Traffic congestion is a global problem that results in tremendous personal and social costs. In the United States alone, cars stuck in traffic waste billions of gallons of fuel every year; a 2005 study from the Texas Transportation Institute estimated that more than 5 times as much gas is wasted in traffic congestion today compared to twenty years ago in urban areas in the US. Although several Web sites provide driving directions, they don’t take traffic conditions into account (or when they do, the information they use is sparse and notoriously out-of-date).

Our research objective is to provide an effective navigation system for cars that uses both historical and real-time traffic data to determine good driving directions. Our algorithms take traffic delay distributions as the primary input, deriving these from historical data, and using real-time notifications as additional information. By collecting traffic delay data and categorizing it according to location (road segment), time of day, day of week, and time of year, as well as correlating it with the occurrence of various events (e.g., concerts, sporting events, etc.), we can model the statistics of traffic delays on road segments. These statistical models will then be used for two purposes: first, to identify the locations and times of traffic “hot spots” that drivers should avoid, and second, to use in new algorithms for computing efficient paths according to various delay criteria. In particular, our interest is in developing algorithms that use delays rather than simply distance.

The first part of the system requires a data gathering system to obtain traffic delay data. As part of the MIT CarTel project (http://cartel.csail.mit.edu), over the past three years we have developed and deployed such a system on a fleet of taxis in the Boston area. CarTel is a mobile sensor network; each car carries a wireless embedded computer and a collection of sensors, including GPS. The result is a network capable of obtaining data about a large metropolitan area at low cost, compared to current approaches that use traffic sensors in roads. We have gathered many gigabytes of traffic delay information over the past year, covering several thousand hours of driving.

The second part of the system converts this data into a statistical model of traffic delays segmented according to various temporal and spatial categories. This task is challenging for several reasons: first, data volumes are quite high, so efficient algorithms and indices are needed. Second, the data is noisy, so algorithms that match GPS points to an underlying road database in a robust and error free way are essential. Third, determining how to segment the data into a collection of road segments non-trivial; simply computing statistics on a per-intersection granularity leads to a huge road graph with questionable statistics for each segment. However, making the segments too large loses information about alternate path segments to route cars along.

The third part of the system uses these statistical distributions of road delays to produce delay- and congestion-aware paths. We have developed new algorithms that optimize various delay criteria. Our contributions here focus on developing algorithms that are tractable in a real implementation, and in coping with the non-idealities (such as non-independence of delays on different segments) that arise in practice. In the worst case, stochastic shortest path problems are horribly intractable, requiring the exploration of a search space that is super-exponential in the number of road segments.

In addition to finding paths that minimize shortest expected time, we also allow users to find paths that take a desired deadline into consideration. For example, if you want to go from your home to the airport tomorrow morning such that you reach the airport by 8am, what is the latest time you can leave home and what route should you use, such that you make it by that time with high probability?

We will demonstrate our system, which is a browser accessible web application, with data from our taxi-cab deployment in the Boston area. Our demonstration will include:

1. A live view of the current data from cabs, showing their current position and speed.
2. An illustration of the processing pipeline that transforms this live data into delay statistics that are used for our routing algorithms.
3. A web service that allows users to select start and destination routes and observe paths subject to a variety of delay criteria. The interface includes the ability to compare predicted routes to those produced by the Google Maps.
4. “MyDrive”, a feature that allows users to upload their own drives and visualize them in the system as color coded routes on top of a map, where colors indicate the speed on a particular section of road.