1. ABSTRACT
In this work, we present an empirical analysis of a real-world taxi fleet system, based on GPS location data. Our current dataset records the movement of 6,230 taxis (3.6 million observations), as well as the time and location of 38,048 booking requests. Both datasets were collected over a 24-hour period in the city-state of Singapore. Each taxi reports its GPS coordinates, vehicle speed, and operating state at regular intervals. The main goals of our initial analysis were (a) to characterize the efficiency of the taxi system, and (b) to explore the sources of inefficiency, with a view to mitigating them.

1.1 Locations and Bookings
Figure 1 shows a subset of the observed taxi movements (0.3% of the data set sampled uniformly at random) plotted by their geographic coordinates. We divided the city into four zones based on exogenously specified local parameters, e.g., the nature of households and relative density of business establishments. The points within each zone are shaded according to the density of bookings in that zone, with the highest shaded black.

The figure helps to validate the accuracy of our GPS data, as the overall picture closely resembles a map of Singapore. It shows the major roads and highways, which, unlike the road networks typically considered in simulation studies, are not arranged in a grid. It also suggests patterns of traffic congestion that appear as differences in density, providing motivation for using GPS data to estimate real-time traffic flow in the future.

1.2 Occupancy and Waiting time
To reason about the efficiency of the taxi-dispatch system, we say that an efficiently balanced taxi system is one in which the satisfaction of passengers and drivers is inversely correlated. That is, when passengers are happy because of short waiting times, drivers are unhappy because of low occupancy, and vice versa.

Our index of driver satisfaction is the ratio $O/(O+A)$, where $O =$ the number of minutes the taxi was occupied in the given zone and hour, and $A =$ the number of minutes it was available. We call this ratio the occupancy rate.

We measure passenger satisfaction by computing the median waiting time for a booking request using the following method: First, we measure the time lag between a booking request and the pickup that satisfies it. Second, we observe that, especially during peak periods, some booking requests may never be satisfied. From the customer’s perspective, this is equivalent to an infinite waiting time. As long as fewer than half of the booking requests in a zone-hour are unsatisfied, the median waiting time will be finite. Otherwise we arbitrarily define it as 20 minutes, which is five minutes longer than the longest waiting time for a successful pickup.

Figure 2 shows the graph of occupancy rate and median waiting time in the given zone for each hour in the day. It reveals systematic differences by zone in the efficiency of taxi allocation. It also highlights situations in which either waiting time spikes without a similar rise in occupancy, or occupancy drops without a similar reduction in waiting time. In the former case, customers are worse off but drivers are no happier. In the latter case, drivers are worse off but customers are no happier. Overall, across all zones, waiting time and occupancy are positively correlated; i.e., longer waiting times (unhappy customers) usually indicates higher occupancy (happy drivers) and vice versa.

Our initial analysis shows the power of our approach in identifying potential trouble spots in a simple and repeatable way. In the future, we plan to do the following: (a) collect a larger dataset, (b) develop mechanisms to improve system performance from multiple perspectives (operators, drivers, and passengers), and (c) test and refine our mechanisms in real situations.
Understanding and Improving a GPS-based Taxi System
Darshan Santani, Rajesh Krishna Balan, and C. Jason Woodard

Objective

- To describe the **spatio-temporal efficiency** of a large GPS-equipped taxi cab fleet.
- To identify potential **hotspots** characterized by unusually long wait times and/or **colds spots** with unusually low taxi occupancy.
- To explore high level system characteristics from the different, possibly conflicting, perspectives of passengers, taxi drivers and system operators.

Summary Statistics

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS Updates per Hour (no.)</td>
<td>33.35</td>
<td>33.70</td>
<td>9.48</td>
</tr>
<tr>
<td>Average time spent on the road (hrs)</td>
<td>17.99</td>
<td>18.38</td>
<td>4.18</td>
</tr>
<tr>
<td>Average Speed (km/h)</td>
<td>38.94</td>
<td>33.00</td>
<td>39.47</td>
</tr>
<tr>
<td>Street Pickups (no.)</td>
<td>21.1</td>
<td>21.0</td>
<td>8.50</td>
</tr>
<tr>
<td>Booking Pickups (no.)</td>
<td>4.6</td>
<td>4.0</td>
<td>3.64</td>
</tr>
<tr>
<td>Total Pickups (no.)</td>
<td>25.7</td>
<td>26.0</td>
<td>9.68</td>
</tr>
<tr>
<td>Average Occupancy Rate (%)</td>
<td>42.46</td>
<td>46.17</td>
<td>17.24</td>
</tr>
</tbody>
</table>

* Per Taxi for Full Day

Can GPS data be used to estimate real time traffic conditions?

Location and Bookings

How can we develop mechanisms to disseminate local information to taxi drivers in real time?

Taxi Bookings

Can GPS data be used to estimate real time traffic conditions?

Location and Bookings

Can GPS data be used to estimate real time traffic conditions?

Location and Bookings

How can we develop mechanisms to disseminate local information to taxi drivers in real time?

Occupancy & Waiting Time

- Reveals systematic differences by zone in the efficiency of taxi allocation, as well as periods in which efficiency deviates positively or negatively from average.
- Clearly highlights situations in which either waiting time spikes without a similar rise in occupancy, or occupancy drops without a clear reduction in waiting time.
- Overall, waiting time and occupancy are **positively correlated**, i.e., longer wait times (unhappy customers) usually indicate higher occupancy (happy drivers) and vice-versa.

Cruising Behavior

How can the cruising behavior of taxi drivers be optimized to gain higher operating efficiency?