Lake Tahoe Case Study Validation: Travel Time Reliability Monitoring

Authors:
Tiffany Barkley, Berkeley Transportation Systems, Inc.
Armand Ciccarelli, Berkeley Transportation Systems, Inc.
Michael Darter, Berkeley Transportation Systems, Inc.
Rob Hranac, Berkeley Transportation Systems, Inc.
Eric Mai, Berkeley Transportation Systems, Inc.
Berkeley, CA; August 2011

George List, North Carolina State University
Nagui Roupail, North Carolina State University
Billy Williams, North Carolina State University
Raleigh, NC; August 2011

Transportation Research Board of the National Academies Privileged Document

This report, not released for publication, is furnished only for review to members of or participants in the work of SHRP 2. This report is to be regarded as fully privileged, and SHRP 2 must approve dissemination of the information included herein.
Acknowledgements:

The Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials sponsored this work. It was conducted in the second Strategic Highway Research Program (SHRP 2), which is administered by the Transportation Research Board of the National Academies. Bill Hyman, Program Officer for SHRP 2 Reliability, managed this project.
# TABLE OF CONTENTS

1. MONITORING SYSTEM ........................................................................................................ 4  
   STUDY DESCRIPTION ........................................................................................................ 4  
   SITE OVERVIEW .............................................................................................................. 5  
   SENSORS .......................................................................................................................... 6  
   DATA MANAGEMENT ....................................................................................................... 9  
   SYSTEMS INTEGRATION ................................................................................................. 9  
   SUMMARY ....................................................................................................................... 9  
   REFERENCES .................................................................................................................. 9

2. METHODOLOGICAL EXPERIMENTS ................................................................................. 14  
   OVERVIEW ..................................................................................................................... 14  
   BLUETOOTH DEVICE DATA ........................................................................................... 14  
   CALCULATING TRAVEL TIMES BASED ON BLUETOOTH DEVICE DATA ..................... 20  
   SUMMARY ..................................................................................................................... 36  
   REFERENCES .................................................................................................................. 36

3. USE CASE ANALYSIS ...................................................................................................... 37  
   A. IMPACT OF ETC READER DEPLOYMENT CONFIGURATION ON DATA QUALITY .. 38  
   B. IMPACT OF BLUETOOTH READER DEPLOYMENT CONFIGURATION ON DATA QUALITY ........................................................................................................... 48  
   C. USING BLUETOOTH AND ELECTRONIC TOLL COLLECTION DATA TO ANALYZE TRAVEL TIME RELIABILITY IN A RURAL SETTING ........................................ 64  
   REFERENCES .................................................................................................................. 78

4. PRIVACY CONSIDERATIONS .......................................................................................... 79  
   INTRODUCTION .............................................................................................................. 79  
   ELECTRONIC TOLL TAG-BASED DATA COLLECTION .................................................. 79  
   BLUETOOTH-BASED DATA COLLECTION .................................................................. 81  
   APPLICATION OF PRIVACY PRINCIPLES .................................................................... 82

5. LESSONS LEARNED ....................................................................................................... 84  
   OVERVIEW ..................................................................................................................... 84  
   METHODOLOGICAL EXPERIMENTS .............................................................................. 84  
   USE CASE ANALYSIS ................................................................................................... 85  
   PRIVACY CONSIDERATIONS ......................................................................................... 86
1. MONITORING SYSTEM

STUDY DESCRIPTION

This case study is the third of five, performed by the project team in order to validate the approaches to travel time reliability monitoring described in the Travel Time Reliability Monitoring Guidebook. The goal of each case study is to illustrate how agencies apply best practices for: monitoring system deployment; travel time reliability calculations; and agency use and analysis of the system. To accomplish this goal, the team is implementing prototype travel time reliability monitoring systems at each of the five sites. These systems take in sensor data from a variety of transportation networks, process this data inside a large data warehouse, and generate reports on travel time reliability for agencies to help them better operate and plan their transportation systems. This case study consists of the following sections:

- Monitoring System
- Methodological Experiments
- Use Case Analysis
- Privacy Considerations
- Lessons Learned

These sections map to the master system components, as shown below in Figure 1-1.

![Figure 1-1: Travel Time Reliability System Description](image)

This monitoring system description section details the reasons for selecting the Lake Tahoe region as a case study and provides an overview of the region. It briefly summarizes agency monitoring practices, discusses the existing sensor network, and describes the software system that the team used to analyze the use cases. Specifically, it describes the steps and tasks that the research team completed in order to transfer data from the data collection systems into a travel time reliability monitoring system.
The section concerning *methodological experiments* describes the manner in which different types of filtering techniques might be applied at different stages of the analytical process to further refine the travel times estimates generated from Bluetooth-based datasets.

*Use cases* are less theoretical, and more site specific. The first two use cases assess the impact of detector network configuration on the data ultimately available for use by travel time reliability monitoring systems. The third use case attempts to quantify the impact of adverse weather and demand–related conditions on travel time reliability using data derived from the Bluetooth and electronic toll collection-based systems deployed in rural areas as part of this case study.

The section on *Privacy Considerations* addresses the challenges associated with collecting data using toll tag and Bluetooth-based technologies in a manner that respects the privacy of the individuals from whom the data is being collected.

*Lessons Learned* summarizes the lessons learned during this case study, with regard to all aspects of travel time reliability monitoring: sensor systems, software systems, calculation methodology, and use. These lessons learned will be integrated into the final guidebook for practitioners.

**SITE OVERVIEW**

The team selected the Lake Tahoe region located in Caltrans District 3 in order to provide an example of a rural transportation network with fairly sparse data collection infrastructure. Caltrans District 3 encompasses the Sacramento Valley and Northern Sierra regions of California. Its only metropolitan area is Sacramento. The District DOT is responsible for maintaining and operating 1,410 centerline miles and 4,700 lane-miles of freeway in eleven counties. District 3 includes urban, suburban, and rural areas, including areas near Lake Tahoe where weather is a serious travel time reliability concern and there is heavy recreational traffic. The District also contains 64 lane-miles of HOV lanes, with more than 140 more lane-miles proposed, all within the greater Sacramento region. Two major interstates pass through the District, Interstate-80, which travels from east to west, and Interstate 5, which travels from north to south. Other major freeway facilities include US-50, which connects Sacramento and South Lake Tahoe, and SR-99.

Built in 2000, the District 3 Regional Traffic Management Center (RTMC) is located in Rancho Cordova, 15 miles east of Sacramento. The RTMC serves as the focal point for traffic information within District 3. RTMC staff are responsible for managing [1]:

- Regional network of sensors, cameras, CMS, HARs, and RWIS
- Delivery of traveler information
- Dispatch of other Caltrans resources

As mentioned above, weather-related conditions contribute to serious travel time reliability concerns in District 3, including [2]:

- Fog/Visibility – The region is prone to thick ‘tule’ fog during periods after heavy rain;
- High Winds - Several bridges in the District are exposed to high winds;
- Frost/Ice – Freezing can occur on longer viaduct sections during cold weather; and,
• Snow in Sierras - High winds combined with snow accumulation create white out conditions over mountain roadways.

Caltrans and its regional partners are pursuing the creation of Corridor System Management Plans (CSMPs) for the most heavily congested transportation corridors in the region, “aimed at increasing transportation options, reducing congestion, and improving travel times. A CSMP is a comprehensive, integrated management plan for increasing transportation options, decreasing congestion, and improving travel times in a transportation corridor. A CSMP includes all travel modes in a defined corridor – highways and freeways, parallel and connecting roadways, public transit (bus, bus rapid transit, light rail, intercity rail) and bikeways, along with intelligent transportation technologies. CSMP success is based on the premise of managing a selected set of transportation components within a designated corridor as a system rather than as independent units. Each CSMP identifies current management strategies, existing travel conditions and mobility challenges, corridor performance management, planning management strategies, and capital improvements. In District 3, six CSMPs have been developed along I-80, I-5/SR-99, US-50, SR 99 North, SR 49, and SR 65.” [3]

SENSORS

Caltrans District 3 currently only collects traffic data along freeway facilities. It operates a total of 2,251 point detectors (either radar or loop detectors) located in over 1,000 roadway locations across the District. Point detection infrastructure in the mountainous regions of the District is sparser, with detectors often miles apart. To supplement the point detection network in rural portions of the Sierra Nevada Mountains near Lake Tahoe, the District has installed electronic toll collection (ETC) readers on I-80 and Bluetooth-based data collection readers along I-5 and US-50 (see Figure 1-2). These readers register the movement of vehicles equipped with FasTrak tags (Northern California’s ETC system) and Bluetooth-based devices (e.g., Smart Phones) for the purpose of generating roadway travel times.
Both ETC and Bluetooth-based data collection technologies utilize vehicle identification technologies to record the presence of vehicles as they pass instrumented points along a roadway. Field controllers typically record location, time, and vehicle identification information for each vehicle to support the calculation of travel times. By knowing the length of the road segment between two instrumented points, and the starting and ending times at which travel between those points took place, the travel time for that section of roadway can be determined.

**Table 1-1: Breakdown of Deployed ETC Readers**

<table>
<thead>
<tr>
<th>Figure 1-3 ID</th>
<th>Roadway / Direction of Travel</th>
<th>ETC Reader ID</th>
<th>Nearest Crossroad</th>
<th>Postmile</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETC 1</td>
<td>I-80 E</td>
<td>42003</td>
<td>Auburn</td>
<td>123.1</td>
</tr>
<tr>
<td>ETC 2</td>
<td>I-80 W</td>
<td>42035</td>
<td>Baxter</td>
<td>148.5</td>
</tr>
<tr>
<td>ETC 3</td>
<td>I-80 W</td>
<td>42041</td>
<td>Kingale</td>
<td>168.0</td>
</tr>
<tr>
<td>ETC 4</td>
<td>I-80 E</td>
<td>42036</td>
<td>Rainbow</td>
<td>168.1</td>
</tr>
<tr>
<td>ETC 5</td>
<td>I-80 E</td>
<td>42042</td>
<td>Rest Area</td>
<td>176.2</td>
</tr>
<tr>
<td>ETC 6</td>
<td>I-80 E</td>
<td>42044</td>
<td>Donner Lake</td>
<td>179.9</td>
</tr>
<tr>
<td>ETC 7</td>
<td>I-80 W</td>
<td>42006</td>
<td>Prosser Village</td>
<td>189.0</td>
</tr>
<tr>
<td>ETC 8</td>
<td>I-80 W</td>
<td>42015</td>
<td>Hirschdale</td>
<td>193.4</td>
</tr>
</tbody>
</table>
Table 1-2: Breakdown of Deployed Bluetooth Readers

<table>
<thead>
<tr>
<th>Figure 1-3 ID</th>
<th>Roadway / Direction of Travel</th>
<th>Bluetooth Reader ID</th>
<th>Nearest Crossroad</th>
<th>Postmile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluetooth 1</td>
<td>I-5 N</td>
<td>1005</td>
<td>Elk Grove</td>
<td>506.4</td>
</tr>
<tr>
<td>Bluetooth 2</td>
<td>I-5 N</td>
<td>1011</td>
<td>Pocket</td>
<td>511.5</td>
</tr>
<tr>
<td>Bluetooth 3</td>
<td>I-5 S</td>
<td>2101</td>
<td>Florin</td>
<td>512.4</td>
</tr>
<tr>
<td>Bluetooth 4</td>
<td>I-5 S</td>
<td>2009</td>
<td>Gloria</td>
<td>513.5</td>
</tr>
<tr>
<td>Bluetooth 5</td>
<td>I-5 N</td>
<td>1039</td>
<td>Vallejo</td>
<td>517.2</td>
</tr>
<tr>
<td>Bluetooth 6</td>
<td>I-5 N</td>
<td>1004</td>
<td>L St.</td>
<td>518.9</td>
</tr>
<tr>
<td>Bluetooth 7</td>
<td>US-50 E</td>
<td>1054</td>
<td>Placerville</td>
<td>48.4</td>
</tr>
<tr>
<td>Bluetooth 8</td>
<td>US-50 E</td>
<td>2055</td>
<td>Twin Bridges</td>
<td>87.1</td>
</tr>
<tr>
<td>Bluetooth 9</td>
<td>US-50 W</td>
<td>2058</td>
<td>Echo Summit</td>
<td>94.9</td>
</tr>
<tr>
<td>Bluetooth 10</td>
<td>US-50 E</td>
<td>2056</td>
<td>Meyers</td>
<td>98.7</td>
</tr>
</tbody>
</table>

ETC-based Data Collection in District 3

The ETC-based data collection infrastructure deployed along I-80 consists of eight (8) Fastrak toll tag reader stations, installed and operated by Caltrans District 3. The readers were initially installed to provide the Bay Area’s 511 system with travel times to Lake Tahoe, but have not yet been used for that purpose.

According to Caltrans, each reader is either mounted on an overhead Changeable Message Sign (CMS) or other fixed overhead sign. Each reader station consists of a cabinet mounted to the sign pole, which is connected to antennas mounted on the edge of the sign closest to the roadway; directed such that they monitor traffic in each lane of travel. All of the readers are deployed at roadway sections that have two lanes of travel in each direction, with the exception of one location, where there are three lanes of travel in each direction. ETC transponders passing these readers are each encoded with a unique identification number. Data from these transponders is collected via Dedicated Short-Range Communication (DSRC) radio by the reads and assigned time/date stamps, as well as an antenna identification stamp for use in calculating travel time.

Bluetooth-based Data Collection in District 3

This case study also leverages data from Bluetooth readers (BTRs) deployed on I-5 in Sacramento and along US-50 between Placerville and Lake Tahoe; these BTRs were installed by Caltrans’ research division.

From a travel time data collection standpoint, Bluetooth readers are typically placed on the side of a roadway, ideally at a vehicle windshield height or higher to minimize the obstructions between the reader and the in-vehicle Bluetooth-enabled devices. In Caltrans’ case, each BTR was mounted inside an equipment cabinet strapped to poles along the freeway.

The BTRs deployed by Caltrans used the standard Bluetooth device inquiry algorithm, scanning all 32 available channels every 5.12 seconds (split into two 2.56 phases of 16 channels each). Each Bluetooth reader records the unique Media Access Control (MAC) address generated by every Bluetooth device it detects during each scan cycle for use in calculating travel time.
DATA MANAGEMENT

The primary data management software system in the District 3 region is Caltrans’ Performance Measurement System (PeMS). All Caltrans districts use PeMS for data archiving and performance measure reporting. PeMS integrates with a variety of other systems to obtain traffic, incident, and other types of data. It archives raw data, filters it for quality, computes performance measures, and reports them to users through the web at various levels of spatial and temporal granularity. It reports performance measures such as speed, delay, percentage of time spent in congestion, travel time, and travel time reliability. These performance measures can be obtained for specific freeways and routes, and are also aggregated up to higher spatial levels such as county, district, and state. These flexible reporting options are supported by the PeMS web interface, which allows users to select a date range over which to view data, as well as the days of the week and times of the day to be processed into performance metrics. Since PeMS has archived data for Caltrans dating back to 1999, it provides a rich and detailed source of both current travel times and historical reliability information.

PeMS integrates, archives, and reports on incident data collected from two different sources: the California Highway Patrol (CHP) and Caltrans. CHP reports current incidents in real-time on its website. PeMS obtains the text from the website, uses algorithms to parse the accompanying information, and inserts it into the PeMS database for display on a real-time map, as well as for archiving. Additionally, Caltrans maintains an incident database, called the Traffic Accident Surveillance and Analysis System (TASAS), which links to the highway database so that incidents and their locations can be analyzed. PeMS obtains and archives TASAS incident data via a batch process approximately once per year. Incident data contained in PeMS has been leveraged to validate use cases associated with how different sources of congestion impact travel time reliability.

PeMS also integrates data on freeway construction zones from the Caltrans Lane Closure System (LCS), which is used by the Caltrans districts to report all approved closures for the next seven days, plus all current closures, updated every 15 minutes. PeMS obtains this data in real-time from the LCS, displays it on a map, and lets users run reports on lane closures by freeway, county, district, or state. Lane closure data in PeMS was used in the validation of the use cases associated with how different sources of congestion impact travel time reliability.

SYSTEMS INTEGRATION

Data Acquisition Prior in Support of Travel Time Reliability Analysis

PeMS can calculate many different types of performance measures; and as such, the requirements for linking PeMS with an existing system depend on the features being used. The following bullet points describe the basic data that PeMS requires from the source system to support these functions:

- Metadata on the roadway linework of facilities being monitored
- Metadata on the detection infrastructure, including the types of data collected and the locations of equipment (configuration)
- Real-time traffic data in a constant format at a constant frequency (such as every 30-seconds or every minute)

Traffic data are generally unusable for travel time calculation purposes if not accompanied by a detailed description of the configuration of the system. Configuration information provides the contextual and spatial information on the sensor network needed to make sense of the real-time data. Ideally, these two types of information should be transmitted separately (i.e., not in the same file or data feed). Roadway and equipment configuration information is more static than traffic data, as it only needs to be updated with changes to the roadway or the detection infrastructure. Keeping the reporting structure for these two types of information separate reduces the size of the traffic data files, allowing for faster data processing, better readability, and lower bandwidth cost for external parties who may be accessing the data through a feed.

To represent the monitored roadway network and draw it on maps, PeMS requires Geographic Information System (GIS) type roadway polylines defined by latitudes and longitudes. To help the agency link PeMS data and performance metrics with their own linear referencing system, PeMS also associates these polylines with state roadway mileposts. In most state agencies, mileposts are a reference system used to track highway mileage and denote the locations of landmarks. Typically, these mileposts reset at county boundaries. In cases where freeway alignments have changed over time, it is likely that the difference between two milepost markers no longer represents the true physical distance down the roadway. For this reason, PeMS adds in a third representation of the roadway network, called an absolute postmile. These are akin to mileposts, but they represent the true linear distance down a roadway, as computed from the polylines. They do not reset at county boundaries, in order to facilitate the computation of performance metrics across long sections of freeway. In PeMS, this information is ultimately stored in a freeway configuration database table that contains a record for every 10th of a mile on every freeway. Each record contains the freeway number, direction of travel, latitude and longitude, state milepost, and absolute postmile.

PeMS also requires metadata concerning the detection equipment from which the source system is collecting data. This is due to the need to standardize data collection and processing across all agencies, regardless of their source system structures. Configuration information ultimately populates detector, station, and controller configuration database tables in PeMS, and is used to correctly aggregate data and run equipment diagnostic algorithms.

Finally, the data acquisition step often involves reconciliation between the framework of the source system and the monitoring system. For example, different terminology can lead to incorrect interpretations of the data. As such, this step often requires significant communication between the system contractor and the agency staff who have familiarity with the data collection system, in order to resolve open questions and make sure that accurate assumptions are being made.

Integration of District 3 Case Study Data Sources into PeMS

Keeping the above in mind, the two sources of data utilized in support of this case study, based on the movement of vehicles equipped with Electronic Toll Collection (ETC) and Bluetooth devices, are extremely new and not currently integrated into Caltrans District 3’s existing PeMS data feed. Consequently, it was necessary to ingest these data sets into project–specific instances of PeMS for analysis as part of this project. This
section provides an overview of the resources needed to conduct the pre-requisite data collection through monitoring system integration-related activities, as well as discusses some of the challenges likely to be encountered when developing such a monitoring system. Such activities included:

- **ETC Data** - With the Tahoe area ETC data, the goal was to use pre-existing PeMS ETC processing and equipment configuration software, as well as the road network definitions in use by PeMS for Caltrans. This effort proved to be fairly straightforward and no special accommodations were required, other than dealing with detectors that would occasionally go off-line during real-time collection. As per public agency policy, all individual toll tag identifiers from the ETC readers were deleted every 24 hours.

For each ETC reader station, the research team was provided with information regarding the county in which it was deployed, freeway on which it was located, a direction of travel for which it was collecting data, milepost, textual location, and the Internet Protocol (IP) address used to communicate with it to obtain data. To integrate each reader into PeMS so that data could be collected in real-time, the research team assigned each reader a unique ID and determined its latitude and longitude. Software was then developed to communicate with each reader’s IP address, obtain its data, and incorporate that data into the PeMS database.

- **Bluetooth Data** – With the Lake Tahoe area Bluetooth data, the goal was to configure PeMS so that the Bluetooth readers and data they produced could be utilized as if it was from standard ETC reader stations. For each BTR, the research team received configuration data in a text file, with fields for the node (reader) ID, a textual location, and a latitude/longitude. Configuration data was provided for a total of 26 Bluetooth readers. Caltrans also provided the research team with a 2 gigabyte SQL file containing all of the Bluetooth data collected by the BTRs between December 25, 2010 and April 21, 2011. The research team subsequently integrated this data into PeMS and processed it to compute travel times between each pair of BTRs.

### Analyzing ETC and Bluetooth Data

PeMS collects sensor data, either by directly polling each detector, obtaining it from an existing data collection system, or via integration of data from another archival resource, and stores it in an Oracle database. Reliability measures available based on this data will depend on the type of detector from which it has been collected – e.g., loop detectors will provide different raw data for analysis than ETC or Bluetooth-based data collection systems. Reliability metrics available in PeMS based on data from the ETC and Bluetooth systems are as follows:

- **Min** - The fastest vehicles that traveled across a roadway segment during a given period of time.
- **25th**: The 25th percentile travel time during a given period of time.
- **Mean**: The mean travel time during a given period of time.
- **Median**: The median travel time during a given period of time.
- **75th**: The 75th percentile travel time during a given period of time.
Max: The slowest moving vehicles that traveled across a roadway segment during a given period of time. It is likely that much (if not all) of this data is composed of outliers that made at least one stop between two consecutive readers before completing their trip.

Each of the reliability measures described above is available for analysis based on five-minute and hourly time periods.

As stated above, the research team utilized pre-existing PeMS ETC processing and equipment configuration software to support the development and deployment of ETC and BTR instances of PeMS. Existing PeMS analysis tools create reports of travel time versus starting time. For a given starting (or source) tag reader, the travel time to a destination tag reader is defined as the amount of time it takes for a specific tag to be seen at the destination tag reader. Due to public agency policy, PeMS does not store travel times for individual ETC tag reads, only recording summary statistics for all of the tags that traversed the distance between each consecutive pair of readers during a given period of time. That said, similar regulations do not currently exist regarding the use of data collected from Bluetooth devices. As such, the research team had access to a much wider variety of raw and summary data concerning the movement of Bluetooth-enabled vehicles for use as part of this case study.

It is important to note that the algorithm currently used by PeMS to calculate travel times based on ETC and Bluetooth data is fairly simple, with its only purpose being to identify travel times for vehicles that pass between consecutive readers regardless of whether the resultant travel time makes logical sense. For example, there is no way of knowing if a given vehicle got off the freeway in between reader stations. We only know when they were seen at each station. As a result, the travel times produced by PeMS based on this data have the potential to be significantly influenced by outliers and can at times be quite “noisy.”

Lastly, there are two key differences between the ETC and Bluetooth technologies that needed to be accounted for as part of the research team’s efforts to utilize the BTR data available as part of this project:

- **Directionality** – ETC detectors are aimed in such a way as to sense traffic flowing in a particular direction. In most cases, well over 95% of data collected by an ETC device is from traffic flowing in the direction that the detector is anticipated to be measuring. The Bluetooth readers do not have this directional bias. Both ETC and Bluetooth readers are capable of recording the presence of a single vehicle multiple times as it passes through the reader’s detection zone. In the case of ETC readers, a vehicle is seldom detected more than twice, due to the limited range and directionality (aimed down on a spot on the road, not parallel to the ground) of the ETC antenna. However, Bluetooth readers can record any device generating a Bluetooth signal within its sensing radius, sometimes from 100 meters away. This can result in a single Bluetooth device being detected many times as it passes through the reader’s detection zone, especially in cases where it is traveling slowly or is stopped.

Keeping the above in mind, PeMS expects data to come from devices that have a directional bias. To accommodate this issue, the research team configured PeMS to view each Bluetooth reader as generating data for two directions of travel and fed the data into PeMS twice, assigning it first to one detector in one direction of
travel and then assigning a copy of that data to the other direction of travel as well.

- **Background noise** - Several Bluetooth readers deployed as part of this project are located within a few dozen meters of office buildings, homes, or parking lots. Consequently, there are many stationary (or nearly, so) Bluetooth devices residing within these locations that produce a reading every few seconds for hours on end. This data has the potential to overwhelm legitimate vehicular data, sometimes by as much as a factor of 10 times or greater. The research team’s initial solution for dealing with this issue in order to generate roadway travel times for analysis was to eliminate all subsequent reports of unique Bluetooth media access control (MAC) addresses collected within one hour of its initial reporting.

Additional information concerning activities undertaken by the project team to optimize the usefulness of these data sets is contained in the section entitled “METHODOLOGICAL EXPERIMENTS” and the first two use cases.

**REFERENCES**


2. METHODOLOGICAL EXPERIMENTS

OVERVIEW

Due to the significant amounts of Bluetooth-based travel time data available for analysis as part of this case study, the research team elected to focus its methodological efforts on this dataset rather than on data generated by the ETC-based system. This stems from an awareness that Bluetooth-based systems, while new, have been rapidly embraced by a wide range of transportation agencies interested in identifying low-cost, easy to deploy solutions for collecting roadway travel times. A great deal remains largely unknown regarding the underlying nature of this data, including how filtering techniques might be applied at different stages of the analytical process to further refine generated travel times. As such, this section focuses on the evaluation of methods for identifying individual vehicle trips between Bluetooth readers, followed by a statistical analysis of procedurally generated vehicle travel times. Filtering techniques at both the procedural and statistical levels are also explored as methods for improving the quality of travel time estimates.

The primary output of this section is a methodology for obtaining filtered travel time histograms that depict the distribution of travel times within a sample of Bluetooth data. It is possible to generate parameterized probability distribution functions (PDFs) from these histograms as was done in the San Diego case study, however this step is omitted here in favor of analysis of the underlying data issues.

BLUETOOTH DEVICE DATA

Impact of Bluetooth Reader Hardware on Data Available for Analysis

The characteristics of Bluetooth device data available for analysis are determined largely by the capabilities of the Bluetooth reader (BTR) deployed at the roadside. For example, only five of the 10 BTRs deployed by Caltrans had the ability to read and store signal strength measurements for each observation. Signal strength measurements are important because they provide the ability to determine the relative distance of each Bluetooth enabled mobile device from the reader. Whether or not a specific BTR has the ability to read and report signal strength values for each mobile device depends on the nature of the BTR's Host Controller Interface (HCI). The HCI is an interface between the Bluetooth protocol stack and the device's controller hardware. BTRs that reported signal strength values were based on Linux boards using the BlueZ protocol stack, while units not reporting signal strength values used a microcontroller-based implementation.

In the case of the Bluetooth Class I devices deployed by Caltrans, which have a range (radius) of detection of approximately 100m (see Figure 2-1), knowing the signal strength of each mobile device observation can be important to accurately calculate the travel times of those devices to the next BTR. To clarify, if a vehicle is traveling at 40 MPH, it will pass through the device's full 200m detection zone in approximately 10 seconds. However, if heavy congestion is present and the BTR zone traversal speed is only 5 MPH, it will take approximately 82 seconds. In cases where BTRs are fairly close together, the accurate calculation of travel time can be significantly affected by whether or not the travel time analysis system has the ability to determine the time at which each Bluetooth device is closest to each BTR; the impact is even greater during periods of congestion when vehicles are moving slowly and generating many more observations.
This issue is underscored by the fact that within the Caltrans dataset, Bluetooth-enabled mobile devices each generated approximately 1 observation (on average) per second (see Table 2-1), resulting in a mean number of observations per mobile device, per visit to each BTR of between 1.06 and 21.30.

Figure 2-1: Bluetooth and ETC Reader Detection Zones

Figure 2-1 provides a graphical depiction of the nature of the detection zones generated by Bluetooth and ETC-based data collection technologies.
Table 2-1: BTR Detection Zone Traversal Times and Observations

<table>
<thead>
<tr>
<th>Bluetooth Reader ID</th>
<th>Mean Zone Traversal Time (Secs.)</th>
<th>St. dev. of Zone Traversal Time (Secs.)</th>
<th>Mean Observations per Visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTR 1</td>
<td>1.43</td>
<td>12.02</td>
<td>1.06</td>
</tr>
<tr>
<td>BTR 2</td>
<td>0.88</td>
<td>9.88</td>
<td>1.10</td>
</tr>
<tr>
<td>BTR 3</td>
<td>1.27</td>
<td>25.45</td>
<td>1.16</td>
</tr>
<tr>
<td>BTR 4</td>
<td>4.20</td>
<td>48.56</td>
<td>1.44</td>
</tr>
<tr>
<td>BTR 5</td>
<td>0.58</td>
<td>10.91</td>
<td>1.09</td>
</tr>
<tr>
<td>BTR 6</td>
<td>7.93</td>
<td>48.01</td>
<td>1.38</td>
</tr>
<tr>
<td>BTR 7</td>
<td>8.77</td>
<td>37.49</td>
<td>11.49</td>
</tr>
<tr>
<td>BTR 8</td>
<td>8.73</td>
<td>65.38</td>
<td>11.44</td>
</tr>
<tr>
<td>BTR 9</td>
<td>5.77</td>
<td>36.78</td>
<td>8.19</td>
</tr>
<tr>
<td>BTR 10</td>
<td>23.10</td>
<td>116.65</td>
<td>21.30</td>
</tr>
</tbody>
</table>

Table 2-2 provides examples of mean, maximum, and standard deviation of mobile device signal strengths collected by the various BTRs involved in this study. BTRs with signal strength characteristics noted as “N/A” did not have the capability to collect signal strength data. Signal strength readings are proportional to the distance between a BTR and each mobile device. A BTR’s mean signal strength is therefore a function of the location of the BTR relative to the roadway. In addition, BTR antenna gain varies as a function of manufacturer and type which affects mean signal strength [1].
Figure 2-2 compares observed signal strengths over time for 3 vehicles traveling through BTR detection zones; each plot is centered (from a temporal perspective) on the time at which the peak signal strength was detected for each vehicle. The first vehicle arrives in the detection zone, travels past the reader and stops for approximately 11 minutes within the detection zone. The second vehicle passes through the detection zone in approximately 17 seconds, traveling at 24 MPH. The third vehicle enters the BTR’s detection zone, pauses for approximately 18 seconds, passes the BTR, and then departs the detection zone.

<table>
<thead>
<tr>
<th>Bluetooth Reader ID</th>
<th>Number of Observations</th>
<th>Mean Signal Strength</th>
<th>Maximum Signal Strength</th>
<th>Signal Strength St. Dev.</th>
</tr>
</thead>
</table>

Table 2-2: BTR Signal Strength Characteristics
Bluetooth device data collected as part of this case study exhibited a number of characteristics that should be understood prior to attempting the calculation of roadway travel times; these are discussed below.
Device Visiting Only One BTR. One way to classify mobile devices is by the total number of unique BTRs they visit. For the purposes of calculating segment (BTR to BTR) travel times, observations generated by devices that visit only a single BTR can be ignored. Based on the team’s analysis, approximately 29% of all mobile devices represented in the Caltrans dataset visited only a single BTR during a given trip; these devices contributed 12.5% of all mobile device observations (Table 2-3).

Table 2-3: Observations Generated by Devices – By Number of BTRs Visited

<table>
<thead>
<tr>
<th>Number of Devices</th>
<th>Visited 1 BTR</th>
<th>Visited &gt; 1 BTR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Devices</td>
<td>146,075 (29%)</td>
<td>356,408 (71%)</td>
<td>502,483 (100%)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>2,315,389 (13%)</td>
<td>16,176,143 (88%)</td>
<td>18,491,532 (100%)</td>
</tr>
</tbody>
</table>

Variable BTR Detection Zone Traversal Times. As discussed above, mobile devices take varying amounts of time and generate unpredictable numbers of observations each time they pass through a given BTR’s detection zone. Generally, the number of observations generated by a device is proportional to the amount of time the vehicle is present within the detection zone; which is proportional to the vehicle’s speed. Based on analysis conducted as part of this case study, the research team believes that the “Mean Zone Traversal Time” (see Table 2-1) is affected by a combination of the physical location of the reader relative to the roadway and other roadway characteristics. For example, BTR #2 (see Table 1-2 for the location of each BTR) is located at the end of an entrance ramp and is isolated from nearby arterials and buildings. It has a mean detection zone traversal time of .88 seconds with approximately 1.10 observations per visit. This can be seen in the zone traversal time frequency distribution (top distribution in Figure 2-) with no delay time for vehicles passing through the detection zone. This reader contrasts with BTR #10, which has a mean detection zone traversal time of 23.1 seconds and 21.3 observations per visit (bottom distribution Error! Reference source not found.). This reader is located on one leg of a T-intersection with a single stop sign. Consequently, cars queuing at the stop sign may be contributing significantly to the long zone traversal times.
Multiple Mobile Device Observations Per BTR. Individual mobile devices can enter and exit a single BTR's detection zone multiple times during a sufficiently lengthy period of time. Depending on the size of the window of time, these individual observations have the potential to be matched with a significant number of observations from other BTRs. Table 2-4 displays the results of one vehicle visiting BTR #10 four times during one day. The final 2 visits are separated by just 10 minutes (the 3rd and 4th visit are shown in Figure 2-44. This demonstrates that a travel time algorithm that processes device observation data must have the ability to aggregate and differentiate between clouds of such observations separated in time as a step in the process of calculating travel times between BTRs.

Table 2-4: Multiple Device Observations for One Device at BTR #10

<table>
<thead>
<tr>
<th>Visit Number</th>
<th>Time</th>
<th>Number of Observation</th>
<th>Time Delta (Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>09:50 am</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>15:00 pm</td>
<td>5</td>
<td>2.54</td>
</tr>
<tr>
<td>3</td>
<td>16:33 pm</td>
<td>2</td>
<td>18.05</td>
</tr>
<tr>
<td>4</td>
<td>16:43 pm</td>
<td>39</td>
<td>42.98</td>
</tr>
</tbody>
</table>
CALCULATING TRAVEL TIMES BASED ON BLUETOOTH DEVICE DATA

The primary goal of BTR-based data analysis is to characterize segment travel times between BTRs based on the re-identification (re-id) of observations derived from unique mobile devices. Generally, the data processing procedures associated with the calculation of BTR to BTR travel times can be broadly broken down into 3 processes, as shown below. The first two processes are procedural. The third is statistical:

1. Identification of Passage Times
   A. Aggregating device observations into visits
   B. Selecting BTR passage time
2. Generation of Passage Time Pairs
   A. Method 1: Maximum origin and destination permutations
   B. Method 2: Use of all origin visits
   C. Method 3: Aggregation of visits
3. Generation of Segment Travel Time Histograms
   A. Filter outliers across days
   B. Filter outliers across time intervals
   C. Remove intervals with few observations
   D. Remove highly variable intervals

The steps involved in these 3 processes are discussed below.

**Process 1: Identification of Passage Times**

The first step in the process of calculating segment travel time PDFs for a roadway is the calculation of segment travel times for individual vehicles. A vehicle segment travel
time is calculated as the difference between the vehicle’s Passage Times at both the origin and destination BTRs. Passage Time is defined as the single point in time selected to represent when a vehicle passed through a BTR’s detection zone. As previously discussed, mobile devices typically generate multiple observations as they pass through a BTR’s detection zone. Consequently, selection of appropriate passage times is an important step in maximizing the accuracy of calculated segment travel times for individual vehicles.

Aggregating Device Observations into Visits. The goal here is to identify clusters of observations that represent a vehicle’s continuous presence in the detection zone. Each group of observations is referred to as a visit. For example, Error! Reference source not found. displays numerous observations during a single vehicle’s visit over the course of several minutes. Figure 2-4 displays two separate visits by a single vehicle (each with multiple observations) separated in time by a stop outside of the detection zone. Identifying unique visits is an important step in increasing the accuracy of segment travel time calculations. Associating multiple observations clustered in time as part of a single visit rationalizes the selection of a single passage time for calculating the vehicle’s travel time to a destination BTR. The alternative, which makes little sense, would be to calculate a travel time for each origin observation to the destination BTR. Identifying visits also enables the assessment of arrival (first mobile device observation) and departure (last mobile device observation) times for each mobile device, which can, depending on the circumstances, be used as the passage time. Figure 2-5 depicts clusters of observations associated with two distinct visits by a single vehicle to the same BTR.

The method used to aggregate visits is a causal sliding time window filter. It is a filter in that it removes unnecessary observations during the aggregation process. It is causal in that it uses only past and present observations to support its decision-making. This filter discards all subsequent observations that are within a fixed time span from the time of a prior observation. However, when the time between an observation and the prior observation with which it is being compared exceeds this time span, it is considered to be part of a new visit. This has the effect of aggregating observations into visits by arrival time (or departure time, depending on how it is implemented) and discarding all other
observations. Figure 2-3 displays 11 observations that have been aggregated into 2 visits due to a sufficiently large time gap between the 5th (part of visit #1) and 6th (part of visit #2) observations. This filter is an efficient method of processing real-time observations and compressing large quantities of observation data for efficient storage.

The size of the filter interval time depicted in Figure 2-3 determines the granularity of identified visits. The effect of different sized interval times is shown in Figure 2-. In general, selecting the largest reasonable interval time is desirable because it results in more accurate estimates for arrival and departure times (and hence, passage times, depending on the method used). However, over-aggregating visits is potentially problematic. The research team has identified the following error types to consider when selecting an interval time:

- **Observation over-aggregation**: When observations belonging to multiple visits are incorrectly aggregated as a single visit, the arrival passage time and departure passage time may be calculated as too early or too late, depending on the method used. This may also result in the classification of stopped non-delay time as stopped delay time because the vehicle is incorrectly identified as being continuously in the detection zone. For example, if a filter time interval of 20 minutes is used and the vehicle leaves the detection area and returns 10 minutes later, this 10-minute absence would be classified as having been spent within the zone. If the distance to adjacent BTRs is close, over-aggregation risks subsuming valid origin visits, resulting in the deletion of valid trips. For these reasons, under-aggregation is preferred to over-aggregation.

- **Observation under-aggregation**: Incorrectly sub-dividing observations from a single visit into multiple visits may result in the incorrect calculation of passage time, depending on the method used. Under aggregation is less problematic because multiple sequential visits that are not interwoven with visits to other BTRs can be aggregated and considered to be a single visit (discussed below).
The Caltrans Bluetooth data was processed by storing the arrival, departure, and maximum signal strength (where available) for each identified visit. Observations were aggregated using a 120-second time window. The 120-second window size was selected due to the small distance between BTRs #9 and #10 (about 3.8 miles) and the preference for under-aggregating visits. Moreover, using a smaller, 60-second time interval was found to be under-aggregating observations in too large a number of cases. Other researchers appear to be using a 5-minute interval [2], which may be appropriate for large BTR to BTR distances. When deploying permanent travel time data collection systems based on BTR (or related) technologies, it is likely that the filter interval should be adjusted for each BTR as a function of its location and the characteristics of the surrounding region. For example, if a snow chain fitting area is nearby, longer interval times may be optimal.

Vehicles that are continuously within a BTR’s detection zone (and generating observations) are either in travel mode (e.g. driving, in congestion, at a stop light) or trip mode (e.g. stopped at a fuel station, parked at the side of the road). Without more information, distinguishing between trip and travel behavior within a single visit is difficult. In contrast, distinguishing between trip and travel behavior across multiple visits is possible. Repeat visits to the same BTR (without visiting any other BTR) can be assumed to be non-travel time oriented and therefore eliminated. For example, if the vehicle in
Table 2-4 did not visit other BTRs between 09:50 am and 16:43 pm, then these visits can be eliminated from travel time calculations.

**Selecting BTR Passage Time.** The precise methodology used to determine a vehicle’s passage time depends on the availability of signal strength data, the distance to adjacent BTRs and traffic flow patterns in the area surrounding a BTR. When signal strength data is available, passage time can be considered as corresponding to the mobile device observation with the greatest signal strength. In cases where signal strength data is not available and the distance to adjacent BTRs is large, the arrival, mean, or departure time may be used as the passage time without introducing a significant bias. However, if traffic through the detection zone is subject to stop delay time (e.g., traffic signals, stop signs, congestion, etc.) then use of arrival or departure times may either introduce or eliminate significant bias. This is illustrated with BTRs #9 and #10, below.

BTR #10 provides an example of how the use of arrival vs. departure times as a proxy for passage time (in cases where no signal strength data is available) can influence the calculation of segment travel times. BTR #10 is located on one leg of a T-intersection with a single stop sign, as shown in Figure 2-. Its nearest neighboring BTR is 3.8 miles away. The mean detection zone traversal time for BTR #10 is 23.1 seconds, which is likely due at least in part to vehicles queuing at the nearby stop sign. Vehicles queued at the stop sign either turn right, away from BTR #10, or turn left and pass it. The free-flow speed of traffic passing the BTR is approximately 45 MPH. At this speed, traffic passes through the detection zone in 9 seconds. As such, for vehicles not queued at the stop sign, arrival times are (on average) 4.5 seconds earlier and departure times 4.5 seconds later than when the mobile device passes the BTR. If BTR #10 is used as the point of origin for generating a segment travel time, e.g., with BTR #9 as the destination, left-turning vehicles proceeding from the stop sign will pass the reader and generate an arrival time (and consequently a passage time) that is approximately $23.1 - 4.5 = 18.6$ seconds early, or nearly 7% of the travel time to the next BTR (3.8 miles away with a free flow speed of 45 mph). This error may be further compounded by heavy traffic, causing longer queues at the stop sign within the detection zone of BTR #10. In contrast, basing vehicle passage time on departure time, thereby removing the delay associated with the presence of the stop sign, would introduce only about 4.5 seconds of error, representing a substantial improvement over use of
arrival time. This example demonstrates why significant attention needs to be paid to the process used to calculate passage time in situations where signal strength data is not available.

Figure 2-8: BTR #10 Geometry in Relation to Adjacent BTR #9

In addition to considering the impact of BTR passage times, users of Bluetooth data must also consider that the accurate calculation of segment travel time is a function of the relationship between BTR-to-BTR distance and the maximum speed error. Following on the analysis performed by Haghani, et. al. [2], this relationship is depicted in Figure 2-4 as the maximum error in segment speed versus BTR-to-BTR distance ("Distance between Nodes") for 4 speeds. For this analysis, BTR-to-BTR distance is L, vehicle speed is S, and the travel time between adjacent BTRs is T, such that:

1) \( L = S \times T \)

2) \( L + \Delta L = (S + \Delta S)(T + \Delta T) \)

3) \( \Delta S_{\text{max}} \leq (\Delta L_{\text{max}} \cdot \Delta T \cdot S) / (L / S + \Delta T) \)

Equation #2 introduces error terms for L, S, and T. As per Equation #3, the maximum error in distance, \( \Delta L_{\text{max}} \), is assumed to be 600 ft. (the diameter of each BTR’s detection zone).

As per Figure 2-4, if time error (\( \Delta T \)) is 0, then speed error is maximized as vehicle speed increases. As a result, for BTRs spaced less than 2 miles apart collecting Bluetooth data from vehicles traveling at high rates of speed, the maximum speed error becomes quite significant. However, due to a combination of clock synchronization error and/or Bluetooth time stamp inaccuracies, it is highly unlikely that \( \Delta T \) will often (if ever) equal 0.
As per Figure 2-5, if time error ($\Delta T$) is greater than zero, then both slower, as well as faster vehicle speeds have the potential to maximize speed errors. Within the context of this graph, the influence of time errors has a tendency to negate the effect of distance errors. A time error of 4 seconds was used based on clock synchronization error; associated with Caltrans’ method for synchronizing BTRs when local time differed from network time by more than 2 seconds.

**Process 2: Generation of Passage Time Pairs**

It is common for vehicles to generate multiple sequential visits per BTR, which may be interwoven in time with visits at other BTRs (see Table 2-5). For BTRs with a significant mean zone traversal time, it is common for vehicles to generate multiple visits close in time. The motivation for grouping visits is evident in Table 2-5, where the vehicle was at the origin BTR multiple times (see rows 1-3) before traveling to the destination BTR (see row 4). Based on this data, 3 different travel times could be calculated: 1 to 4; 2 to 4; or 3 to 4. Which pair or pairs represent the most likely trip? The benefit of performing more complex analysis of visits is that many likely false trips can be eliminated, increasing the quality of the calculated travel time. Three methods of identifying segment trips are discussed below.
Method 1: The first potential method for identifying segment trips is simple: create an origin and destination pair for every possible permutation of visits, except those generating negative travel times (Figure 2-6). For example, the visits in Table 2-5 show 6 origin and 2 destination visits, resulting in 12 possible pairs. Five pairs can be discarded because they generate negative travel times. Even so, this approach will generate many passage time pairs that do not represent actual trips. Using this method,
243,777 travel times were generated between one pair of BTRs over a three-month period.

**Method 2:** The second potential method for identifying segment trips is also simple, but represents an improvement from the first method. It creates an origin and destination pair for every origin visit and the closest (in time) destination visit, as shown in Figure 2-7. Multiple origin visits would therefore potentially be paired with a single destination visit. Using this method with the data in Table 2-5 would generate 4 pairs: 1-4, 2-4, 3-4, 5-6. This method generated 60,537 travel times between the origin and destination BTRs, eliminating 183,240 (75%) potential trips compared with the first method.

**Method 3:** Vehicles frequently make multiple visits to an origin BTR before traveling to a destination BTR. The third method of eliminating invalid segment trips aggregates those origin visits that would otherwise be interpreted incorrectly as multiple trips between the origin and destination readers. This method can be described as aggregating visits at the BTR network level. This is an additional level of aggregation beyond aggregating individual observations into visits, as discussed in the previous section. Logically, a single visit represents a vehicle's continuous presence within a BTR's detection zone. In contrast, multiple visits aggregated into a single grouping represent a vehicle's continuous presence within the geographic region around the BTR, as
determined by the distance to adjacent BTRs. This method is an example of using knowledge of network topology to identify valid trips.

This method can be applied to the data displayed in Table 2-5, which shows 3 origin visits in rows 1, 2, and 3. The question is whether any of these visits can be aggregated or should each be considered a valid origin departure? The distance from the origin (BTR #7) to the destination (BTR #10) is 50 miles (or 100 miles for the round-trip). Driving at some maximum reasonable speed (for that road segment, anything over 80 MPH is unreasonable) a vehicle would take 76-minutes for the round-trip. Therefore, if the time between visits at the origin is less than 76-minutes, they can be aggregated and considered as a single visit. In table 2-5, visits 2 and 3 (rows 2 and 3) meet this criterion and can therefore be aggregated (Figure 2-13). This eliminates 1 of 3 potential origin visits that could potentially be paired with the destination visit in row 4. Again, the idea is to identify when the vehicle was continuously within the geographic region around the origin BTR and eliminate departure visits wherever possible. When this method was applied to the data set, it generated 39,836 travel times, eliminating 20,701 (34%) potential trips compared with the second method discussed above.

Figure 2-8: Trips Generated from Aggregating Origin Visits

Additional filters could be used to identify and eliminate greater numbers of trips. For example, an algorithm could take advantage of graph topology and interspersed trips to other BTRs to aggregate larger numbers of visits. In addition, the algorithm could potentially track which destination visits had previously been paired with origin visits, eliminating unlikely trips. If PDFs are developed based on historical data, selection among multiple competing origin visits paired with a single destination visit could be probabilistic. These are potential topics for future research.

**Process 3: Generation of Segment Travel Time Histograms**

Previous sections described methods for determining travel times based on Bluetooth data. This was done by first identifying vehicle passage times at each of the Bluetooth readers, then pairing those passage times from the same vehicle at origin and destination locations. These techniques were developed with the goal of maximizing the validity of the travel times. However, because of Bluetooth data's susceptibility to erroneous travel time measurements, even the most careful pairing methodology will still
result in trip times (which could include stops and/or detours) that need to be filtered in order to obtain accurate ground truth travel times (the actual driving time).

This section of the methodology describes a four-step technique for filtering travel times, presents travel time histograms before and after filtering, and compares the effects of two passage time pairing techniques (“Method 2” and “Method 3” from Generation of Passage Time Pairs) on the resulting travel time histograms. The underlying parameterized travel time PDFs could be approximated from the filtered travel time histograms presented here. However, this step is omitted from this methodology section in order to more closely focus on the low-level issues associated with obtaining travel time distributions from Bluetooth data.

To begin, the distribution of raw travel times obtained from two different passage time pairing methods can be seen in Figure 2-9. The data presented here as “Method 2” was developed using the second passage time pairing method described in Generation of Passage Time Pairs. Data labeled as belonging to “Method 3” was built according to the third passage time pairing method in that same section. No “Method 1” analysis is included here due to that method’s lack of sophistication. In Figure 2-9, the unfiltered travel time distributions appear similar apart from the quantity of data present. Both distributions have extremely long tails, with most trips lasting an hour or less and many taking months. It is clear from these figures that even the carefully constructed “Method 3” data is unusable before filtering.

Several plans have been developed to filter Bluetooth data. Here, we adopt a four-step method proposed by Haghani, et. al. [2]. In Haghani’s filtering plan, points are discarded based on their statistical characteristics, such as coefficient of variation and distance from the mean. The four data filtering steps are:

1. **Filter outliers across days.** This step is intended to remove measurements that do not represent an actual trip but rather a data artifact (i.e., the case above of a vehicle being missed one day and detected the next). Here, we group the travel times by day and plot PDFs of the speeds observed in each day (rounded to the nearest integer). To filter the data, thresholds are defined based on the moving average of the distribution of the speeds (with a recommended radius of 4 miles per hour). The low and high thresholds are defined as the minima of the moving average on either side of the modal speed (i.e., the first speed on either side of the mode in which the moving average increases with distance from the mode). All speeds above/below these values are discarded (see Figure 2-15).

2. **Filter outliers across time intervals.** For the remaining steps 2-4, time intervals smaller than one full day are considered (we use both 5-minute and 30-minute intervals). In this step, speed observations beyond the range mean ±1.5σ within an interval are thrown out. The mean and standard deviation are based on the measurements within the interval.

3. **Remove intervals with few observations.** Haghani determines the minimum number of observations in a time interval required to effectively estimate ground truth speeds. This is based on the minimum detectible traffic volume and the length of the interval. Based on this, intervals with fewer than 3 measurements per 5-minutes (or 18 measurements per 30-minutes) were discarded.
4. **Remove highly variable intervals.** In Step 4, the variability among speed observations is kept to a reasonable level by throwing out all measurements from time intervals whose coefficient of variation (COV) is greater than 1.

![Figure 2-9: Unfiltered Travel Times Between One Pair of Bluetooth Readers, February 6, 2011 to February 12, 2011](image1)

To carry out Step 1, the moving average (with radius of 4 mph) is computed over the speed distribution for each day (note that speeds are found by simply dividing route length by travel time). The moving average and distribution of speeds from a single day can be seen in Figure 2-15. To exclude unreasonably low speeds, the modal speed is defined as the speed corresponding to the peak of the moving average above 20 mph (53 mph in this case). On this day, as a result of filtering Step 1, the upper threshold was set to the maximum observed speed (the minimum of the moving average above the modal speed), and the lower threshold was set to 25 mph (the minimum of the moving average below the modal speed). Thus, on this day, all data points representing speeds below 25 mph or above 62 mph were discarded as a result of Step 1.

![Figure 2-15: Distribution of Speeds](image2)
While Step 1 is carried out across days, Steps 2-4 are carried out across 5-minute and 30-minute intervals. The 5-minute interval was chosen to match what was done by Haghani [2] and represents a standard, baseline filter. Filtering results based on a 30-minute interval are also included to compare the effects of a wider filtering interval. A wider filtering interval may be more appropriate for sparse data sets like that available in the rural Lake Tahoe area where many 5-minute intervals contain no measurements at all. The particular details of steps 2-4 are more straightforward and are omitted from further discussion.

The results of Haghani’s four-step filtering method on the data obtained using passage time pairing method 2 are presented in Figure 2-16 for the week beginning on February 6, 2011. The white points are points identified to be thrown out in that step, and the black points are points to be kept following that step. Note that the steps are performed sequentially, so that points discarded after Step 1 are not considered in Step 2, and so on. Higher traffic volumes due to weekend traffic in the area can clearly be seen on Friday and Saturday.

After the data has been filtered, the travel time distributions over the week appear much more meaningful, as can be seen by comparing Figure 2-17 with Figure 2-14. The filtered travel times have lost their unreasonably long values and in both cases, a nicely shaped distribution is visible. Note that the earlier comparison of the data sets remains true: both have similarly shaped distributions, but the data prepared using passage time pairing method 2 contains a greater quantity of data. This is because that data set was larger initially, and also a larger percentage of points from it survived filtering.
Figure 2-16: Four-step filtering on passage time pairing method 2 with 5-minute intervals.
Table 2-6 presents a summary of the filtering results on both data sets using 5-minute and 30-minute intervals. It can be seen that Step 1, which removes outliers by day, takes out a much smaller percentage of the data from Method 2. This is because the data in Method 2 was prepared in a way such that the resulting data is grouped more closely, even though it was not prepared as carefully. For example, if a particular O-D pair contained three vehicle passage times at the origin and one at the destination, Method 3 would report the single travel time from the latest origin timestamp to the destination timestamp. Method 2, on the other hand, would report this as three separate travel times, all likely with similar magnitudes. Thus, the data in Method 3 is of higher quality, but is not as closely grouped, and it is penalized for this in filtering Step 1.

Additionally, Method 3, which has fewer points, is much more vulnerable to overaggressive filtering in Step 3 (which removes sparse intervals). This can be seen in the larger bands of dark gray in the method 3 columns of Figure 2-18. This is because the data sets prepared using Method 3 were much sparser initially. As a result, filtering routines that discard intervals with sparse detection may be overaggressive for sparse data sets such as those prepared by Method 3, even if the data itself is more meaningful.

Overall, data sets constructed with passage time pairing Method 2 had a higher percentage of points survive the filtering process when using both 5-minute and 30-minute intervals (see Table 2-6 and Figure 2-18), although both data sets performed more poorly when 30-minute intervals were used. This could be because the longer time intervals do not allow for quickly changing conditions such as weekend traffic congestion or adverse weather events.
Table 2-6: Comparison of Passage Time Pairing Methods

<table>
<thead>
<tr>
<th></th>
<th>5-Minute Intervals</th>
<th>30-Minute Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Method 2</td>
<td>Method 3</td>
</tr>
<tr>
<td>Total points</td>
<td>5,886</td>
<td>4,185</td>
</tr>
<tr>
<td>Removed at step 1</td>
<td>3,118 (53%)</td>
<td>2,687 (64%)</td>
</tr>
<tr>
<td>Removed at step 2</td>
<td>117 (2%)</td>
<td>20 (0%)</td>
</tr>
<tr>
<td>Removed at step 3</td>
<td>915 (16%)</td>
<td>883 (21%)</td>
</tr>
<tr>
<td>Removed at step 4</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Total points</td>
<td>4,150 (71%)</td>
<td>3,590 (86%)</td>
</tr>
<tr>
<td>Remaining points</td>
<td>1,736 (29%)</td>
<td>595 (14%)</td>
</tr>
<tr>
<td>Mean after filtering</td>
<td>58.3 min</td>
<td>58.4 min</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>5.9 min</td>
<td>5.7 min</td>
</tr>
</tbody>
</table>

Figure 2-18: Proportions of Data Points Discarded
SUMMARY

This section has evaluated various methodological approaches and processes for estimating ground-truth segment travel times based on Bluetooth data. The characteristics of Bluetooth data at each node were found to vary significantly, as a function of the surrounding roadway configuration. In cases where inter-node distances were small, the availability of signal strength was determined to be an important factor in increasing the accuracy of calculated travel times. Methods were also explored for identifying invalid segment trips, most especially via the analysis of network topology. In turn, this facilitated the generation of fewer and higher quality segment trips for use in statistical analysis.

The generation of travel time histograms used filters proposed by Haghani, et al. A comparison of two passage time pairing methods was made through histograms of filtered and unfiltered data. Potential pitfalls of using standard filtering procedures on Bluetooth data (such as discarding sparsely populated intervals) were also identified. The filtering methodology demonstrated herein was statistical in nature, in the sense that data points were discarded based on their statistical characteristics, such as coefficient of variation and distance from the mean. By comparison, passage time pairing strategies were based on the physical characteristics of the network. This exercise showed that to obtain valid travel times, knowledge of the characteristics of the network should be leveraged to the greatest extent possible, although there will still likely be a need for statistics-based filtering due to the nature of Bluetooth data.

REFERENCES


3. USE CASE ANALYSIS

This case study explores the use of two vehicle re-identification technologies in support of travel time reliability monitoring within a rural setting. These data collection technologies (ETC and Bluetooth-based) work by sampling the population of vehicles along the roadway, subsequently matching unique toll tag ids or Bluetooth MAC addresses between contiguous reader stations. Their effectiveness in accurately calculating roadway travel times is dependent on a number of factors, including:

- The percentage of the total traffic stream sampled at individual readers
- The re-identification rate between pairs of readers

In general, the percentage of the vehicle population sampled by individual readers depends on the penetration rate of the technology within the vehicular population, the positioning and mounting of the reader, and the roadway configuration at the reader’s location. The re-identification rate between pairs of readers can depend on all of the above factors, as well as the distance between readers and the likelihood of a trip diverting between the origin and destination reader. Since little can be done to increase the technology’s penetration rate when deploying a reliability monitoring system, locating, positioning, and configuring readers to maximize their collection of quality data is crucial to the success of the system.

As this case study leveraged data generated by networks of existing data collection devices, the research team could not evaluate the process for installing and configuring detection infrastructure. However, this case study did provide the opportunity to analyze the impacts of the configuration of existing ETC and Bluetooth reader networks on the nature of the data ultimately collected for use by a travel time reliability monitoring system. Based on this concept, the team developed two network configuration-related use cases. The first use case details the findings of the research team’s investigation into the configuration of the Lake Tahoe ETC network, and discusses issues including: time-of-day dependency of the toll tag penetration rate, the number of lanes that can be monitored using ETC infrastructure, and re-identification of toll tags between readers separated by different distances. The second use case details the team’s investigation into configuration-related issues associated with the Bluetooth reader network, including the relationship between reader location and the number of lanes monitored and the sample sizes measured between readers on different freeways over varying distances.

A third use case seeks to quantify the impact of adverse weather and demand–related conditions on travel time reliability using data derived from the case study’s Bluetooth and ETC-based systems deployed in rural areas. To examine travel time reliability within the context of this use case, methods were developed to generate probability density functions (PDFs) from large quantities of travel time data representing different operating conditions. To facilitate this analysis, travel time and flow data from ETC readers deployed on I-80W and Bluetooth readers deployed on I-50E and I-50W were obtained from PeMS and compared with weather data from local surface observation stations. PDFs were subsequently constructed to reflect reliability conditions along these routes during adverse weather conditions, as well as according to time-of-day and day-of-week. Practical data quality issues specific to Bluetooth and ETC data were also explored.
A. IMPACT OF ETC READER DEPLOYMENT CONFIGURATION ON DATA QUALITY

This first use case details the findings of the research team’s investigation into the impact of the configuration of the Lake Tahoe ETC network on the quality of travel time data collected.

INTRODUCTION

In this case study, the ETC detection network consisted of eight Fastrak readers located on Interstate-80 between the eastern outskirts of Sacramento and North Lake Tahoe. For each reader, Caltrans provided us with its county, freeway, a single direction of travel, mile post, a textual location, and the IP address that could be used to communicate with it and obtain its data. To place data from this network into PeMS, the research team assigned each reader a unique ID and determined its latitude and longitude from the provided mile post. Code was then written to connect with each reader’s IP address, obtaining its data once per minute for storage onto the PeMS database.

METHODOLOGY

The configuration data obtained from Caltrans was sufficient to place each reader at a location alongside the roadway. Based on that information, the team sought to answer the following questions in order to more fully understand the impacts of the network’s configuration on the characteristics of the reported travel times:

1. Are the readers where they are reported to be?
2. Are any of the readers monitoring multiple directions of travel?
3. What percentage of total traffic is being detected?
4. What percentage of toll tags is matched between pairs of readers?

The first question, which addresses where the readers are located, appears straightforward, but agencies often struggle to track detection equipment in the field. This is especially problematic with vehicle re-identification technologies, which can be easily moved from location to location. While one solution to this problem is to equip readers with GPS units, this is not common practice. The issue is compounded when multiple departments within a single agency, or multiple agencies, are using the data from these readers for their own purposes, and are not informed in a timely manner of configuration changes. To verify reader locations, the team evaluated the travel time data reported between each pair of readers to make sure that the travel times and the number of samples reported within a given time period were reasonable given the distance between the readers and the direction of travel for which they were supposed to be collecting data.

Answering the second question is important because, in some cases, ETCs can be deployed such that they monitor two directions of travel. This question was addressed by examining the roadway configuration of each reader deployment and evaluating the ETC data collected to determine whether a significant number of toll tag matches occurred between that reader and the neighboring reader in each direction of travel.

The third question addresses the “hit rate” occurring at each reader. The team calculated this by comparing hourly ETC tag reads against hourly volumes collected from nearby loop detectors. In an effort to relate mounting configuration to the percentage of traffic sampled, hit rates were subsequently compared between readers.
The final question relates to the quality of travel times being reported. As the higher the percentage of matches, the more accurate a travel time estimate is likely to be, the research team assessed the percentage of tags matched between all possible combinations of upstream and downstream ETC readers. Results from each combination of ETC readers were then compared to determine how the percentage of matches is impacted by the hit rates of each individual reader as well as the distance between readers.

**ANALYSIS**

This subsection documents the process used and analysis conducted by the team to develop answers to the aforementioned four questions.

**Are the readers where they are reported to be?**

According to Caltrans, each reader is either mounted to an overhead Changeable Message Sign (CMS) or an overhead fixed sign. Each reader consists of a cabinet mounted to the sign pole, which is connected to two antennae mounted on the edge of the sign closest to the roadway. Figure 3-1 shows a photograph (courtesy of Caltrans District 3), taken during installation, of the reader at the Donner Lake exit on eastbound I-80.

![Figure 3-1: ETC Installation](image)

Using the information provided by Caltrans, the team verified that there was a CMS or overhead sign at the latitude/longitude reported for each ETC station. Photographs of each deployment were obtained to determine each ETC’s mounting configuration, its positioning over the roadway, and the roadway geometry at that location. Photographs of each reader’s mounting structure, as indicated by Caltrans, are displayed in Figure 3-2.

The team next evaluated the minimum travel times reported between each pair of readers in order to ensure they were reasonable given the distances involved. All travel times were determined to be reasonable with the exception of trips that originated or
ended at the Kingvale reader, stated by Caltrans as being located on I80-W, adjacent to the Rainbow reader on I80-E. Results of the team’s analysis indicated that the Kingvale reader was actually located on I80-E, approximately 3 minutes downstream of the Donner Lake reader, which was later confirmed by Caltrans (see Figure 3-3).

Figure 3-2: ETC Locations
Are any of the readers monitoring multiple directions of travel?

The next step in understanding the impact of the various ETC reader configurations on the nature of the data collected was to determine whether any readers were capturing traffic in both directions of travel. The photographs in Figure 3-2 indicated that the eastbound and westbound directions of travel at the Rainbow, Rest Area, and Donner Lake reader deployments were completely separated from one another. As a result, it was not possible for these readers to monitor the opposite direction of travel. For the other readers, their ability to capture bi-directional traffic depended on the size and orientation of the detection zone generated by their antennae.

To conduct this analysis, the research team calculated the minimum and median travel times and the number of matches reported between each pair of adjacent readers monitoring opposite directions of travel along I-80. In cases where the minimum travel times reported between two readers approximated the free-flow speed given their geographic distance, and significant numbers of matches were generated that approximated that speed, the research team determined that the destination reader was likely capable of monitoring bi-directional traffic. Alternatively, if the minimum travel times were high and the number of matches low, then the matches likely represented vehicles making a round-trip (see Figure 3-4 for a graphical depiction of a one-way trip vs. a round trip).

For example, Figure 3-5 display the hourly travel times and tag matches from Friday May 27th through Saturday May 28th, 2011 between Auburn/Bell on I80-E (origin) and Prosser Village on I80-W (destination). The Prosser Village reader is 66 miles east of the Auburn/Bell reader, and is deployed in the freeway median. As indicated in Figure 3-4, the minimum travel times between Auburn/Bell on 80-E and Prosser Village on 80-W ranged between 60-80 minutes, which is reasonable given the 60 mile distance between them. The 75th percentile travel times are higher, likely reflecting the travel times of vehicles detected passing the Prosser Village reader on 80-W as part of a round trip after having first passed both the Auburn/Bell reader, as well as the Prosser Village reader, but not being detected by it, while traveling east. Finally, the median travel times more closely reflect the minimum travel times, indicating that the Prosser Village reader was matching more toll tags traveling past it along 80-E than were being generated based on 80-W round trips as reflected in the 75th percentile travel time. Overall, the research team’s analysis indicated that only the Prosser Village reader was capable of monitoring bi-
directional traffic; Caltrans later indicated that the Prosser Village reader had been deployed with antennae facing in both directions of travel.

**Figure 3-4: Graphical Depiction of a One-way Trip vs. a Round-Trip**

**Figure 3-5: Travel Times Between Origin I80-E at Auburn/Bell and Destination I80-W at Prosser Village**

What percentage of total traffic is being detected?

As mentioned previously, the percentage of the vehicle population sampled by individual readers depends on a number of factors, including the penetration rate of the technology within the vehicular population and the positioning and mounting of the reader.
The Bay Area Toll Authority (BATA) reported in January 2011 that 53% of drivers passing through its toll plazas were equipped with FasTrak tags, with that percentage increasing to 65% during weekday peak periods [1]. Even so, the ETCs in the Tahoe area are more than 100 miles from the nearest toll plaza. Consequently, the percentage of vehicles equipped with FasTrak tags depends, to a great extent, on traffic patterns between the Bay Area and Lake Tahoe.

With respect to the mounting configuration of the readers, previous ETC-based travel time data collection deployments noted that a number of configuration-related factors have the potential to impact the quantity and quality of tag reads [2]. For example, when readers are positioned directly overhead, such as at tolling facilities, they reliably capture data from almost all toll tags. That said, in many real-world traffic monitoring deployments, such as in Lake Tahoe, ETC readers are placed at the side of the road, increasing their distance from vehicles and reducing the efficiency of their tag reads. Such configurations also make it more difficult for readers to capture traffic across all lanes of travel, particularly when there are multiple lanes of traffic.

To calculate the percentage of vehicles sampled at each reader, the research team compared ETC tag reads with the traffic flows measured at nearby loop detectors; the result is referred to as the hit rate. The hit rate at Prosser Village was not analyzed since there were no working loop detectors nearby.

Table 3-1 displays the average daily hit rates, by day of the week, for each of the ETCs along I-80, collected during the week of May 9th to May 15th, 2011. Low hit rates on Sunday and Monday were common to all of the eastbound readers, especially on the eastern end of the monitored corridor. Another trend common across all readers, though especially marked at the Auburn/Bell Road reader, was the spike in the hit rate during the overnight hours (see Figure 3-6). This could be due to the higher percentage of freight traffic during these hours, which may be more likely to be equipped with FasTrak tags.

Table 3-1: Average ETC Hit Rates by Day of Week (7:00 AM-8:00 PM)

<table>
<thead>
<tr>
<th>Reader</th>
<th>Average</th>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday-Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auburn/Bell Road (80-E)</td>
<td>3.4%</td>
<td>2.9%</td>
<td>2.6%</td>
<td>4.0%</td>
<td>4.0%</td>
<td>3.6%</td>
</tr>
<tr>
<td>Rainbow (80-E)</td>
<td>5.4%</td>
<td>2.9%</td>
<td>3.3%</td>
<td>6.9%</td>
<td>7.3%</td>
<td>6.7%</td>
</tr>
<tr>
<td>Rest Area (80-E)</td>
<td>3.2%</td>
<td>1.6%</td>
<td>1.8%</td>
<td>5.0%</td>
<td>4.0%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Donner Lake (80-E)</td>
<td>6.3%</td>
<td>3.6%</td>
<td>3.7%</td>
<td>8.9%</td>
<td>8.5%</td>
<td>6.8%</td>
</tr>
<tr>
<td>Kingvale (80-E)</td>
<td>6.4%</td>
<td>3.9%</td>
<td>4.4%</td>
<td>9.6%</td>
<td>7.2%</td>
<td>7.0%</td>
</tr>
<tr>
<td>Hirschedale (80-W)</td>
<td>4.5%</td>
<td>4.1%</td>
<td>3.7%</td>
<td>5.6%</td>
<td>5.0%</td>
<td>4.1%</td>
</tr>
<tr>
<td>Baxter (80-W)</td>
<td>6.7%</td>
<td>6.2%</td>
<td>6.0%</td>
<td>8.1%</td>
<td>7.1%</td>
<td>5.9%</td>
</tr>
</tbody>
</table>
Comparing average hourly hit rates for all of the readers on I-80-E from Tuesday through Friday makes it clear that some readers are sampling a significantly higher percentage of traffic than others. An examination of photographs of the signs onto which each reader was mounted provides no clear explanation for why the hit rates at some readers are approximately double those at other readers. The hit rate at Auburn/Bell Road may be lower because it is the only location with three lanes of travel (all other readers only monitor two) being monitored by two antennae. The Rest Area reader, though appearing to be optimally positioned above of the roadway, also has a low hit rate. Another possible reason for the low hit rate is that the antennae here are not properly aligned with the two lanes of travel, resulting in a reduced number of toll tag reads.

To gauge the sampling rate in another way, we also looked at the raw number of hourly tag reads reported by each reader, the results of which are displayed in Figure 3-7 for the eastbound direction of travel. Despite its low percentage of tag reads, the reader at Auburn/Bell Rd. still records a large number of reads, simply due to the fact that traffic volumes are higher here than at any other reader. The highest number of reads recorded across readers is on Friday, due to the recreational pattern of weekend trips to Lake Tahoe.
What percentage of toll tags is matched between pairs of readers?

For the purpose of calculating travel time reliability-related metrics it is most important to have the ability to quantify the typical percentages and volumes of toll tags matched between multiple readers, as this directly impacts the quality of aggregated travel times. To quantify typical tag match rates between readers, the team looked at the percentage of vehicles being matched between the furthest upstream readers (Auburn/Bell Road in the eastbound direction and Hirschdale Road in the westbound direction) and all subsequent downstream readers between May 9, 2011 and May 15, 2011 (see Figure 1-2 for the deployment layout).

Figure 3-8 shows the percentage of toll tags detected at the Auburn/Bell reader that are re-identified at each downstream 80-E reader (ordered from left to right by distance from origin). If each reader’s data collection capabilities, and therefore their hit rates, were identical, we would expect to see the percentage of matched tag reads decrease with distance from the origin reader as vehicles detected at Auburn/Bell Road deviated from 80-E. However, this trend does not hold for these readers. Instead, the highest matching percentage (91%) is seen between Auburn/Bell Rd. and the Kingvale reader, which are separated by a distance of 50 miles. At the same time, matches between Auburn/Bell Rd. and the Rainbow and Rest Area readers are much lower, which is consistent with the low hit rates measured at these three stations.

Figure 3-8: Average Percentage of Toll Tags Matched on EB I-80 From Origin Auburn/Bell

Figure 3-9 displays the total number of hourly matches measured on eastbound I-80 between the origin reader at Auburn/Bell Rd. and all downstream readers. At all readers, matches are the highest on Fridays, which is supported by local traffic patterns between the Bay Area and Lake Tahoe. Again, despite being the second furthest reader from the origin, the Kingvale reader often sees the most matches on eastbound I-80. Daytime matches between Auburn/Bell and Kingvale generally exceed 30 per hour (3 per five minutes or eight per fifteen minutes). While this number of samples is likely too low to compute accurate five-minute average travel times, this data has the potential to be used to generate average fifteen-minute or hourly statistics throughout the week.
FINDINGS

There are two primary variables that impact hit rate: (1) the total number of ETC tags in the population of vehicles that pass a reader; and, (2) the number of tags actually read by a specific reader. The product of these factors has a significant influence on the accuracy of travel time data generated between any two ETC readers. With that in mind, this use case has sought to demonstrate how the hit rates and matching percentages generated by the Tahoe region ETC network may be impacted by the configuration of individual ETC readers. Table summarizes the configuration and average data collection results for each of the single-directional ETC readers deployed along I80-E during Friday afternoons/evenings (12:00 – 8:00 PM); considered the peak period for this roadway due to weekend traffic between the Bay Area and Lake Tahoe. All of the readers included in this table were deployed along I-80, are mounted on overhead signs, and are monitor two or three (in one case) lanes of traffic. Despite these readers being deployed under what would appear to be such similar conditions, this table’s content indicates that there are a number of differences in the percentage of the traffic stream sampled at the different locations that are worth noting.

To begin with, hit rates (% of Traffic Sampled) for the Donner Lake and Rainbow readers are more than twice those generated by the Auburn/Bell Road and Rest Area readers. Although the team was not able to investigate the underlying reason for these differences, we believe they are most likely the result of:

- Reader antennae being misaligned at some locations.
- The reader at Auburn/Bell Road attempting to collect data from three lanes of traffic using only two ETC readers.

As seen in the table, differences in hit rate of only 2-3% can make a significant difference in the number of tag reads collected, which is crucial for ensuring that a sufficient number of samples are re-identified downstream to generate accurate travel times and travel time distributions.

As expected, the hit rate for an individual reader has a profound impact on that reader’s ability to re-identify vehicles initially detected at upstream readers. For example,
as shown in Table 3-2, even though the Auburn/Bell Road reader is 45 miles and 24 exits from the downstream reader at Rainbow, the high hit rate at this downstream reader enables it to re-identify 83% of vehicles initially detected at Auburn/Bell Road. Given the number of opportunities to exit the freeway, this likely represents nearly all of the ETC-equipped vehicles that pass between the readers. Conversely, despite the fact that the Rest Area reader is only 8 miles from the Rainbow reader with only one exit ramp in between, it is only able to match 42% of vehicles initially identified at Rainbow. Overall, at least on rural roads experiencing fairly significant through traffic, the readers’ hit rates appear to impact the percentage of matched vehicles to a greater extent than the distance between the readers.

However, even with ideal reader placement and configuration, the primary constraint on the percentage of traffic sampled will always be the penetration rate of toll tags in the population. In rural areas, it is uncommon to have electronic tolling infrastructure, so deploying ETCs in these locations requires that at least some portion of the traffic stream be composed of vehicles equipped with toll tags used in nearby urban areas. The results of this use case show that this penetration rate can vary by time of day and day of week; for example, on I80-E, far fewer Fastrak-equipped vehicles travel the corridor on Sundays and Mondays than during the rest of the week.

Table 3-2: I-80 E ETC Reader Summary, Fridays, 12:00 PM-8:00 PM

<table>
<thead>
<tr>
<th>Reader</th>
<th>Mounting Type</th>
<th>Lanes</th>
<th>Tag Reads</th>
<th>% of Traffic Sampled</th>
<th>Distance to Next Reader (mi.)</th>
<th>Exits Between Readers</th>
<th>% Hits Re-identified Downstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auburn/Bell Rd</td>
<td>EB Roadside VMS</td>
<td>3</td>
<td>648</td>
<td>3.5%</td>
<td>45</td>
<td>24</td>
<td>83%</td>
</tr>
<tr>
<td>Rainbow</td>
<td>EB Roadside VMS</td>
<td>2</td>
<td>789</td>
<td>7.4%</td>
<td>8</td>
<td>1</td>
<td>42%</td>
</tr>
<tr>
<td>Rest Area</td>
<td>EB Roadside Sign</td>
<td>2</td>
<td>380</td>
<td>3.6%</td>
<td>4</td>
<td>1</td>
<td>99%</td>
</tr>
<tr>
<td>Donner</td>
<td>EB Roadside Sign</td>
<td>2</td>
<td>785</td>
<td>7.4%</td>
<td>3</td>
<td>2</td>
<td>96%</td>
</tr>
<tr>
<td>Kingvale</td>
<td>EB Roadside Sign</td>
<td>2</td>
<td>696</td>
<td>6.6%</td>
<td>6</td>
<td>--</td>
<td>61%</td>
</tr>
</tbody>
</table>
B. IMPACT OF BLUETOOTH READER DEPLOYMENT CONFIGURATION ON DATA QUALITY

This use case details the findings of the research team’s investigation into the impact of the configuration of the Lake Tahoe Bluetooth network on the quality of travel time data collected.

INTRODUCTION

The Bluetooth reader network leveraged in this case study was deployed along Interstate 5 (I-5) in Sacramento and US 50 between Placerville and Lake Tahoe. For each BTR, the research team received configuration data in a .CSV file, with fields for the node ID, a textual location, and a latitude/longitude. The research team was also provided with a 2-gigabyte SQL file containing all of the Bluetooth data collected at the readers between December 25, 2010 and April 21, 2011; this use case only utilizes the eight BTRs that provided more than a week’s worth of data. This data was downloaded into PeMS for use in computing travel times between each BTR pair.

METHODOLOGY

In evaluating the impact of the network’s configuration on the characteristics of the reported travel times, the team sought to answer questions of a similar nature to those explored as part of the ETC use case, including:

1. Are the readers where they are reported to be?
2. Which facilities is each reader monitoring?
3. What percentage of total traffic is being detected?
4. What percentage of Bluetooth devices is matched between pairs of readers?

The first question was particularly important for this use case as the BTRs were deployed as part of a test, and not as permanent data collection infrastructure. As a result, each BTR changed locations multiple times over a span of several months. To compute accurate travel times, the team had to ensure that the locations provided in the configuration file matched the data delivered in the SQL file. This was achieved by mapping the latitude and longitude provided by Caltrans to determine whether the data matched the textual locations provided.

Answering the second question is critical to all Bluetooth studies. Class I Bluetooth devices, like the ones used in this case study, have a detection radius of 300’. As a result, the potential exists for the BTRs to monitor bi-directional traffic along a roadway, as well as traffic along parallel facilities, which presents challenges when trying to compute accurate travel times. As such, the research team evaluated the reader locations and data to approximate the lanes of travel they each monitored, whether they were monitoring traffic bi-directionally, whether they might also be detecting vehicles on on-ramps, off-ramps, or frontage roads, and whether the potential existed for them to capture data concerning the movement of other modes of travel, such as from bicyclists.

The third question addresses the “hit rate” occurring at each reader. The team calculated this by comparing hourly BTR reads against hourly volumes collected from nearby loop detectors. In an effort to relate mounting configuration to the percentage of traffic sampled, hit rates were subsequently compared between readers.
The final question relates to the quality of travel times being reported. As the higher the percentage of matches, the more accurate a travel time estimate is likely to be, the research team assessed the percentage of Bluetooth devices matched between all possible combinations of upstream and downstream BTRs. Results from each combination of BTRs were then compared to determine how the percentage of matches is impacted by the hit rates of each individual reader, as well as the distance between readers.

**ANALYSIS**

This subsection documents the process used and analysis conducted by the team to develop answers to the aforementioned four questions.

**Are the readers where they are reported to be?**

Using the information provided by Caltrans, the team verified the location of each BTR according to both its latitude/longitude and textual description. While a number of the readers represented in the configuration file were erroneously located (for example, the latitude/longitude of one placed it in a lake), the eight readers used as part of this use case all appeared to be in roughly the correct location. Photographs of each BTR station used as part of this use case are displayed in Figure 3-10. One BTR, deployed on US-50 at Echo Summit is not visible as a result of being buried in snow. Despite this, the team was able to use the data collected from this station as part of its analysis.

As a final location confirmation step, the team evaluated the minimum travel times computed between each BTR to ensure they were reasonable given the distances involved. All minimum travel times were subsequently determined to be reasonable, and the BTR locations deemed accurate.

**Which facilities is each reader monitoring?**

The next step in understanding the impact of each BTR’s configuration on the nature of the data collected was to determine which readers might be capturing traffic data for multiple directions of travel. The BTRs evaluated as part of this use case were deployed as follows:

- Three of the readers on US-50 were deployed in locations where there is one lane of travel in each direction.
- The reader at US-50 and Placerville monitored two lanes in each direction.
- The reader at US-50 and Meyers was located near an intersection that might result in it picking up MAC addresses from vehicles turning onto US-50 from a cross street.
- The reader at I-5 and Vallejo potentially monitored up to five lanes in each direction.
- The reader at I-5 and Gloria (the only BTR along I-5 located on southbound side of the roadway) had the potential to monitor up to four lanes of travel in each direction.
- The reader at I-5 and Florin was located in the middle of the clover-leaf on-ramp of Florin Road to I-5 North. It was adjacent to four mainline northbound and southbound lanes. Given the reader’s location, it was likely detecting significant numbers of vehicles entering and exiting I-5, both traveling at slower speeds and
being detected earlier (for on-ramp vehicles) or later (for off-ramp) than if they were actually traveling on I-5.

- The reader at I-5 and Pocket was located some distance from the northbound side of the roadway. It had the potential to monitor two on-ramp lanes to I5-N, three mainline lanes in each direction, and a clover-leaf off-ramp from I5-S.

Based on this analysis, the research team concluded that all BTRs were likely monitoring at least bi-directional traffic, a conclusion that was confirmed by the data analysis conducted in support of the following subsection. This effort also provided some insight into how the reader’s locations have the potential to impact the sampling of non-representative trips.

What percentage of total traffic is being detected?

As with the ETCs, the percentage of the traffic stream monitored by a Bluetooth reader depends on the penetration of Bluetooth-enabled devices within the vehicle population. Although it is estimated that 20% of travelers now have Bluetooth devices with them in their vehicles, at least a quarter of them do not have the device set to discoverable mode.

The detection rate also depends on the reader’s configuration. Class I Bluetooth readers have a 300’ detection radius. Based on this, a single BTR could easily monitor all lanes of a freeway that has four lanes of traffic in each direction of travel and is barrier-separated. That said, it might also collect a number of undesired samples, such as Bluetooth devices on parallel facilities or within office buildings. All readers used in this case study had approximately the same average signal strength, so this variable was not a factor.

To calculate the percentage of vehicles sampled at each BTR, the research team compared Bluetooth mobile device reads with the traffic flows measured at nearby loop detectors; the result is referred to as the hit rate. Hit rates were computed for the four readers on I-5 (there were no working loop detectors near the US-50 readers). Because all readers were presumed to monitor both directions of travel along I-5, and as it is impossible to assign a direction of travel to unmatched Bluetooth reads, hit rates were calculated by comparing hourly detections at each reader with hourly volumes summed up from nearby northbound and southbound loop detectors over a week-long period (Monday, February 28th to Sunday, March 6th, 2011). In addition, because the Florin and Pocket readers clearly detected traffic on roadway on-ramps, hit rates at these readers were computed by comparing the hourly reader detections with the hourly volumes summed up from the mainline and on-ramp loop detectors (so as not to upwardly bias the hit rates at these readers).

Bluetooth hit rates were first evaluated to determine if they exhibited any time of day or day of week patterns. As with the ETC readers, hit rates were lowest during the early morning hours. There were no other discernable patterns. Figure 3-11 compares the hourly hit rates measured over three days (Tuesday-Thursday) across the four readers. The hit rates at all readers generally ranged between 6% and 10%. Hit rates were usually highest at the Gloria reader, which was directly adjacent to the southbound lanes; hit rates between 8% and 10%. The reader at Pocket typically had the lowest hit rates, between 6% and 8%, possibly due to its setback from the roadway.
US50/Meyers: off of US50-E

US50/Echo Summit: not visible in photograph

US50/Twin Bridges: off of US50-E

US50/Placerville: off of US50-E

I5/Vallejo: off of I5-N
Figure 3-10: Bluetooth BTR cabinet locations

Figure 3-11: Hourly Hit Rates, I-5 Bluetooth Readers
The data displayed in Figure 3-12 presents the raw number of hourly MAC address reads at each reader listed. The reads shown in this plot are based on the number of MAC address reads remaining following filtering to remove duplicate IDs at the same reader during the same hour. While the reader at I-5 and Vallejo does not have the highest hit rate, it generally records the largest number of MAC addresses per hour, reaching nearly 1,000 reads per hour during the weekday PM peak. In contrast, the reader at I-5 and Pocket has both the lowest hit rate and the lowest number of reads, with between 500 and 600 MAC address reads per hour during the peak hours and only 300 to 400 per hour during the midday.

While the hit rate could not be computed for the readers on US-50, the research team did evaluate the raw number of hourly MAC address reads at each reader on this road (see Figure 3-13). The pattern of reads on US-50 differs from that on I-5, which follows the more typical AM/PM peak commute pattern. On US-50, each reader detects the most reads on Fridays, Saturdays, and Sundays, due to recreational traffic patterns near Lake Tahoe. At the Meyers, Echo Summit, and Twin Bridges readers, which are all relatively closely spaced (within 12 miles of one another) near South Lake Tahoe, the number of hourly reads are fairly similar, and are quite low (30-50 per hour, or 2 to 4 per five minutes) from Monday through Thursday. The number of reads at the Placerville reader, which is closer to Sacramento, are higher, especially during the work week, when this location has higher traffic volumes.
What percentage of Bluetooth devices is matched between pairs of readers?

For the purpose of supporting the calculation of travel time reliability-related metrics it is most important to have the ability to quantify the typical percentages and volumes of Bluetooth devices matched between multiple readers, as this directly impacts the quality of aggregated travel times. The first step in performing this analysis was to evaluate the percentage of each reader’s Bluetooth MAC address reads re-identified at downstream readers. Results for the readers along I-5 are shown in Figure 3-14 and for the US-50 readers in Figure 3-15.

On I-5, the Vallejo (northern-most) and Pocket (southern-most) readers only have downstream readers in one direction of travel (see Figure 1-2 for the deployment layout). Re-identification of devices between these readers occurred as follows:

- For the Vallejo reader, approximately 42% of its MAC address reads were re-identified at the Gloria reader located about 4 miles to the south.
- For the Gloria reader, about 48% of its reads were re-identified in the northbound direction at Vallejo, while 50% of its reads were re-identified in the south-bound direction at Florin; 2% were not re-identified at all.
- For the Florin reader, 53% of reads were re-identified in the northbound direction at Gloria, while 39% were re-identified in the southbound direction at Pocket; 8% were not re-identified in either direction.
- For the Pocket reader, 48% of reads were re-identified in the northbound direction at Florin.

Overall, the rate of matching between readers was very high, with the vast majority of Bluetooth devices matched at another sensor for use in generating travel times.

Re-identified rates were also high between the readers along US-50, particularly the three deployments closest to Lake Tahoe. Re-identification of devices between these readers occurred as follows:
For the Meyers (eastern-most) reader, for which there is no downstream reader in the eastbound direction, 50% of reads were re-identified at the Echo Summit reader, four miles to the west.

For the Echo Summit reader, 57% of reads were re-identified at Meyers and 43% were re-identified at Twin Bridges, 8 miles to the west. Virtually none of the reads captured at Echo Summit went unmatched, likely due to its location at a point on the roadway that has no parallel facilities, and fact that there were few possible exits between Echo Summit and Meyers or Echo Summit and Twin Bridges.

For the Twin Bridges reader, 40% of reads were re-identified to the east at Echo Summit and 47% were re-identified at Placerville, 39 miles to the west. 13% of reads captured at Twin Bridges are not re-identified.

For the Placerville (western-most) reader, 22% of reads were re-identified downstream at Twin Bridges.

Based on these high re-identification rates, the team concluded that the readers on US-50 were capable of detecting and re-identifying a very high proportion of the Bluetooth devices that pass through their detection zones, likely due to the narrow roadway width at these locations and the limited options available to exit the roadway.

![Figure 3-14: MAC address matching rates, I-5 readers](image-url)
The next technique for evaluating Bluetooth device reidentification between readers was to examine the raw number of matches between readers in order to assess whether the match volumes were sufficient to yield accurate average travel times. This was carried out by selecting an origin reader and computing the hourly matches to a series of destination readers. Figure 3-16 displays the results of this analysis in the southbound direction along I-5 for the Gloria, Florin, and Pocket readers from the origin reader at Vallejo. Highlights of this analysis included:

- As the team expected, the greatest number of matches occurred with the closest downstream reader, Gloria, and the fewest matches with the reader furthest away, Pocket. These matches differed by about 100 during each PM peak hour, representing a difference of 25%.
- Matches between Vallejo and all downstream readers averaged about 16 per five-minutes during daytime hours, likely sufficient for obtaining five-minute travel times.
- The number of matches peaked during the PM period, at around 350 to 500 per hour, when travelers were departing Sacramento for its southern suburbs.
- Volumes were lower on weekends, but still sufficient to support average travel time computations at a fine granularity.

Figure 3-17 displays results of this analysis for hourly northbound matches at the Florin, Gloria, and Vallejo readers from the origin reader at Pocket. Highlights of this analysis included:

- As Pocket had the lowest hit rate of the readers on I-5, a smaller percentage of vehicles were available for re-identification when using this reader as an origin.
- The numbers of matches at each of the three destination readers were very similar, and generally differed by less than 25 per hour, representing a difference of about 10%.
- The number of matches peaked during the AM period, at around 350 to 400 per hour, when the majority of traffic was commuting north to Sacramento.
- As in the southbound direction, matches were lower on Saturdays and Sundays, but still likely sufficient to calculate fine-grained average travel times.
Results for the number of hourly matches between the Meyers reader (eastern-most) and downstream readers on US-50 are displayed in Figure 3-18. Highlights include:

- Although the number of matches decreased with distance from the origin reader, the number of matches was similar between the three destinations due to the fact that a significant amount of traffic on US-50 travels its entire length from Lake Tahoe to Sacramento.
- The number of matches was much lower than along I-5, likely due to its rural characteristics and lower traffic volumes.
- The number of matches was highest on Saturdays and Sundays, due to recreational traffic, peaking on Sunday afternoons, when travelers are returning from Lake Tahoe to Sacramento and the Bay Area.
- During the peak hours on Sunday, there are 100 to 140 hourly matches (8 to 12 per 5 minutes or 25 to 35 per 15 minutes) between the Meyers reader and the Placerville reader, 50 miles away, likely sufficient to calculate 15-minute travel times, and possibly 5-minute travel times, for this facility’s peak hour.
- During the rest of the week, the number of hourly matches ranged from around 20 to 50 during the daylight hours (2 to 4 per five minutes or 5 to 12 per fifteen
minutes). This number of matches is not likely sufficient to compute average travel times every five-minutes, though it might be used to compute fifteen-minute or hourly average travel times.

The number of hourly matches for traffic between the origin reader at Placerville (western-most) and the destination readers at Twin, Echo, and Meyers, is shown in Figure 3-19.

- Matched peaked on Fridays and Saturdays as vehicles traveled from the Bay Area and Sacramento to Lake Tahoe.
- During the peak hours on Friday and Saturday afternoons, matches between the Placerville reader and the Meyers reader, near South Lake Tahoe, were around 100 per hour (8 per five minutes or 25 per 15 minutes), likely enough to compute average travel times at a five-minute or a 15-minute granularity. During the rest of the week, however, there are only about 20 matches per hour.

![Figure 3-18: MAC address matches, US50-W from Meyers reader](image)

![Figure 3-19: MAC address matches, US50-E from Placerville reader](image)

As a number of travelers make trips between Sacramento and Lake Tahoe, there is potentially value in knowing the travel times between the two. For this reason, the team also examined the number of hourly matches between the readers along I-5 and the
readers on US-50. Figure 3-18 shows the number of matches between the reader at Meyers (closest to South Lake Tahoe) and other readers along I-5. As these readers are on different freeways and are about 100 miles apart, the key question is whether there are sufficient matches to compute travel times at any level of granularity. Figure 3-19 displays the results of this analysis, representing trips along US-50W, exiting onto I-5S.

- The peak number of matches occurred on Sunday afternoon, when some hours had up to 16 - 18 matches. However, even during this peak, there were some hours when the number of matches dipped to only 5 per hour. Consequently, it does not appear possible to consistently calculate travel times at a fine granularity even on, Sunday afternoons. However, there are sufficient matches to compute hourly travel times, which could provide a reasonable indication of travel time reliability for who want to make this return trip from Lake Tahoe.
- During the rest of the week, there were insufficient matches to compute accurate average travel times by time of day and day of week, though travel times could be collected over a period of many weeks to compute average travel times and travel time variability.

Figure 3-20 displays the hourly matches between the Pocket reader (southern-most) on I-5 and each of the readers along US-50. These matches most likely represent vehicles traveling north on I-5 towards Sacramento, and then exiting onto US-50 E in the direction of Lake Tahoe. The number of matches peaked at between six to 12 per hour during Friday afternoon, and was also higher on Saturday morning, at around 8 per hour. Matches during the rest of the week were lower, but could potentially be studied over time to better understand the variability of travel times by time period.

![Figure 3-20: MAC address matches, I5-S from Meyers on US-50](image-url)
FINDINGS

This use case provided the opportunity to assess the performance of a Bluetooth-based travel time monitoring system deployed in an urban environment with that deployed in a rural environment, while simultaneously demonstrating how sensor configuration impacts both the amount and quality of data collected.

One overarching finding of this use case is that the potential exists to use Bluetooth readers to generate travel times over long distances between urban and rural settings based on travel along adjoining roadways; though this is heavily dependent on the presence of the right set of conditions. For example, as indicated in Table 3-3, during an average Friday afternoon/evening, 132 vehicles detected at the Vallejo reader on I-5 (2% of the Bluetooth reads at this location) are later re-identified at the Placerville reader on US 50, more than 46 miles away. For this origin-destination pair, this degree of mobile device re-identification is sustained only on Fridays and Saturdays; on other days of the week, far fewer matches are registered.

Within urban environments, Bluetooth readers placed along the same freeway have the capacity to produce sufficient numbers of matches to continuously compute fine-grained five-minute travel times. In contrast, due to lower overall traffic volumes in rural areas, fewer travel time matches are generated and this capacity is therefore reduced. Even so, at least in the area around Lake Tahoe, sufficient matches were generated to compute fine-grained travel times during peak days.

The research team’s results also indicate that a single Bluetooth reader can typically be used to monitor bi-directional traffic. Although the number of lanes at each reader used as part of this case study ranged from two to ten, data indicate that each reader was able to capture traffic in most, if not all, of the lanes at its location.

Finally, this use case enabled the research team to compare hit rates and matching percentages for readers located in both urban and rural environments. In this study, as is typical for urban versus rural settings, the biggest differences between the readers deployed on I-5 and US 50 included:
- The number of lanes at each reader;
- The distance between readers; and,
- The traffic volumes at each reader.

The I-5 readers were all configured to monitor traffic across six or more lanes of bi-directional traffic. Although three of the four I-5 readers were placed adjacent to the northbound lanes, the content of Tables 3-3 and 3-4, demonstrates that they generate significant hit rates in both directions of travel; a similar situation exists for the Gloria reader deployed adjacent to the southbound lanes. This demonstrates that Bluetooth readers have the potential to monitor wide bi-directional freeway segments.

The content of these tables also indicates that directional traffic patterns have a significant degree of influence on Bluetooth device matching patterns. For example, on I-5, where northbound and southbound traffic volumes are comparable throughout the day, none of the readers re-identify more than 50% of the hits from upstream readers. In contrast, 68% of the hits from the Twin Bridges reader on rural US 50 are re-identified at the Placerville reader (39 miles away); see Table 3-4. These higher rural matching percentages, despite longer distances between readers, are in part due to US 50 exhibiting much stronger directional trends (e.g., eastbound US 50 carrying the majority of traffic on Friday afternoons/evenings). Despite this, volumes of Bluetooth reads along I-5 are several times greater than those along US 50, facilitating the calculation of more granular travel time reliability metrics.

Finally, each of the eight Bluetooth readers from which data was collected as part of this use case were mounted on roadside controller cabinets, and used directional antennae to focus signal strength toward the roadway. The fact that each of the readers had high hit rates and produced significant matching percentages with downstream readers demonstrates that this is an effective configuration for capturing multi-lane, bi-directional traffic. However, the team also found that the readers are most effectively used when deployed in locations where they only monitor traffic in the mainline lanes. This is particularly a problem with readers placed adjacent to on- or off-ramps, such as the readers on I-5 at Florin and Pocket, as the travel time re-identification between the vehicle’s timestamp at the on-ramp and its timestamp at the next downstream reader will be higher than the true travel time on the freeway; this is especially true if the ramp is congested or has ramp metering. For agencies using Bluetooth networks already in the field, it is important to determine which readers may be monitoring ramp traffic so that these travel time biases can be understood and mitigated.
Table 3-3: BTR Reader Summary, I-5N to US 50E, Friday 12:00 PM-9:00 PM

<table>
<thead>
<tr>
<th>Reader</th>
<th>Mounting Type</th>
<th>Lanes</th>
<th>MAC ID Reads</th>
<th>% of Traffic Sampled</th>
<th>Distance to Next Reader (mi.)</th>
<th>Exits Between Readers</th>
<th>% Hits Re-identified Downstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pocket (I-5)</td>
<td>NB roadside controller cabinet</td>
<td>3</td>
<td>4208</td>
<td>7.2%</td>
<td>0.9</td>
<td>1</td>
<td>43%</td>
</tr>
<tr>
<td>Florin (I-5)</td>
<td>NB roadside controller cabinet</td>
<td>4</td>
<td>5402</td>
<td>8.1%</td>
<td>1.1</td>
<td>0</td>
<td>47%</td>
</tr>
<tr>
<td>Gloria (I-5)</td>
<td>SB roadside controller cabinet</td>
<td>4</td>
<td>5843</td>
<td>9.6%</td>
<td>4</td>
<td>2</td>
<td>45%</td>
</tr>
<tr>
<td>Vallejo (I-5)</td>
<td>NB roadside controller cabinet</td>
<td>5</td>
<td>6642</td>
<td>7.7%</td>
<td>46</td>
<td>27</td>
<td>2%*</td>
</tr>
<tr>
<td>Placerville (US 50)</td>
<td>EB roadside controller cabinet</td>
<td>1</td>
<td>1676</td>
<td>Not available</td>
<td>39</td>
<td>7</td>
<td>34%</td>
</tr>
<tr>
<td>Twin Bridges (US 50)</td>
<td>EB roadside controller cabinet</td>
<td>1</td>
<td>882</td>
<td>Not available</td>
<td>8</td>
<td>3</td>
<td>55%</td>
</tr>
<tr>
<td>Echo Summit (US 50)</td>
<td>WB roadside controller cabinet</td>
<td>1</td>
<td>771</td>
<td>Not available</td>
<td>4</td>
<td>2</td>
<td>74%</td>
</tr>
<tr>
<td>Meyers (US 50)</td>
<td>EB roadside controller cabinet</td>
<td>1</td>
<td>1059</td>
<td>Not available</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Table 3-4: BTR Reader Summary, US 50W to I-5S, Sunday 12:00 PM-9:00 PM

<table>
<thead>
<tr>
<th>Reader</th>
<th>Mounting Type</th>
<th>Lanes</th>
<th>MAC ID Reads</th>
<th>% of Traffic Sampled</th>
<th>Distance to Next Reader (mi.)</th>
<th>Exits Between Readers</th>
<th>% Hits Re-identified Downstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meyers (US 50)</td>
<td>EB roadside controller cabinet</td>
<td>1</td>
<td>936</td>
<td>Not available</td>
<td>4</td>
<td>2</td>
<td>52%</td>
</tr>
<tr>
<td>Echo Summit (US 50)</td>
<td>WB roadside controller cabinet</td>
<td>1</td>
<td>771</td>
<td>Not available</td>
<td>8</td>
<td>3</td>
<td>66%</td>
</tr>
<tr>
<td>Twin Bridges (US 50)</td>
<td>EB roadside controller cabinet</td>
<td>1</td>
<td>968</td>
<td>Not available</td>
<td>39</td>
<td>7</td>
<td>68%</td>
</tr>
<tr>
<td>Placerville (US 50)</td>
<td>EB roadside controller cabinet</td>
<td>1</td>
<td>1495</td>
<td>Not available</td>
<td>46</td>
<td>27</td>
<td>6%*</td>
</tr>
<tr>
<td>Vallejo (I-5)</td>
<td>NB roadside controller cabinet</td>
<td>5</td>
<td>4940</td>
<td>8.1%</td>
<td>4</td>
<td>2</td>
<td>45%</td>
</tr>
<tr>
<td>Gloria (I-5)</td>
<td>SB roadside controller cabinet</td>
<td>4</td>
<td>4352</td>
<td>8.7%</td>
<td>1</td>
<td>1</td>
<td>48%</td>
</tr>
<tr>
<td>Florin (I-5)</td>
<td>NB roadside controller cabinet</td>
<td>4</td>
<td>4003</td>
<td>8.5%</td>
<td>1</td>
<td>1</td>
<td>42%</td>
</tr>
<tr>
<td>Pocket (I-5)</td>
<td>NB roadside controller cabinet</td>
<td>3</td>
<td>3208</td>
<td>7.5%</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

* Readers should note that the low re-identification rates for Vallejo on Table 3-3 and Placerville on Table 3-4 are primarily the result of the next downstream reader for each being 46 miles away and on an adjoining roadway; I-5 to US 50 and US 50 to I-5.
C. USING BLUETOOTH AND ELECTRONIC TOLL COLLECTION DATA TO ANALYZE TRAVEL TIME RELIABILITY IN A RURAL SETTING.

This use case aims to quantify the impact of adverse weather and demand–related conditions on travel time reliability using data derived from Bluetooth and electronic toll collection-based systems deployed in rural areas.

INTRODUCTION

At present, loop detectors provide the majority of transportation data used for highway analysis. These detectors must be embedded in the roadway and require regular quality checking and often-costly maintenance. Bluetooth and electronic toll collection-based systems, on the other hand, can be mounted onto existing infrastructure either overhanging or adjacent to the roadway, thereby reducing the costs of deployment, reconfiguration, repair, and replacement. These systems work by scanning compatible devices deployed inside passing vehicles for unique identification information (i.e., Media Access Control (MAC) ids for Bluetooth devices and tag id numbers for toll transponders). If multiple readers detect identification information for a uniquely identifiable device, a record of that vehicle’s travel time can be constructed. Because these devices do not need to be permanently fixed on the roadway, they offer a more flexible and often more cost effective method for detection, especially in rural locations.

To examine travel time reliability within the context of this use case, methods were developed to generate probability density functions (PDFs) from large quantities of travel time data representing different operating conditions. To facilitate this analysis, travel time and flow data from ETC readers deployed on I-80W and Bluetooth readers deployed on I-50E and I-50W were obtained from PeMS and compared with weather data from local surface observation stations. PDFs were subsequently constructed to reflect reliability conditions along these routes during adverse weather conditions, as well as according to time-of-day and day-of-week. Practical data quality issues specific to Bluetooth and ETC data were also explored.

This use case has value to a broad range of user groups. Transportation agencies with data collection needs in rural areas will benefit from seeing a travel time reliability analysis of real world data obtained from Bluetooth and ETC devices. This type of data is still fairly uncommon in practice and this use case should help to demystify it, demonstrating how such data sets compare to more commonly available types of traffic data. Operators and analysts will benefit from a discussion of the quality and typical characteristics of this type of data. Transportation agencies with specific data needs and cost constraints seeking a flexible sensor deployment may find Bluetooth or ETC-based systems more attractive based on the results of this analysis.

This use case also has value to operators who are interested in the effects of varying weather conditions and weekend travel on travel time reliability within a rural setting. Understanding the historical effects of different weather and demand conditions on the performance of a given roadway enables operators to respond to similar conditions as they occur, for example, by posting the expected range of travel times on dynamic message signs located at key decision points.
USE CASE ANALYSIS SITES

Two sites were used in the validation of this use case to compare similar phenomena in different locations, as well as to highlight the different types of data available in this region (see Figure 3-22). Site 1 is a 45.2-mile stretch of primarily 4-lane divided highway along I-80W with an estimated free flow travel time of 46 minutes. It begins east of the Truckee-Tahoe Airport weather station and ends just after I-80 exits the western border of the Tahoe National Forest. This roadway is instrumented with ETC readers mounted on sign structures overhanging the roadway.

Site 2 is a shorter 10.8-mile stretch of 2-lane highway along US 50 with an estimated free flow travel time of 14 minutes. This site was examined in both the Eastbound and Westbound directions of travel. It approaches the South Lake Tahoe airport on its eastern end and terminates in the West just outside of Twin Bridges. Site 2 is instrumented with Bluetooth readers deployed along the side of the roadway.

Table 3-5: Site Characteristics

<table>
<thead>
<tr>
<th>Highway</th>
<th>Distance</th>
<th>Estimated Travel Time</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1 I-80 W</td>
<td>45.2 miles</td>
<td>46 minutes</td>
<td>4 lane, divided</td>
</tr>
<tr>
<td>Site 2 US 50 E &amp; US 50 W</td>
<td>10.8 miles</td>
<td>14 minutes</td>
<td>2 lane</td>
</tr>
</tbody>
</table>

These two sites were selected due to their strong weekend traffic patterns, as well as their proximity to local weather observation stations. They were made as short as possible (within the constraints of the detection infrastructure) in order to enable the research team to closely tie its analysis to the data generated by the weather stations, thereby maximizing the relevance of the weather data. Both Sites 1 and 2 are rural and receive relatively little commute or intercity traffic during the week. However, Lake Tahoe is a popular weekend and holiday destination for residents of the Bay Area, which is just a 3.5-hour drive away. I-80 and US 50 are both popular routes to take to get to Lake Tahoe from the Bay Area, and they are known for their heavy weekend traffic as large numbers of people enter and leave the area at nearly the same time.
Figure 3-22: Use Case Site Map

Figure 3-23: Example of Site 1, I-80
ANALYSIS METHODOLOGY

The routes included as part of this use case were analyzed to determine the effects of weather and weekend travel conditions on travel time reliability. To do this, travel time PDFs that isolate certain operating scenarios (e.g., snow on a weekday) were constructed. Time-of-day, day-of-week, and weather conditions based PDFs were generated for Site 1, and time-of-day and day-of-week conditions based PDFs were generated for Site 2.

To begin the analysis, travel time statistics for 5-minute windows at both sites were obtained from PeMS. For Site 1, where weather conditions were considered, travel time data was matched with weather data from the nearby AWOS-III surface observation station. Each 5-minute time interval was marked with its corresponding weather event if any (rain, snow, fog, or thunderstorm), visibility distance (0 to 10 miles), and precipitation (in inches). For site 2, where only weekend travel effects were considered, intervals were grouped into three categories. Travel times were labeled as belonging to a weekday (Monday through Thursday), a Friday, a Saturday, a Sunday, or a holiday.

With the travel time data collected and labeled, an effort was made to determine which data points, if any, should be thrown out. As was discussed previously and will be explored further in the Data section of this use case, travel time data obtained from Bluetooth and ETC readers can, depending on a number of variables, contain artificially long travel vehicle times.

The travel time data, labeled with weather condition and day-of-week, was used to construct PDFs of travel times under varying operating conditions. The effects of weather and weekend travel can be seen in the differences in travel time variability as indicated in the PDFs reflecting different conditions. Finally, aggregate travel time reliability statistics such as the 95th percentile travel time were computed for different conditions.

Data Collected

To complete this use case, Bluetooth and ETC data was retrieved from PeMS for the two sites described above. ETC data was obtained at Site 1 and Bluetooth data at Site
To benefit from the availability of this rich data set, all available data was used in both cases. This was particularly desirable, as the available data does not span seasonal changes. The data obtained from PeMS included:

- Minimum travel time,
- Average travel time,
- Maximum travel time,
- 25th, 50th, and 75th percentile travel times, and
- Flow (number of vehicles observed during the window).

Each of these metrics was collected for a series of consecutive 5-minute windows. It should be noted that not all 5-minute windows during the periods of observation contained usable data, so some gaps exist in the data do exist.

Table 3-6: Dataset Descriptions by Site

<table>
<thead>
<tr>
<th>Roadway</th>
<th>Data Type</th>
<th>Date Range</th>
<th>Data Completeness</th>
<th>Quantity of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-80 West</td>
<td>ETC</td>
<td>4/25/2011 to 6/29/2011</td>
<td>59.5%</td>
<td>11,071 points</td>
</tr>
<tr>
<td>US 50 West</td>
<td>Bluetooth</td>
<td>1/28/2011 to 4/21/2011</td>
<td>35.9%</td>
<td>8,576 points</td>
</tr>
<tr>
<td>US 50 East</td>
<td>Bluetooth</td>
<td>1/28/2011 to 4/21/2011</td>
<td>38.9%</td>
<td>9,376 points</td>
</tr>
</tbody>
</table>

To examine the effect of weather on travel times across Site 1, weather data was obtained from the nearby AWOS-III surface observation station located at the Truckee-Tahoe Airport. This data was available in windows ranging between 5 and 20 minutes, fine grained enough to match well with the 5-minute travel times. Here, the research team focused on optional event tags (fog, rain, snow, or thunderstorm), visibility (0 to 10 miles), and precipitation (in inches).

After the weather and travel time data sets were obtained, the travel time data was next quality checked to ensure no erroneous data points were included. As mentioned previously, Bluetooth and ETC-based data collection systems are susceptible to data errors due to the way they measure travel times. These detectors work by recording the MAC address or toll tag id of vehicles that pass them on the roadway, along with a timestamp. This identification data is matched between detectors such that a vehicle passing multiple BTRs produces a travel time for that link. However, if that vehicle stops somewhere along the roadway after passing the first BTR before it continues on to the second, an artificially large travel time will be seen. Similarly, if a vehicle visits the first BTR, then travels to an adjacent, but unmonitored roadway prior to returning to the monitored roadway and passing the second BTR, the travel time for that trip will be artificially large. Additionally, vehicles traveling past the same BTR more than once in different directions can also cause data errors when readers are capable of measuring multiple directions of travel simultaneously.

To prepare the raw data for analysis, these inaccurate travel times should typically be removed individually. The data set for this use case was composed of aggregate statistics that had already been computed based on all available travel time data; including data that is potentially inaccurate. To prevent the analysis conducted as part of
this use case from being skewed by those values, the research team used the median travel time for each 5-minute interval. In this case, working with the median as opposed to the mean has a significant effect on the analysis, reducing the appearance of implausible extreme values. This works well for periods of time with significant traffic flow because unreasonably long travel times are muted as the sample size increases. However, the problem remains when the sample size is small, as a time interval containing a single extreme value will still result in an unreasonable median. As can be seen in Figure 3-25, below, representing conditions for Site 1, virtually all "long" travel times in the data occur during low volume time periods. It should be noted that the flows shown in Figure 3-25 are not sustained, but rather 5-minute aggregates).

However, this does not mean that poorly represented time intervals should be discarded. While it is true that median travel times from sample intervals with larger numbers of vehicles should better conform to the expected value, and medians of smaller samples are more likely to contain outliers, this phenomenon is also representative of the fundamental behavior of traffic: both high (uncongested) and low (congested) speeds are seen at low flows. Thus, points from sparsely populated time intervals should not necessarily be discarded on those grounds alone (as long as the points can be assumed to be valid). Plotting the data for Site 1 from Figure 3-25 another way yields an empirical fundamental diagram for speed and flow (see Figure 3-26). The expected triangle shape can be seen with congested conditions represented by the points sloping down and to the left from the peak flow (seen around 60 mph) and uncongested conditions represented by the points sloping down and to the right from the peak flow (with speeds above 60 mph). When viewed like this, all points appear to be valid as they are behaving according to basic traffic flow theory. Because longer travel times can be reasonably expected for time periods with few observations (i.e., during congested flow), it is determined that for the purposes of this use case no data points will be excluded.
Travel Time Analysis: Site 1

Site 1, which lies on I-80W and begins just North of Lake Tahoe, is known to receive heavy traffic from vehicles returning to the Bay Area from weekend trips on Sunday evenings. As such, the breakdown of travel times by day-of-week from April 25 to June 29, 2011 shown Figure 3-27 indicates that the Sunday 95th percentile travel time exceeds that of a normal weekday by ~34%. This difference indicates increased travel time unreliability on Sundays, whereas the rest of the week appears fairly consistent. Since Sundays exhibit a significantly different pattern of traffic, they were considered separately as part of the research team’s weather analysis.

Having assessed travel time reliability trends over the entire week, we next examined travel times within individual days to determine if it was necessary to handle AM and PM peak conditions separately during the analysis. To facilitate this, the distribution of travel times for each 5-minute interval over the course of a full day was plotted (see Figure 3-27: 95th Percentile and Median Travel Times for Site 1).
This figure demonstrates that no significant time-of-day trends exist on this route. If the typical day had shown some periodicity, it would have been necessary to examine weather effects during peak and off-peak hours separately. However, as travel times appear to be consistently between 30 and 45 minutes throughout the day the research team was able to conduct its weather-related travel time reliability analysis without accounting for differences between daily peak conditions.

Figure 3-28 also indicates the presence of significantly longer travel times (hovering near the top of the chart). As these travel times do not appear to follow any time-of-day trends, it was surmised that they might be the result of adverse weather conditions. The team explored this idea further by generating PDFs of travel times collected during varying weather conditions. These PDFs were built by placing each 5-minute median travel time into a bin, each of which is 5 minutes wide. To define discrete weather conditions, we adopted five (5) labeled event categories (baseline, snow, rain, fog, and thunderstorm). We then broke the quantitative measures precipitation and visibility down into categories. For precipitation, we created “no precipitation” and “some precipitation” cases, and for visibility, we defined “low visibility”, “medium visibility,” and “high visibility” cases which corresponded to 0-3, 3-7, and 7-10 miles of visibility, respectively. The event conditions were all mutually exclusive, as were the visibility and precipitation categories. Note that the “baseline” event condition does not necessarily mean that driving conditions were ideal, but that no event was associated with that time (there may have been precipitation or low visibility). The resulting weather PDFs can be seen in Figure 3-29 and their effects on travel time are summarized in Figure 3-30. Note that the scale of the vertical axis in Figure 3-29 is not consistent across each of the graphs. This is due to the variable quantities of data available for each condition.
Figure 3-29: Site 1 Travel Time PDFs During Various Weather Conditions
It is clear from Figure 3-30 that snow, low to moderate visibility, and precipitation have a measurable effect on travel time reliability. The 95th percentile travel times during those weather conditions are significantly higher than their median travel times, indicating that the distribution of travel times is skewed toward the high end.

Figure 3-30: Site 1 Summary of Weather Effects

Another way to explore this data is to assess which conditions were present during the longest travel times occurring on this route. The results of this analysis are presented in Table 3-7. This perspective complements that of the PDFs displayed above by revealing that adverse weather events are present during many more long travel times than short travel times. In fact, the research team’s analysis indicated that if a travel time exceeded the 95th percentile for this route, there was nearly a 50% chance that it was snowing, despite snow accounting for only 5% of all trips.

Table 3-7: Weather Conditions Active During Long Travel Times

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Active</th>
<th>Active when travel time exceeded 85th percentile</th>
<th>Active when travel time exceeded 95th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Precipitation</td>
<td>90.3%</td>
<td>84.3%</td>
<td>76.6%</td>
</tr>
<tr>
<td>Precipitation</td>
<td>9.7%</td>
<td>15.7%</td>
<td>23.4%</td>
</tr>
<tr>
<td>Baseline</td>
<td>90.7%</td>
<td>74.9%</td>
<td>54.7%</td>
</tr>
<tr>
<td>Snow Event</td>
<td>5.8%</td>
<td>23.0%</td>
<td>45.3%</td>
</tr>
<tr>
<td>Rain Event</td>
<td>1.6%</td>
<td>1.6%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Fog Event</td>
<td>1.7%</td>
<td>0.5%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Thunderstorm Event</td>
<td>0.3%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
Travel Time Analysis: Site 2

Site 2 was similar to Site 1 in that it is subject to periodic spikes in demand due to weekend travel. However, whereas Site 1 is a 4-lane divided highway, Site 2 is a 2-lane highway (with only intermittent passing opportunities) and thus not as well equipped to handle the additional demand. Site 2 was equipped with Bluetooth detectors that were used to construct travel times in a similar manner to the ETC readers used for Site 1. The goal of the Site 2 travel time analysis was to determine the effects of the weekend travel on this site.

We began by examining a typical day on US 50 to check for the presence of AM or PM peak conditions, which would have to be controlled for as part of day-of-week analysis. Similarly to Site 1, 5-minute median travel times were obtained from PeMS. The time-of-day average of these median travel times is presented in Figure 3-31, which does not appear to show any true peak conditions. While the maximum daily travel time appears to occur at around 5:00 AM, this does not appear to be a true AM peak, likely being attributable to artificially high travel times occurring during low volume periods as discussed in the Data Collection section.

![Figure 3-31: I-50W Average Travel Time by Time-of-Day](image)

This assessment is supported by the average daily flow data displayed in Figure 3-32. Due to the absence of daily peak conditions at this site, the research team decided to consider each day as a whole. If strong peak conditions had been observed, it would have been necessary to develop travel time distributions for peak and off-peak conditions separately.
If this site were indeed subject to heavy weekend demand, it would be expected that travel times would be less reliable during the weekend. In order to explore whether the data supported this, the team first plotted the average vehicle flow over the course of the week for each direction of traffic for this site. It can be seen in Figures 3-33 and 3-34 that weekend demand dominates the traffic profile for this section of roadway. As a result, we would expect travel time unreliability to follow a similar pattern.

Figure 3-32: I-50W Average Flow by Time-of-Day

Figure 3-33: Weekly Flow on US 50E
Figure 3-34: Weekly Flow on US 50W

To visualize travel time unreliability for this site, the research team constructed a travel time density plot representing a full week for US 50 East (see Figure 3-35). This figure is a collection of PDFs for each 5-minute period over the course of the entire week. Since this figure represents travel times in the Eastbound direction, we would expect more unreliability on Friday and Saturday as weekend travelers are making their way to Lake Tahoe from the Bay Area. The PDF appears to confirm this, as it can be seen at a glance that Friday is the day with the most severe unreliability, with Sunday through Thursday exhibiting much more consistent travel times in comparison.
If this variation in travel time reliability over the course of the week is the result of weekend travel patterns and not adverse weather or some other factor, we would expect to see a complementary trend on US 50 West. Sunday should have been the least reliable day in this direction of travel as heavy traffic caused unreliability for travelers returning to the Bay Area from Lake Tahoe at the end of the weekend. After constructing PDFs for the opposite direction of travel, we see that this is in fact supported by the data (Figure 3-36).

The travel time variability by weekday on US 50 can be expressed in terms of the 95th percentile of travel time. This is presented for both directions of travel along with the mean by day in Table 3-8 below. Weekend travel patterns appear in the longer 95th Percentile travel times seen on Friday in the Eastbound direction and Sunday in the Westbound direction.

Figure 3-35: Week-long Distribution of Travel Times on US 50E
Table 3-8: Travel Time Reliability By Weekday

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>US 50 E - Mean Travel Time</td>
<td>9.6 min</td>
<td>8.1 min</td>
<td>8.9 min</td>
<td>8.3 min</td>
<td>12.4 min</td>
<td>14.3 min</td>
<td>10.4 min</td>
</tr>
<tr>
<td>US 50 E - 95th Percentile Travel Time</td>
<td>25.4 min</td>
<td>22.0 min</td>
<td>18.2 min</td>
<td>23.7 min</td>
<td>36.0 min</td>
<td>41.1 min</td>
<td>32.2 min</td>
</tr>
<tr>
<td>US 50 W - Mean Travel Time</td>
<td>12.7 min</td>
<td>11.5 min</td>
<td>10.7 min</td>
<td>11.3 min</td>
<td>14.4 min</td>
<td>12.3 min</td>
<td>13.1 min</td>
</tr>
<tr>
<td>US 50 W - 95th Percentile Travel Time</td>
<td>40.4 min</td>
<td>30.7 min</td>
<td>20.8 min</td>
<td>30.8 min</td>
<td>41.0 min</td>
<td>37.7 min</td>
<td>31.0 min</td>
</tr>
</tbody>
</table>

REFERENCES


4. Privacy Considerations

INTRODUCTION

As discussed in previous sections of this case study, innovations in data collection technology are providing exciting opportunities in the area of roadway travel time measurement. At the same time, use of these technologies is not without challenges, some technical, others related to protecting the confidentiality of personal information contained in ETC toll tag and Bluetooth mobile device datasets. As individual drivers’ privacy has the potential to be compromised when others have the ability to track their movements across the public roadway network, users of this data, both public and private, have developed a variety of plans and programs to ensure that data gathered in support of the generation of roadway travel times cannot be linked back to individuals. Recognizing that the data collection technologies described in this case study have the potential to raise public concerns over privacy, this section provides examples of the types of privacy protection policies and procedures currently in use by both public agencies and private sector companies to guard against the misuse of drivers’ personal information.

ELECTRONIC TOLL TAG-BASED DATA COLLECTION

Overview of Personal Privacy Concerns

When used for toll collection purposes, toll transponders are automatically identified whenever they pass within the detection zone of a compatible ETC reader. Every time this occurs, the ETC reader prompts the tolling system to deduct a pre-determined amount of money from the prepaid debit account associated with that transponder’s unique ID number. Recognizing that this technology would make it possible to track the path of each transponder-enabled vehicle between successive ETC readers, a number of agencies have deployed supplemental (non-revenue generating) ETC readers and back-office data analysis systems to facilitate the calculation of point-to-point travel times based on this data.

Although not instantaneous, direct connection exists between a toll transponder’s unique ID and the personal information of the transponder user, such data does exist within agency databases. As a result, this creates concerns for some users stemming from the potential loss of anonymity associated with their travel behavior.

Policies and Procedures in Place to Protect the Privacy of ETC Transponder Data

Two of the agencies best known for making use of anonymous ETC transponder data in support of travel time data collection are:

- Houston TranStar (Houston, TX)
- Metropolitan Transportation Commission (San Francisco Bay Area, CA)

Whereas both agencies have made significant efforts to protect the personal information of ETC toll tag users, only MTC has developed detailed guidelines concerning the use, archiving, and dissemination of this data.
**Houston TranStar.** Houston was the first city in the United States to apply ETC-based tolling technology to the collection of data concerning travel times and average speeds. The toll tag data on which this system is based is collected from ETC reader stations deployed at one to five mile intervals along over 700 miles of Houston area roads. Traffic Management Center (TMC) staff use this system to detect congestion along area freeways and high occupancy vehicle (HOV) lanes; this data is also provided to the public via media reports, travel times posted to roadside changeable message sign (CMS), and the Houston TranStar website. In an effort to protect the privacy of the driver’s from which travel time data is being collected, TranStar has configured its ETC readers to only store the last four digits of each toll tag’s ID number. Truncating ID numbers in this way creates an environment where the agency’s automated systems can track, but not identify, individual vehicles as they move across the data collection network. TranStar staff are acutely aware of drivers’ concerns regarding the protection of their personal information and have made efforts to inform the public that not only do they collect just a portion of each toll tag’s ID number, but also that none of the information concerning the movement of individual transponders is available for use by agency staff or law enforcement.

**Metropolitan Transportation Commission.** In support of its 511-traveler information service, the Metropolitan Transportation Commission (MTC) operates a travel time data collection system based on information collected from the region’s FasTrak toll system. As part of this effort, MTC takes the following steps to ensure the protection of toll tag users’ personal information [1]:

- Encryption software in the central software system encrypts each toll tag ID before any other processing is carried out to ensure that the toll tags are treated anonymously.
- Encrypted toll tag IDs are retained for no longer than twenty-four hours before being discarded. No historical database of encrypted IDs is maintained beyond that time period.

In addition to establishing the guidelines described above concerning the management of toll tag data, MTC has also developed the following principles regarding the protection of personal privacy [2]:

1. All traffic data collection activities will be implemented in a manner consistent with Federal and California laws governing an individual’s right to privacy;
2. The tag users’ consent will be secured before the operation of any data collection system based on toll tags;
3. No information about, or that is traceable to, any individual person will be collected, stored, or manipulated;
4. Information on the data collection, aggregation and storage practices will be available at the 511.org website, which will include traffic data collection methods, privacy policy, and full disclosure on the use of the data;
5. Members of the public will be given the ability to contact the program to discuss any privacy questions or concerns;
6. All recipients of the data shall comply with these privacy principles; and,
7. An annual evaluation will be conducted to assure that individual privacy is protected.
Although MTC provides the third-party contractors who operate its 511 and related services with access to the toll tag data collected as part of this system, as is indicated in items #6 and #7, above, these firms are required to observe all of MTC's privacy principles and are subject to an annual evaluation to verify their compliance.

**BLUETOOTH-BASED DATA COLLECTION**

**Overview of Personal Privacy Concerns**

Bluetooth-based travel time data collection systems operate, similarly to ETC-based systems, via the re-identification of mobile device ID data at successive locations along a roadway. However, whereas other technologies used to calculate roadway travel times based on the movement of probe vehicles (e.g., toll tag and license plate reader-based systems) have the potential, if abused, to directly link a specific user to the movement of their vehicle, identification of an individual based on their Bluetooth signature (i.e., MAC address) is much less straightforward. In theory, if the MAC address of the mobile device has been set by its manufacturer, the possibility exists, however remote, for a link to be established between the product part number and its owner via a product registration database or product warranty. Even so, the MAC addresses of mobile devices, though unique, are not linked to specific individuals or vehicles via any type of central database or user account.

Despite these facts, public perception regarding this method of data collection varies widely and has the potential interfere with its implementation. As a result, users of this technology have implemented a range of procedures to minimize the possibility of infringing on users’ privacy.

**Policies and Procedures in Place to Protect the Privacy of Bluetooth ID Data**

Two of the entities currently deploying Bluetooth-based data collection technologies for the purpose of calculating roadway travel times are:

- Post Oak Traffic Systems (Company utilized technology developed at the Texas Transportation Institute)
- Traffax (Company utilizing technology developed at the University of Maryland)

Users of Bluetooth-based data collection technologies stress that the MAC addresses collected by their systems are not directly associated with a specific user and do not contain any personal data or information that could easily be used to identify or “track” an individual person’s whereabouts. That said, all recommend taking additional steps to further ensure that the information collected from individual Bluetooth devices is kept as anonymous as possible.

**Post Oak Traffic Systems.** Techniques used by this firm to help protect the personal privacy of drivers include:

- Only polling the Bluetooth device information necessary to facilitate the calculation of travel times, including:
  - MAC address;
  - Device reader location; and,
Although other data can be accessed as part of the Bluetooth device polling process (e.g., device name and packets of information concerning data exchanged between a mobile phone and its associated Bluetooth headset), Post Oak staff recommend only collecting the data absolutely necessary to calculate segment travel times.

To further address potential privacy concerns, Post Oak field processors are programmed to encrypt all Bluetooth ID data immediately upon receipt. Doing so ensures that the actual device ID is not sent or stored anywhere.

**Traffax.** Company staff recommend implementing the following additional measures to ensure that no unauthorized use of data occurs. This includes [3]:

- Implementation of policies concerning the retention and dissemination of Bluetooth MAC address data, including:
  - Destroy or encrypt any base level MAC address information after processing.
  - Use industry standard encryption and network security. Proper security protocols, passwords, encryption and other methods should be incorporated into the data systems that store and process the MAC address data.
- Establishment of data processing safeguards (encryption and randomization) to prevent the recovery of unique MAC addresses:
  - Encryption methods transform MAC address data (at the sensor level) into an output form that requires special knowledge (such as an encryption key) to recover the original information. This activity preserves the uniqueness of the ID so that matching can still be performed without risking exposure of actual device ID data.
  - Randomization methods deliberatively degrade the data such that individual observations are no longer globally unique and the ability to track individuals based on their MAC addresses becomes theoretically impossible. A simple example of this would be to truncate the final 3 characters of the MAC ID.

All of the privacy protection methods recommended by Traffax are implemented at the sensor level, not the central processing station. Doing so makes it virtually impossible to obtain the complete and globally unique MAC address of any particular device.

**APPLICATION OF PRIVACY PRINCIPLES**

It has been amply demonstrated that travel time data collection technologies based on device re-identification (e.g., ETC toll tags and Bluetooth devices) have the potential to be abused in such a way as to cause significant privacy-related concerns. Although this section of the case study has reviewed a number of techniques currently
being utilized to further ensure that drivers’ anonymity is preserved, long-term acceptance of these technologies will ultimately rely on maintenance of the public’s trust. To that end, the Intelligent Transportation Society of America has established a set of Fair Information and Privacy Principles aimed at safeguarding individual privacy within the context of the deployment and operation of Intelligent Transportation Systems. Although advisory in nature, these principles are intended to act as guidelines for use by public agencies and private entities to protect drivers’ right to privacy. Principles include [4]:

- Individual Centered: Intelligent Transportation Systems must recognize and respect the individual’s interests in privacy and information use;
- Visible: Intelligent Transportation Information Systems will be built in a manner "visible" to individuals;
- Comply: Intelligent Transportation Systems will comply with applicable state and federal laws governing privacy and information use;
- Secure: Intelligent Transportation Systems will be secure;
- Law Enforcement: Intelligent Transportation Systems have an appropriate role in enhancing travelers' safety and security interests, but absent consent, statutory authority, appropriate legal process, or emergency circumstances as defined by law, information identifying individuals will not be disclosed to law enforcement;
- Relevant: Intelligent Transportation Systems will only collect personal information that is relevant for ITS purposes;
- Anonymity: Where practicable, individuals should have the ability to utilize Intelligent Transportation Systems on an anonymous basis;
- Commercial or Secondary Use: Intelligent Transportation Systems information stripped of personal identifiers may be used for non-ITS applications;
- Federal and State Freedom of Information Act (FOIA): FOIA obligations require disclosure of information from government maintained databases. Database arrangements should balance the individual’s interest in privacy and the public’s right to know; and,
- Oversight: Jurisdictions and companies deploying and operating Intelligent Transportation Systems should have an oversight mechanism to ensure that such deployment and operation complies with their Fair Information and Privacy Principles.

REFERENCES

5. LESSONS LEARNED

OVERVIEW

The team selected the Lake Tahoe region located in Caltrans District 3 in order to provide an example of a rural transportation network with fairly sparse data collection infrastructure. The data used as part of this case study was generated by electronic toll collection (ETC) readers on I-80 and Bluetooth-based data collection readers along I-5 and US 50 (see Figure 1-2). These readers register the movement of vehicles equipped with FasTrak tags (Northern California’s ETC system) and Bluetooth-based devices (e.g., Smart Phones) for the purpose of generating roadway travel times.

METHODOLOGICAL EXPERIMENTS

This case study examined vehicle travel time calculation and reliability using Bluetooth and RFID re-id systems. A number of factors were identified that influence travel time reliability and guided the development of methods for processing re-id observations and calculating segment travel times. The results show that smart filtering and processing of Bluetooth data to better identify likely segment trips increases the quality of calculated segment travel time data. This approach helps preserve the integrity of the data set by retaining as many points as possible, and basing decisions to discard points on the physical characteristics of the system, rather than their statistical qualities.

It is important to only filter out unlikely trips, so that the correctly measured variability of the data is not lost. The benefit of a more careful accounting procedure during the vehicle-identification stage allows for later statistical filtering of the data to be milder, preserving more meaning. Filtering trips based on statistical properties is less desirable because criteria for eliminating points are not based on the physical system. If all of the data points in an interval are valid, it does not make sense to discard that entire interval simply because it does not contain very many points. It is important to be aware of the interactions between preprocessing procedures. Future research may explore other smarter methods for filtering out unlikely segment trips. For example, considering observations across the entire BTR network would be useful for identifying unlikely segment trips.

A number of factors were found to influence vehicle segment travel times. For example, if the distance between BTRs was small, errors in calculated travel times may be significant and methods for determining passage time must be carefully considered. Signal strength availability enables easy and accurate determination of passage times. Without signal strengths, using arrival and departure times for passage times may improve travel time accuracy. This was found to be likely for BTR #10 based on the location of the reader relative to an intersection, the intersection configuration, and the short distance to the nearest BTR. Aggregating observations into visits was also found to be useful for distinguishing between trip and travel time for individual vehicles at a BTR.
USE CASE ANALYSIS

This case study explored four aspects of the ETC and Bluetooth reader networks used in the Lake Tahoe case study: (1) detailed locations and mounting structures; (2) lanes and facilities monitored; (3) percentage of traffic sampled; and (4) percentage and number of vehicles re-identified between readers. As a whole, it showed that vehicle re-identification technologies are suitable for monitoring reliability in rural environments, provided that traffic volumes are high enough to generate a sufficient number of samples. For rural areas that have heavy recreational or event traffic, vehicle re-identification technologies such as ETC and Bluetooth can provide sufficient samples to calculate accurate average travel times at a fine granularity during high-traffic time periods. During these high-volume periods, vehicle re-identification technologies can be used to monitor travel times and reliability over long distances, such as between the rural region and nearby urban areas.

For agencies deploying vehicle re-identification monitoring networks, it is necessary to understand that the quality of the collected data is highly dependent on the decisions made during the design and installation process. The mounting position and antennae configuration of ETC readers impacts the number of lanes sampled at a given location. The positioning of Bluetooth readers, which have a large detection radius, dictates whether ramp, parallel facility, or multi-modal traffic are also sampled, which can introduce large errors into travel time and reliability computations. In addition to choosing an optimal positioning of readers, it is also important to place and space them appropriately. Readers should be placed where they can provide travel time information for heavily traveled origins and destinations. Because vehicle re-identification readers can be easily moved, there are opportunities to do pilot tests to evaluate the quality, quantity, and value of collected data, so that the final deployment robustly supports the desired measures.

For agencies leveraging existing networks, it is important to fully understand the configuration of the network before using its data. At a minimum, this should include taking steps to verify that reader locations are correct and that the computed travel times and number of matches are reasonable given the distance and known traffic patterns between reader pairs. In locations where readers are closely spaced, computing reader hit rates and comparing between readers can help identify the reader most suited for monitoring travel times at a given location. Finally, evaluating percentage and volume of matched reads between each reader pair by time of day and day of week can indicate which time periods typically have sufficient matches to support average travel time computations at different granularities.

This case study also explored an approach for isolating and exploring the effects of weather and weekend travel on travel time reliability. As implemented here, the analysis should be fairly straightforward to replicate with data from a travel time reliability monitoring system such as PeMS and the appropriate weather data. The PDFs of travel times under different operating conditions consistently demonstrated the unreliability associated with low visibility, rain, and travel under high-demand conditions. This use case also described the travel time unreliability associated with such events in terms of 95th percentile travel time. Taken together, these tools should be valuable to planners, operators, and engineers interested in analyzing and communicating the travel time reliability of a section of roadway, especially one of a rural nature. Finally, application of the research team’s approach has revealed several insights into the nature of working
with Bluetooth and ETC-based sources of data. Specifically, that due to the nature of how these data collection technologies calculate travel time, it is necessary to account for artificially long travel times likely contained in the data set prior to conducting any analysis. Despite this shortcoming, these technologies both provide users with the potential to effectively assess roadway travel times and consequently, reliability of travel, in rural areas where the cost of deploying and maintaining spot-based sensors (e.g., loop detectors) makes their use impracticable.

PRIVACY CONSIDERATIONS

For either of the data collection technologies described in this report to be successful over the long-term, safeguards must be put into place to ensure that the privacy of individual drivers being sampled is protected. With this in mind, we recommend that any probe data collection program implemented by public agencies or by private sector companies on their behalf adhere to a pre-determined set of privacy principles (e.g., ITS America’s Fair Information and Privacy Principles) aimed at maintaining the anonymity of specific users. Additionally, any third party data provider working for a public agency to implement a travel time data collection solution based on either of the technologies described in this Case Study should be required to submit an affidavit indicating that they will not use data collected on the agency’s behalf in an inappropriate manner, including:

- Renting, leasing, selling, or otherwise providing data to any entity without explicit written permission of the agency;
- Using data for any purpose(s) other than those described as part of the project-specific requirements; and,
- Attempting to identify the ownership of individual vehicles or devices whose personal information is collected as part of the system’s data collection infrastructure.