insight into when and how travelers choose to use public transportation. In the long-term, better data will assist planners in understanding how and where people are traveling via non-single occupancy vehicle (SOV) modes, which will enable DOTs, MPOs, and transit agencies to better prioritize infrastructure investments and make operational improvements.

**Objectives:**

The objectives of this project are to improve multimodal infrastructure planning and transit service quality by increasing the cost-effectiveness and quality of data from all modes of transportation, including transit, bike, walk, transportation network companies (TNCs, for example, Uber and Lyft), carshare, and bikeshare. This objective is accomplished via the development of a plugin for the OneBusAway mobile app to collect travel data including origin/destination, transfers, trip travel path, and travel time data. The tool can immediately leverage the large existing user base of OneBusAway instead of trying to recruit users to use a new app dedicated to travel behavior data collection.

**Key objectives:**

1. Reduce data collection cost per completed trip based the cost of existing data collection efforts used by transit agencies, MPOs, and FDOT.
2. Collect travel behavior data from 500 OBA users that opt-in to detailed tracking using OBA.
3. Better understand user travel behavior, in particular the choice of where and when to use public transportation.

**Study design/methodology:**

This project will be a collaboration between USF researchers (“USF research team”) and non-USF researchers (“non-USF research team”).

Non-USF research-team members are:

Candace Brakewood, Ph.D.
Assistant Professor
This IRB application only covers the activities of the USF research team and explains the exchange of data between USF and non-USF research teams. Non-USF research team will pursue IRB review at their own institutions to cover their activities.

Figure 1 summarizes the research-tasks that are going to be done throughout this project in collaboration
with USF and non-USF research teams. As part of this project, travel-behavior data collection feature will be integrated into OneBusAway by the USF research team, adding the ability to collect travel-behavior data from existing OneBusAway users (with their consent). The OneBusAway application will collect the participants’ email addresses and travel-behavior data and share this information with the non-USF research team via a secure, password-protected server. Non-USF research team will prepare surveys, send these surveys to participants through emails, collect raw survey data, and perform data analysis on the raw survey and travel-behavior data in collaboration with the USF research team. The non-USF research team will also share the raw survey data with the USF research team via secure and password-protected servers.

Figure 2: The flow of the travel-behavior study design.
Figure 2 illustrates the flow of the travel-behavior data collection with OneBusAway. As shown in Figure 2, once the new travel-behavior data collection feature becomes available in the new version of the OneBusAway (OBA) application, and a user opens the application for the first time after the new feature update, OBA will show a confirmation dialog in the screen. This dialog will ask the user to participate in this study which aims to improve public transportation. If the user taps the “No” button, the user will not be enrolled in the study.

If the user taps the “Yes” button, the user will be prompted to confirm his/her age. If the user is not an adult and taps the “No” button, the user will not be enrolled in the study.

If the user confirms his/her age is over 18 by tapping the “Yes” button, the user will be prompted to provide their consent (the informed consent document is attached to this IRB application) for sharing their travel-behavior information as illustrated in Figure 1. If the user taps the “I Decline” button, the user will not be enrolled in the study.

Upon providing consent by tapping the “I Agree” button, OneBusAway application will ask an email address as contact information from the user. This email address will be the only identifying information among all other information collected within the mobile app. This email address will be used by the non-USF research team to send surveys about the user’s demographic and travel behavior information. Furthermore, this email address will be associated with an anonymous unique identifier. This unique identifier will be used to match the travel behavior information that will be collected by the OneBusAway application with the online survey data. The mapping between email addresses and unique identifiers will be stored in a separate and secure and password-protected online database (e.g., Google Firebase Cloud Storage), will only be accessible by the USF and non-USF research teams and will not be published.

Upon providing the email, the user’s raw multimodal travel-behavior information, including the anonymous unique identifier, will be automatically and securely collected and forwarded to a secure and password-protected online server (e.g., Google Firebase Cloud Storage). A participant can anytime stop sharing his/her travel-behavior information and changing his/her preference from the app settings.

The raw multimodal travel-behavior data collected from users of the OneBusAway app will only be accessible by the USF and non-USF research teams. Although there is no user-identifying information in the automatically collected data, location path information when combined with other data sources can potentially be revealing - users may be identified by cross-referencing with other public data source such as property appraisal databases. Therefore, we will not share or publish any raw travel behavior data – it will be accessible only to the research team.

An example of the type of raw data that will be collected as a part of the study is given below:

0, 0, 35833084925236,1468797302000,28.034712,-82.49605,-1,5.37269,265.799988, IN_VEHICLE, ACTIVITY_TRANSITION_ENTER, 1559230206005, 80,3,11, gps

1, 0, 35833084925236,1468797302000, 28.034712,-82.49606,-1,5.37279,265.799988, IN_VEHICLE, ACTIVITY_TRANSITION_EXIT, 1559230206005, 90, 3,11, gps
Starting from the first line, here are the columns that contain the automatically-collected travel behavior data:

1. coordinateID – unique ID for each location fix in the file
2. userID – anonymous unique ID for each user
3. the time in nanoseconds since the application started
4. the time in UTC
5. destination latitude
6. destination longitude
7. altitude
8. speed
9. bearing
10. detected activity
11. detected activity transition type
12. elapsed realtime when this transition happened
13. detected activity confidence level (“the likelihood that the user is performing this activity”)
14. horizontal accuracy
15. number of satellites used in fix
16. location provider

In addition to the location and activity information, OneBusAway will also store information about what the user sees in the app (e.g., arrival and departure information, trip plan information) when they are making travel choices. For example, arrival and departure information that is being shown to the participating user at the time that the user is getting into the bus will be collected. In this example, this information will enable us to determine why users choose a particular route. More specifically, in this example, we can reason that this user chose Route 5 because the user saw that the bus on Route 10 was running late. An example of arrival and departure information is shown in Appendix A; this data is publicly available real-time bus route information and does not contain any user-specific information.

Another example of collected data could be the trip plan information. OneBusAway can store a user’s trip-plan information that was recommended by the OneBusAway application within an hour of getting on a bus. In this example, the trip plan information will enable us to determine whether the users are taking actual trips that were recommended by the OneBusAway application. An example trip plan information is shown in Appendix B; this information contains an itinerary (i.e., directions) from one place to another.
After enrolling in the study, a demographic survey and a one-day travel diary survey will be sent via email to the participating OBA users by the non-USF research team. The demographic survey will be asked once. However, the non-USF research team will ask the one-day travel diary survey in recurring intervals such as every two months. The non-USF research team will store the results of the survey data in a secure and password-protected storage. The non-USF research team will share the raw survey data with the USF research team via the secure and password-protected server to assist with data analysis.

The demographic survey will ask questions about basic socio-demographic information and familiarity with the public transportation network in the city. The one-day travel diary survey will ask information about the user’s recent trips and will be used to validate the accuracy of the travel behavior data collected by the OneBusAway application. Every survey will pertain to factual data about trips taken by the user (i.e., where they traveled on that particular day). Drafts of demographic and travel diary surveys are attached as survey-instruments documents.

**Study Population:**

The study will observe the travel behavior of participating adult OneBusAway Android app users. All active OBA Android users that are over the age of 18 will be eligible to enroll in this study. There are currently approximately 180,000 active users of OBA Android application in six different regions. These regions are:

- Tampa Bay, Florida
- Washington, D.C.
- York, Ontario
- San Diego, California,
- Rogue Valley, Oregon
- Puget Sound, Washington

In all of the regions above, a person who is over the age of 18 is considered an adult.

**Study Schedule:**

This will be a longitudinal study. All active OneBusAway Android users will be notified when the travel-behavior data collection feature is available via a prompt in the app. As shown in Figure 2, we will collect travel-behavior data from the users that opted in to the study, are adults, and have accepted the informed consent for a period of two years. After two years, the OBA app will stop collecting travel-behavior data. The data analysis task will start in parallel with data collection task.

<table>
<thead>
<tr>
<th>Task #</th>
<th>Task end date (month/year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Travel behavior data collection software development and testing completion</td>
<td>July 1(^{st}), 2019</td>
</tr>
<tr>
<td>2 – Collecting travel behavior data from users</td>
<td>July 1(^{st}), 2021</td>
</tr>
</tbody>
</table>
The USF research team will perform Task 1. Additionally, the USF research team will collaborate with the non-USF research team to perform Tasks 2 and 3. The non-USF research team will (1) perform the online survey data collection task by preparing and sending surveys to participants in rolling intervals via email, and (2) perform data analysis on raw travel-behavior and survey data. In order for the non-USF research team to participate in this study, the USF research team will share the email-unique-id mapping data and raw travel-behavior data through a secure and password-protected online server with the non-USF research team. The non-USF research team will (1) use the email addresses to send surveys to participants, (2) use raw travel-behavior data and survey results (matched using the email-unique-id mapping) to perform data analysis.

**Statistical Analysis Plan:**

The non-USF research team in collaboration with the USF research team will collect raw multimodal travel-behavior data and survey results from all participating OBA users. This raw travel-behavior data will contain information about users’ travel origins and destinations, users’ travel activity, users’ GPS locations.

The email address and unique ids will be used to cross-reference the travel-behavior data and the survey data so that the automatically-collected data from the mobile app can be matched with online survey results for a user, including socio-demographics information for that user as well as their one-day travel diaries that can be compared to the automatically-collected data.

The non-USF research team in collaboration with the USF research team will compare the one-day travel diary survey results with the raw multimodal travel-behavior data collected from the OneBusAway application to assess the accuracy of the data collection methodology.

The aggregated, de-identified, and anonymized data will answer research-questions about users’ travel behaviors. Possible research outcomes from the aggregated data are listed below:

- Characteristics of access to information, linked to trip purpose, time of the day, supply features, etc.
  
  - Example result: 30% of the transit riders use the bus to go to school.

- Characteristics of the choice set (number of considered/used stops, number of considered/used lines, accepted walking distance) in relation to access to information (frequency, type, location).
  
  - Example result: 30% of the riders use more than one bus stop in a day

- Additional productive time estimation; Estimation of how long it takes from the first access to information to the time of departing from the location.

- Stop arrival buffer time estimation: Estimation of how long before predicted bus arrival passengers arrive at the stop, depending on information, personal attributes as well as service characteristics of chosen as well as unchosen service (e.g. arrive earlier if service is less
frequent).

- Example result: Users tend to wait 10 minutes before riding their bus. Also, users tend to wait longer in the bus stops that have shelters.

Sample size determination

The study will observe the travel behaviors of participating adult OneBusAway Android app users. OneBusAway is actively used in six regions (i.e., Tampa bay, Washington D.C., York, San Diego, Oregon Rogue Valley, and Puget Sound) and it has approximately 180,000 active users. A goal of this study is to enroll at least 500 users to test the automatic travel behavior data collection technology and compare it to online one-day travel behavior diaries, and to begin to study traveler behavior, especially when and where they use public transportation.

Informed Consent Process:

OneBusAway will provide the informed consent document on a screen within the app. The participating OBA user will have to read and agree (by tapping on the “I agree” button” on the consent) before enrolling the study. The user can only contribute data to the study if they agree to the informed consent.

Privacy and confidentiality:

The mapping of email addresses (used for the online surveys) and the anonymous unique IDs (used in the automated travel behavior data collection within the app) will be stored in a secure, password-protected server that is only accessible by the USF and non-USF research teams and is separate from both the online survey response as well as the app-based data. This data will not be published and only be used to send surveys to participants and pair the two datasets. The raw data will be sent to a secure server that is only accessible by the USF and non-USF research teams.

All raw travel-behavior data from the user device will be (1) stored in the device’s internal memory which is accessible only to the OneBusAway app and (2) forwarded to a secure, password-protected online server (e.g., Google Firebase Cloud Storage). All online survey data will also be stored in a secure, password-protected online server.

We will only publish aggregated, de-identified, and anonymized travel-behavior and survey data resulting from this study.

Risk/Benefit:

Risk to participants:

The potential risks to the subjects are minimal. The travel-behavior data and survey will be anonymous. The raw-travel behavior data and survey alone cannot be used to identify a user. While there is no identifying information in the survey and the raw travel behavior data, users may be identified by cross-referencing the email-unique-id mapping data with other publicly-available data (e.g., property appraisal databases). This cross-referenced data could potentially reveal sensitive locations that the user has visited. Therefore, we will not share the email-unique-id mapping and raw travel-behavior data and the disaggregated survey results – only the research team will have access to this data.
Benefits to Participants

There are no direct benefits to participants. However, this research will assist regions to better understand how and where people are traveling via public transportation, bike, and pedestrian modes, and will help inform DOTs, MPOs, and transit agencies in multimodal infrastructure and program investments, including improvements to bike and pedestrian facilities to access transit and transit rider safety. This research will also inform future automated travel behavior data collection efforts, which could reduce the ongoing costs related to transportation planning and operations, which benefits society as a whole.

Publication and Presentation Plans:

Only aggregated, de-identified, and anonymized travel-behavior and survey data will be presented in publications and conference presentations. Due to the privacy and confidentiality concerns, we will not publish any raw survey or travel-behavior data. For example, we will publish results in the following form: “10% of the users tend to wait 5 mins before riding the bus”.

References:


Appendices

Appendix A - An example arrival and departure information

```json
{
  "code": 200,
  "currentTime": 1558450549472,
  "data": {
    "entry": {
      "arrivalsAndDepartures": [
        {
          "arrivalEnabled": true,
          "blockTripSequence": 7,
          "departureEnabled": true,
          "distanceFromStop": 9184.623602097097,
          "frequency": null,
          "lastUpdateTime": 1558450534000,
          "numberOfStopsAway": 32,
          "predicted": true,
          "predictedArrivalInterval": null,
          "predictedArrivalTime": 1558452164000,
          "predictedDepartureInterval": null,
          "predictedDepartureTime": 1558452164000,
          "routeId": "Hillsborough Area Regional Transit_42",
          "routeLongName": "University Area Connector",
          "routeShortName": "42",
          "scheduledArrivalInterval": null,
```
"scheduledArrivalTime": 1558451504000,
"scheduledDepatureInterval": null,
"scheduledDepartureTime": 1558451504000,
"serviceDate": 1558411200000,
"situationIds": [],
"status": "default",
"stopId": "Hillsborough Area Regional Transit_6666",
"stopSequence": 37,
"totalStopsInTrip": 43,
"tripHeadsign": "North to University Area",
"tripId": "Hillsborough Area Regional Transit_1426835",
"tripStatus": {
  "activeTripId": "Hillsborough Area Regional Transit_1426835",
  "blockTripSequence": 7,
  "closestStop": "Hillsborough Area Regional Transit_7186",
  "closestStopTimeOffset": 14,
  "distanceAlongTrip": 1587.028867066896,
  "frequency": null,
  "lastKnownDistanceAlongTrip": 0,
  "lastKnownLocation": {  
    "lat": 28.02909660393555,
    "lon": -82.45943450927734
  },
  "lastKnownOrientation": 0,
  "lastLocationUpdateTime": 0,
  "lastUpdateTime": 1558451504000,
  "nextStop": "Hillsborough Area Regional Transit_7186",
  "nextStopTimeOffset": 14,
  "orientation": 95.4268124970699,
  "phase": "in_progress",
  "position": {  
    "lat": 28.039780285533773,
    "lon": -82.4593772212571
  },
  "predicted": true,
  "scheduleDeviation": 660,
  "scheduledDistanceAlongTrip": 1587.028867066896,
  "serviceDate": 1558411200000,
  "situationIds": [
    "Hillsborough Area Regional Transit_2feca8bd-14ea-47b8-b45b-c0446dd93cd0"
  ],
  "status": "SCHEDULED",
  "totalDistanceAlongTrip": 13064.578157199137,
  "vehicleId": "Hillsborough Area Regional Transit_1522"
},
"phone": "813-254-4278",
"privateService": false,
"timezone": "America/New_York",
"url": "http://www.gohart.org"
]
],
"routes": [
{
"agencyId": "Hillsborough Area Regional Transit",
"color": "09346D",
"description": "",
"id": "Hillsborough Area Regional Transit_42",
"longName": "University Area Connector",
"shortName": "42",
"textColor": "FFFFFF",
"type": 3,
"url": "http://www.gohart.org/Pages/maps-schedules.aspx"
}
],
"situations": [
{
"activeWindows": [],
"allAffects": [

{"agencyId": "Hillsborough Area Regional Transit",
"applicationId": "",
"directionId": "",
"routeId": "",
"stopId": "",
"tripId": ""
}
],
"consequences": [
],
"creationTime": 1558140800647,
"description": {
"lang": "en",
"value": "On Monday 5/27, all HART bus and van services will operate on a Sunday schedule. TECO Line Streetcar service will operate on a regular Monday schedule. No service on the following routes: 20X, 24LX, 25LX, 31, 60LX between Bloomingdale and Downtown, 360LX between Brandon and Britton Plaza, and all HARTFlex routes. Customer Service will be available via phone from 6:00 AM to 8:00 PM. HART administrative offices and transit center customer service windows will be closed."
},
"id": "Hillsborough Area Regional Transit_2feca8bd-14ea-47b8-b45b-c0446dd93cd0",
"publicationWindows": [
{
"from": 1558071000000,
"to": 1559017800000
}
],
"reason": "HOLIDAY",
"severity": "normal",
"summary": {
"lang": "en",
"value": "Memorial Day"
},
"url": null
}
],
"stops": [
{
"code": "6666",
"direction": "E",
"id": "Hillsborough Area Regional Transit_6666",
"lat": 28.080254,
Appendix B - An example trip-plan information

<?xml version="1.0" encoding="UTF-8"?>
<Response>
:requestParameters>
<date>05-21-2019</date>
<mode>TRANSIT,WALK,BICYCLE_RENT</mode>
<arriveBy>false</arriveBy>
<wheelchair>false</wheelchair>
<optimize>QUICK</optimize>
<showIntermediateStops>true</showIntermediateStops>
<fromPlace>28.0587,-82.4153</fromPlace>
<toPlace>28.0587,-82.4139</toPlace>
<time>11:51AM</time>
</requestParameters>
<plan>
<date>1558453860000</date>
<from>
<name>Origin</name>
<stopid />

"locationType": 0,
"lon": -82.429241,
"name": "Bearrs Av @ Reflections Apts",
"routeIds": [  
  "Hillsborough Area Regional Transit_42"
],
"wheelchairBoarding": "UNKNOWN"
},
{
"code": "7186",
"direction": "E",
"id": "Hillsborough Area Regional Transit_7186",
"lat": 28.0399,
"locationType": 0,
"lon": -82.458861,
"name": "Linebaugh Ave @ Florida Ave",
"routeIds": [  
  "Hillsborough Area Regional Transit_42"
],
"wheelchairBoarding": "UNKNOWN"

"trips": [  
  
  "blockId": "Hillsborough Area Regional Transit_312907",
  "directionId": "0",
  "id": "Hillsborough Area Regional Transit_1426835",
  "routeId": "Hillsborough Area Regional Transit_42",
  "routeShortName": "",
  "serviced": "Hillsborough Area Regional Transit_WE",
  "shapeId": "Hillsborough Area Regional Transit_43449",
  "timeZone": "",
  "tripHeadsign": "North to University Area",
  "tripShortName": ""
  
  ]
],
"text": "OK",
"version": 2
}
<bogusName>true</bogusName>
<lon>-82.415299400000001</lon>
<lat>28.058776000000002</lat>
<elevation/>
</steps>
<steps>
<distance>38.729</distance>
<relativeDirection>RIGHT</relativeDirection>
<streetName>service road</streetName>
<absoluteDirection>SOUTH</absoluteDirection>
<exit/>
<stayOn>true</stayOn>
<area>false</area>
<bogusName>true</bogusName>
<lon>-82.4146683</lon>
<lat>28.058732900000003</lat>
<elevation/>
</steps>
<steps>
<distance>58.36900000000001</distance>
<relativeDirection>LEFT</relativeDirection>
<streetName>path</streetName>
<absoluteDirection>EAST</absoluteDirection>
<exit/>
<stayOn>true</stayOn>
<area>false</area>
<bogusName>true</bogusName>
<lon>-82.41466670000001</lon>
<lat>28.0583846</lat>
<elevation/>
</steps>
</legs>
</legs>
<tooSloped>false</tooSloped>
</itineraries>
</itineraries>
</plan>
<error/>
<debugOutput>
<precalculationTime>70</precalculationTime>
<pathCalculationTime>25</pathCalculationTime>
<pathTimes>
<pathTimes>1</pathTimes>
<pathTimes>18</pathTimes>
<pathTimes>6</pathTimes>
</pathTimes>
<renderingTime>1</renderingTime>
totalTime:96</totalTime>
<timedOut>false</timedOut>
</debugOutput>
<elevationMetadata>
<ellipsoidToGeoidDifference>-25.835867013018333</ellipsoidToGeoidDifference>
<geoidElevation>false</geoidElevation>
</elevationMetadata>
</Response>
Informed Consent to Participate in Research
Information to Consider Before Taking Part in this Research Study

Pro #
Project ID # 2117-9063-19

Researchers at the University of South Florida (USF) study many topics. To do this, we need the help of people who agree to take part in a research study. This form tells you about this research study. We are asking you to take part in a research study that is called: “Improving the Quality and Cost Effectiveness of Multimodal Travel Behavior Data Collection”. The person who is in charge of this research study is Dr. Sean J. Barbeau. This person is called the Principal Investigator.

This study is sponsored by: National Center for Transit Research

Purpose of the Study

The purposes of this project are to improve transit service quality and multimodal transportation planning. More specifically, the objectives of the research project are to:

A. Better understand the travel behavior of transit users, including where and when travelers choose different mode of transportation such as the bus, walking, biking, and a car.
B. Reduce travel behavior data collection cost per completed trip for transit agencies

Why are you being asked to take part?

You are being asked to participate because you use OneBusAway Android mobile application.

Study Procedures

If you take part in this study, your travel behavior information will be automatically collected through the OneBusAway Android application and you will be asked to participate in online surveys about how and when you travel places. The surveys will ask you about your daily travel activities and demographic questions. Each survey will take about 10 minutes to complete it, and you will be asked to complete approximately 4 surveys over a period of two years. The research will be done at the University of South Florida.
Alternatives / Voluntary Participation / Withdrawal
You have the alternative to choose not to participate in this research study.

You should only take part in this study if you want to volunteer; you are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study.

Compensation, Benefits and Risks
You will receive a small incentive such as a $10 Amazon gift card to participate in this study. There is no cost to participate. This research is considered to be minimal risk. Minimal risk means that study risks are the same as the risks you face in daily life.

Privacy and Confidentiality
We will keep your study records as confidential as possible. It is possible, although unlikely, that unauthorized individuals could gain access to your responses because you are responding online.

Certain people may need to see your study records. By law, anyone who looks at your records must keep them completely confidential. The only people who will be allowed to see these records are: principal investigator - Dr. Sean J. Barbeau, the research assistant - Cagri Cetin, research collaborators such as Dr. Candace Brakewood, Dr. Jan-Dirk Schmoecker, Dr. Achille Fonzone, Dr. Kari E. Watkins, and Aaron Brethorst.

It is possible, although unlikely, that unauthorized individuals could gain access to your responses. Confidentiality will be maintained to the degree permitted by the technology used. No guarantees can be made regarding the interception of data sent via the Internet. However, your participation in this study involves risks similar to a person’s everyday use of the Internet.

Contact Information
If you have any questions about your rights as a research participant, please contact the USF IRB at (813) 974-5638 or contact by email at RSCH-IRB@usf.edu. If you have questions regarding the research, please contact the Principal Investigator at (813) 974-7208 or contact by email at barbeau@cutr.usf.edu.

We may publish what we learn from this study. If we do, we will not let anyone know your name. We will not publish anything else that would let people know who you are. You can print a copy of this consent form for your records.

I freely give my consent to take part in this study. I understand that by proceeding with this survey that I am agreeing to take part in research and I am 18 years of age or older.

[User will click “Yes” or “No” within the app]
Researchers at the University of South Florida (USF) study many topics. To do this, we need the help of people who agree to take part in a research study. The person who is in charge of this research study is Dr. Sean Barbeau. This study is sponsored by the National Center for Transit Research.

**Purpose of the Study**

The purpose of this study is to better understand the travel behavior of transit users, including where and when travelers choose different mode of transportation such as the bus, walking, biking, and a car. This data will be used to inform better multimodal planning and policy.

**How do I participate?**

If you take part in this study, your travel behavior information will be automatically collected by the OneBusAway Android application. You will also be asked to participate in a few online surveys about how and when you travel places. Each survey will take about 10 minutes to complete, and you will be asked to complete approximately 4 surveys over a period of two years. You may also be prompted occasionally to provide brief information about your travels (e.g., "Did you ride the bus on your last trip?").

**Withdrawal**

You can choose not to participate in this research study by tapping the "Decline" button at the bottom of this form. If you do participate, you are free to withdraw at any time via the Settings menu. There will be no penalty or loss of benefits if you stop taking part in this study. To participate in this study tap the "Agree" button at the bottom of this form.

**Compensation, Benefits and Risks**

There is no compensation for participating in this study. There is no cost to participate.

**Privacy and Confidentiality**

We will store your data securely in a database accessible only to the research team. Your email address will be used for communication about this study (e.g., a link to online surveys) and will not be used for any other purpose.

We may publish and share what we learn from this study. If we do, we will not let anyone know your name. We will not share anything else that would let people know who you are.

**Contact Information**
If you have any questions about this study, please contact Sean Barbeau at (813) 974-7208 or barbeau@usf.edu.

I freely give my consent to take part in this study. I understand that by proceeding with this survey that I am agreeing to take part in research and I am 18 years of age or older.