Hybrid Data Implementation

Identifying the Need

Every Caltrans district generates a quarterly report, called a Mobility Performance Report (MPR), that summarizes key performance measures such as Vehicle Miles Traveled (VMT), Vehicle Hours Traveled (VHT), and Vehicle Hours of Delay (VHD). To compute VHD and VMT, data from Vehicle Detector Stations (VDS) are used. The VDS include 40,000 individual detection zones. Maintaining such vast infrastructure requires extensive operational and maintenance support. The availability of third-party, vendor-provided data can augment data from VDS to estimate performance measures, as required data (such as speed or travel time) can be obtained from third-party vendors. This research explores the feasibility of incorporating third-party traffic data to enhance or replace the existing infrastructure, and to improve the quality of the MPR.

What is the goal?

The primary purpose of this research is to determine how to incorporate third-party data into the Caltrans PeMS (Performance Measurement System) to provide both real-time and historical performance metrics. This requires evaluating and modifying algorithms currently used in Caltrans PeMS. Key goals include the following:

- Reduce costs and increase coverage of traffic monitoring
- Provide methodology for calculating VHD in a hybrid model (traditional plus third-party data)
- Enable smarter deployment of point-based sensors
- Provide a roadmap strategy for using third-party data

Project Description

This report proposes a methodology for using third-party data to improve Caltrans performance reporting, investigates the advantages and opportunities that come with using this data, and provides a roadmap for Caltrans to move forward with a pilot study.

A survey of data offerings by third-party vendors was performed and used to define the characteristics of third-party data. Third-party data can provide information like date, timestamp, link identifier, link length, speed, and travel time. Travel time data is useful for estimating VHD, a key goal of this research. The PATH research team created detailed guidelines showing what data is required from third-party vendors, helping Caltrans develop a roadmap strategy for comparing these vendors. To test the advantages of using third-party data in addition to the traditional point-based sensors, the PATH team also developed algorithms to develop a reporting method for VHD. This approach involved using third-party data, point-based sensor data, and annual average daily traffic (AADT). These algorithms were evaluated using a microsimulation model of the I-210 Integrated Corridor Management (ICM) Corridor.

Currently, each VDS inherits its freeway association from the controller it is connected to--one single control box may handle multiple freeways at an interchange but can only be associated with one freeway. This makes it difficult to associate the data with its location on the freeway. Point-based sensors collect vital data at points across the network, and the advantages for fixed sensors are necessary for more complex traffic conditions or where traffic control is required. In general, the removal of these locations resulted in an increase in error. One solution to fix the freeway association problem would be to associate the sensors on the pavement with a network map to show exact locations by adding an additional table to PeMS. This approach would clarify the locations of VDS by assigning each sensor its own latitude and longitude coordinates. It would also enable automated checking of configuration and data integrity, improving the ability to fill in any missing data.

There are four main methods that were compared for estimating VHD: (1) Traditional data and calculation, which uses point-sensor data; (2) Third party data and traditional calculation; which is only possible when spatial reference systems match; (3) Hybrid calculation; which is needed when spatial reference systems do not match, and (4) Adjustments for limited instrumentation, which uses rough estimates for flows and third-party data for travel times. The recommended VHD estimation method depends on the infrastructure type and the data available. For freeway mainlines, the best performance was achieved with the hybrid calculation. For HOV lanes, traditional traffic sensing methods must be used until third-party data become available that offer the precision to reliably distinguish HOV lanes from mainline lanes. For connectors, good performance was obtained using third-party data combined with the traditional calculation. For ramps and arterials, further work is required.

Recommendations have been provided to Caltrans for improving the configuration data of the Caltrans PeMS. A framework was presented to determine: (1) where improvements are possible when merging third-party data to determine delay; and (2) the error introduced into VHD estimates as point-based sensor data is removed. This framework was then used to evaluate the performance of the algorithms for a range of operating conditions and a strategic roadmap was proposed for incorporating third-party data into the PeMS so the districts can use the VHD performance measure whenever needed.

Projected Benefits to California

The deployment of a traffic management system that uses hybrid traffic data will provide the opportunity to reduce costs and increase coverage of traffic monitoring, improve the quality and accuracy of performance reports, and provide information upon which to make better prioritization of resources and better investment decisions. A hybrid approach enables the strengths of third-party data to offset the weaknesses of point-based sensors and vice versa.

What is the progress to date?

The project is complete. A huge challenge of this research was to ensure that the data from multiple sources could be projected onto the same domain of analysis -- systems used by PeMS and by multiple vendors do not generally align easily. After careful evaluation, findings suggest that the use of third-party data can compensate for loss of point-based sensors and can be used to roughly estimate delay. Third-party mobile data also offers advantages over point sensors for longer road segments, reducing the need for frequently spaced detector stations and potentially extending coverage to remote roadways with limited equipment. However, point-sensors offer full counts and lane-specific data.

Based on the analysis, the hybrid approach provides the best estimates of performance measures. The analysis was conducted for both freeway-mainline and freeway-freeway connectors, and for five different times of day during weekday hours. For almost all the scenarios, the inclusion of travel time data reduces the estimation errors. These benefits hold when the number of fixed detectors is reduced. Recommendations to Caltrans include pre-selecting sites for a limited pilot and performing an initial assessment of data consistency. Precise VDS location information at freeway-freeway connectors needs to be obtained, and analysis needs to be performed to focus on existing sensors. The next steps would involve launching a limited pilot program by selecting well-studied freeways with excellent data and determining the accuracy with real-world data instead of simulation. This will require comparing the data quality of various third parties to create an initial set of freeways with high-quality, reliable data. Launching a full-scale pilot in the selected district, assessing the cost and difficulty of data integration over a limited geographical region, and assessing the value of a hybrid, integrated traffic data solution would be the final steps. This research was done with sponsorship from Caltrans.

Final Report

Hybrid Data Implementation: Final Report for Task Number 3643 (escholarship.org)

About the Authors

The authors of this report are researchers at California PATH. Dr. Nicholas Fournier is a postdoctoral researcher interested in transportation and planning and demand modeling. Dr. Mahmud Khan is a postdoctoral researcher interested in Hybrid Data and Connected Corridors. Dr. Anthony Patire is a Program Leader and Research and Development Engineer. He builds large-scale traffic microsimulation models and is interested in traffic engineering, big data, and machine learning. Dr. Michael Mauch is a Senior Development Engineer who supports all phases of research projects and is responsible for crafting complete research reports. Dr. Alexander Skabardonis is an internationally recognized expert in traffic flow theory, traffic management, and control systems.

CALIFORNIA PATH Berkeley Institute of Transportation Studies