I-95 Corridor Coalition’s Vehicle Probe Project

A cross-vendor and cross-state analysis of the GPS-probe data latency
OUTLINE

1. INTRODUCTION
2. LITERATURE REVIEW
3. METHODOLOGY
4. CASE STUDY
5. CONCLUSIONS
What is Latency?

- **Latency** is the time difference between GPS-probe data and real traffic condition;
- It describes the punctuality of data;
- It is crucial to real-time applications.
Where does the latency come from?

- GPS-Probe Data
- Sensor Data
- Incident Data
- Historical Data

Real-Time Data Processing Engine

Vendor Data Server

User application

Timeline:

- T1 Field data recorded
- T2 Field data received
- T3 Probe data generated
- T4 Probe data is made available
- T5 Probe data received by user

Delay caused by the telecommunication medium
Delay dictated by the processing time of the data fusion algorithms
Delay caused by the internal processes to generate the real-time feed
Latency due to query time, internet speed, etc.
Definition of Latency

Latency is defined as “the difference between the time the traffic flow is perturbed and the time that the change in speed is reflected in the data.”
Related Research

  - Objective: Maximizing Correlation Coefficient
  - Average latency: 6.8 minutes
  - Maximum latency could exceed 10 minutes in many time periods
    - Reference data: Loop detector
    - Aggregating data into 10 second time interval
Related Research

- INRIX. (2007). Traffic Data and Associated Services along the I-95 Corridor
  - INRIX will deliver current speed, travel time, average speed...with latency on average of 4.5 minutes.

  - Propose methodology to measure GPS-probe data latency in comparing to Bluetooth data. It is shown to be effective, but only on a limited dataset for one GPS-probe data vendor.
Contributions

✓ Applying the latency measurement methodology (Wang et. al 2017) to a larger dataset
  ✓ Expanding to three states
  ✓ Data from three major probe data vendors
✓ Developing a robust algorithm for automating slowdown episode detection
✓ Revisiting conclusions of the previous research based on the extended data set;

✓ Analyzing latency for different vendors to describe:
  ➢ Latency distribution for different vendors;
  ➢ Latency distribution at slowdown and recovery periods;
  ➢ Latency distribution at different times of the day;
  ➢ Latency distribution for segments of various lengths;
  ➢ Statistical comparison of latency across three vendors.
Data Processing

Bluetooth/WiFi data preparation

- Matching and filtering high resolution re-identification based travel time observations to generate segment space mean speed data;
- Aggregating reference data in one minute intervals.

Data Filtering

The following observations are identified and discarded:

- Observations with unreasonably low or high speeds;
- Observations in a particular time interval that are far from the average of the rest of the speeds observed in the same time interval.
The average of the neighboring observations is considered as the travel speed for the missing interval (only applied to less than or equal to 5 mins).

\[ s_{t+i} = s_t + \frac{i}{n+1} \left( s_{t+n+1} - s_t \right) \quad \forall i = 1, 2, 3 \]

Missing Interval: \( n=3 \)

\[
\begin{align*}
\ldots & & \quad s_{t-1} & & s_t & & s_{t+i} & & s_{t+n+1} & & s_{t+n+2} & & \ldots \\
& & & & & & i=1 & & i=2 & & i=3 & 
\end{align*}
\]
Data Processing

Data Smoothing

- Weighted moving average function

\[ y_k = 0.33x_k + 0.27x_{k-1} + 0.20x_{k-2} + 0.13x_{k-3} + 0.07x_{k-4} \]

Arithmetic growth of the weights with the previous five time intervals

Why Zero-phase digital filtering? Smoothing may introduce undesired horizontal shift into the time series, however in Zero-phase digital filtering, forward shift is followed by a reverse shift, so the artificial shift is compensated.
**Data Processing**

**Data Smoothing**

- Less noise
- No shift
Data Processing

Flow chart

Bluetooth Data Processing

1. Bluetooth Field Data
2. Paired Travel Time Data
3. Filter
4. Paired Data without Outliers
5. Convert
6. Segment Travel Speed Data
7. Aggregate
8. One-min Speed Data
9. Interpolate
10. Speed Data without Gap
11. Smooth
12. Smoothed Speed Data

GPS-probe Data Processing

1. GPS-probe Data
2. Filter (rule out data with low confidence score)
3. Path data
4. Interpolate
5. GPS-probe Data without Gap
6. Smooth
7. Smoothed GPS-probe Data

Calculating Latency

I-95 CORRIDOR COALITION

CATTWORKS
Objective

- Find the shift distance that maximizes the overlapping of Bluetooth data and GPS-probe data.

\[
\begin{align*}
\min f_1 &= \sum_{t=1}^{n} \left| S_{t}^{BT} - S_{t-\text{latency}}^{\text{probe data}} \right| & (AVD) \text{ Absolute vertical distance between two curves} \\
\min f_2 &= \sum_{t=1}^{n} \left( S_{t}^{BT} - S_{t-\text{latency}}^{\text{probe data}} \right)^2 & (SVD) \text{ Squared vertical distance, which gives more weights to the points that have bigger difference} \\
\min f_3 &= \text{corr}(S_{t}^{BT}, S_{t-\text{latency}}^{\text{probe data}}) & (COR) \text{ Statistical representation of the linear relationship between two curves}
\end{align*}
\]

\[lb \leq \text{latency} \leq ub\]
Minimize Absolute Vertical Distance \((f^I)\) – as example
# Data Selection

<table>
<thead>
<tr>
<th>State</th>
<th>Road</th>
<th>Start Point</th>
<th>End Point</th>
<th>Length (mile)</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Carolina</td>
<td>I-85</td>
<td>US-276/Exit 48</td>
<td>SC-14/Exit 56</td>
<td>7.15</td>
<td>12/03/2015 - 12/17/2015</td>
</tr>
<tr>
<td></td>
<td>I-26</td>
<td>Bush River Rd/Exit 108</td>
<td>Harbison Blvd/Exit 103</td>
<td>4.47</td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>I-89</td>
<td>I-93</td>
<td>Stickney Hill Rd/Exit 3</td>
<td>3.54</td>
<td>07/10/2016 - 07/24/2016</td>
</tr>
<tr>
<td></td>
<td>I-93</td>
<td>NH-28/Rockingham Rd/Exit 5</td>
<td>NH-102/Nashua Rd/Exit 4</td>
<td>3.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I-26</td>
<td>I-40/ Exit 46 A/US 74</td>
<td>I-26/Exit 37</td>
<td>14.43</td>
<td></td>
</tr>
</tbody>
</table>
Data Selection

Road type: Freeway

Location: South Carolina, New Hampshire, North Carolina

Direction: both directions

Time: 2 weeks for each state

Segment length: 1~3.5 mile
Data Selection

New Hampshire

I-89/93

North Carolina

I-240

I-40/26
Data Preparation
Capturing Slowdown and Recovery Episodes

Travel Speed (mph)

Slowdown period

Recovery period

Transition from speed slowdown to recovery

Slowdown Starts

Speed recovery ends

Bluetooth data

GPS Probe data
Data Preparation: Automated Speed Pattern Recognition Filter (ASPRF)

<table>
<thead>
<tr>
<th>State</th>
<th># of paths</th>
<th># of days</th>
<th># of Slowdown and Recovery Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Carolina</td>
<td>14</td>
<td>15</td>
<td>72</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>12</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>North Carolina</td>
<td>15</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>41</td>
<td>116</td>
</tr>
</tbody>
</table>

Minimum Slowdown Threshold = 0.4 × Free Flow Speed
Sample Result

South Carolina
12/10/2015
Length: 2.3 miles
Data: 3 vendors

Legend
- Reference (Bluetooth)
- Original (GPS-probe data)
- Absolute Vertical Distance
- Squared Vertical Distance
- Correlation coefficient

(a) Vendor 1
(b) Vendor 2
(c) Vendor 3
## Results

Average Latency for all segments over all periods (minute)

<table>
<thead>
<tr>
<th></th>
<th>Vendor 1</th>
<th></th>
<th></th>
<th></th>
<th>Vendor 2</th>
<th></th>
<th></th>
<th></th>
<th>Vendor 3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AVD</td>
<td>SVD</td>
<td>COR</td>
<td>Mean</td>
<td>AVD</td>
<td>SVD</td>
<td>COR</td>
<td>Mean</td>
<td>AVD</td>
<td>SVD</td>
<td>COR</td>
<td>Mean</td>
</tr>
<tr>
<td>SC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRE</td>
<td>3.9</td>
<td>4.4</td>
<td>4.6</td>
<td><strong>4.3</strong></td>
<td>6.1</td>
<td>6.2</td>
<td>6.2</td>
<td><strong>6.2</strong></td>
<td>2.9</td>
<td>3.1</td>
<td>3.1</td>
<td><strong>3.0</strong></td>
</tr>
<tr>
<td>SP</td>
<td>3.6</td>
<td>3.6</td>
<td>4.0</td>
<td><strong>3.7</strong></td>
<td>6.7</td>
<td>6.8</td>
<td>6.6</td>
<td><strong>6.7</strong></td>
<td>4.4</td>
<td>4.3</td>
<td>3.8</td>
<td><strong>4.2</strong></td>
</tr>
<tr>
<td>RP</td>
<td>4.7</td>
<td>5.3</td>
<td>4.0</td>
<td><strong>4.7</strong></td>
<td>5.0</td>
<td>5.2</td>
<td>5.0</td>
<td><strong>5.1</strong></td>
<td>2.0</td>
<td>2.4</td>
<td>2.5</td>
<td><strong>2.3</strong></td>
</tr>
<tr>
<td>NH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRE</td>
<td>3.3</td>
<td>3.4</td>
<td>3.6</td>
<td><strong>3.4</strong></td>
<td>5.6</td>
<td>6.0</td>
<td>6.4</td>
<td><strong>6.0</strong></td>
<td>2.3</td>
<td>2.5</td>
<td>2.7</td>
<td><strong>2.5</strong></td>
</tr>
<tr>
<td>SP</td>
<td>3.3</td>
<td>3.2</td>
<td>3.2</td>
<td><strong>3.2</strong></td>
<td>7.0</td>
<td>7.0</td>
<td>6.7</td>
<td><strong>6.9</strong></td>
<td>3.5</td>
<td>3.0</td>
<td>3.2</td>
<td><strong>3.2</strong></td>
</tr>
<tr>
<td>RP</td>
<td>4.0</td>
<td>4.2</td>
<td>3.9</td>
<td><strong>4.0</strong></td>
<td>3.6</td>
<td>3.7</td>
<td>4.7</td>
<td><strong>4.0</strong></td>
<td>2.9</td>
<td>3.5</td>
<td>2.6</td>
<td><strong>3.0</strong></td>
</tr>
<tr>
<td>NC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRE</td>
<td>3.5</td>
<td>4.1</td>
<td>4.1</td>
<td><strong>3.9</strong></td>
<td>7.3</td>
<td>7.8</td>
<td>7.3</td>
<td><strong>7.5</strong></td>
<td>3.2</td>
<td>3.2</td>
<td>2.7</td>
<td><strong>3.0</strong></td>
</tr>
<tr>
<td>SP</td>
<td>4.6</td>
<td>4.6</td>
<td>3.6</td>
<td><strong>4.2</strong></td>
<td>7.7</td>
<td>7.8</td>
<td>7.4</td>
<td><strong>7.6</strong></td>
<td>3.2</td>
<td>2.9</td>
<td>2.5</td>
<td><strong>2.9</strong></td>
</tr>
<tr>
<td>RP</td>
<td>3.4</td>
<td>3.7</td>
<td>4.1</td>
<td><strong>3.7</strong></td>
<td>5.9</td>
<td>6.5</td>
<td>6.8</td>
<td><strong>6.4</strong></td>
<td>3.2</td>
<td>3.1</td>
<td>3.8</td>
<td><strong>3.4</strong></td>
</tr>
</tbody>
</table>

**SC**: South Carolina, **NH**: New Hampshire, **NC**: North Carolina; **AVD**: Absolute vertical distance; **SVD**: Squared vertical distance; **COR**: correlation coefficient; **SRE**: Slowdown and recovery episode, **SP**: Slowdown period, **RP**: recovery period.
Results

Latency comparison between speed slowdown and recovery periods

(a) SRE

(b) SP

(c) RP

✓ Comparison results vary from vendor to vendor.
✓ Latency is different during slowdown and recovery.

Note: whiskers represent the 1.5 interquartile range of the lower and upper quartiles, corresponding to approximately $+/-2.7$ times standard deviation and 99.3 percent coverage if the latency is normally distributed.
Results

Latency histogram

(a) SRE

(b) SP

(c) RP
Is difference in probe data latency statistically significant between pair of vendors?

<table>
<thead>
<tr>
<th>P-value</th>
<th>t-stat</th>
<th>Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor 1</td>
<td>Vendor 2</td>
<td>Vendor 3</td>
</tr>
<tr>
<td>Vendor 1</td>
<td>-9.370</td>
<td>4.729</td>
</tr>
<tr>
<td>Vendor 2</td>
<td>0.000</td>
<td>14.185</td>
</tr>
<tr>
<td>Vendor 3</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- Statistically speaking, average latency among vendors is different.
- The latency distribution is statistically independent among vendors.
**Result**

**Time of the day**

(a) SRE

(b) SP

(c) RP
There is no strong relationship between the segment length and latency for all vendors.
ConclusionS

- Across all vendors and for all tested freeway segments, average latency is found to be **4.4** minutes with standard deviation of 2.3 minutes;
- Average latency is statistically different among vendors;
- No strong correlation between latency and time of the day, and also latency and segment length were found.
- Instead of interpreting the latency as a single number, distribution of latency should be measured and evaluated
Future Research

- Investigating the latency of GPS-probe data on signalized arterials.
- Studying potential impact of traffic volume on latency of probe data.
- Keep analyzing more cases in future to study long time trend of latency improvement for different vendors.
- Understanding implications of the latency and its distribution on real-time applications and exploring solutions to compensate for latency.
Thank you
Q & A

Masoud Hamedi*
Senior Research Scientist
Center for Advanced Transportation Technology
University of Maryland
5000 College Ave
College Park, MD 20740
Email: masoud@umd.edu | Phone: 301-405-2350

Zhongxiang Wang
Graduate Student
Department of Civil & Environmental Engineering
University of Maryland
Email: zxwang25@umd.edu

Elham Sharifi
Faculty Research Assistant
Center for Advanced Transportation Technology
University of Maryland
Email: esharifi@umd.edu

Stanley Young
Advanced Transportation and Urban Scientist
National Renewable Energy Laboratory
Email: stanley.young@nrel.gov