A comprehensive spatial-temporal connectivity measure for transit systems

Joseph Choi

E-mail: [joecmama@gmail.com](mailto:joecmama@gmail.com)

John C. Handley (Corresponding author)

Palo Alto Research Center, Inc.

800 Philips Rd., MS 128-27E

Webster NY 14450

TEL: 585-265-8900

E-mail: [John.Handley@parc.com](mailto:John.Handley@parc.com)

Brian Roach

E-mail: [brian.roach@tamu.edu](mailto:brian.roach@tamu.edu)

Word count: ???

Number of Figures: ? x 250

Number of Tables: ? x 250

Total words: ????

**ABSTRACT**

One of the major issues with vehicle to infrastructure (V2I) communications is data quality. Based on sampled data from University of Michigan’s Safety Pilot Model Deployment we find issues with dropouts, “stuck at zero”, and trajectories that are not physically plausible. In response, we have developed an algorithm to reconstruct missing data based on a system model coupled with a physical model that constrains the reconstructed data to that which is realistic. The system model is a Kalman filter where the missing data is inferred by expectation-maximization. [Something on the physical model]. We illustrate the efficacy of this approach using 75 time series from the Safety Pilot Model Deployment data. It comprises nine dimensions, latitude, longitude, elevation, speed, heading, longitudinal, lateral acceleration, vertical acceleration, and yaw rate.

**1. INTRODUCTION**

Connected vehicle technology promises to usher in a new era of automobile safety. Using dedicated short range communication (DSRC) at 5.9Ghz, vehicles will soon communicate to each other (vehicle to vehicle or V2V), to the infrastructure (V2I) or to pedestrians (V2P). It is estimated that 80% of serious accidents could be avoided using communication systems. The goal in the US is to have 25% penetration rate by 2025.

In addition to safety, other applications include traffic flow monitoring for traffic management, road condition for asset management and driver alerts, improved signal timing

In the V2I scenario, a vehicle comes within range of road side equipment (RSE), within as much as 300 meters by direct line of sight, performs a handshake, is assigned a temporary, unique identifier, and then pushes a variety of on-board sensor readings to the RSE. Among the safety applications envisioned are curve speed warning, where a RSE is mounted close to a curve and based on the detected speed and heading, a vehicle could be warned, either through road-side signage or an in car display, that the car is approaching the curve too fast for safety. A second scenario is a vehicle approaching a signalized intersection. If the vehicle is approaching too fast based on the signal phase and timing (SPaT) and other vehicles in the intersection, an alarm could be raised to alert the driver to brake, or even brake for the driver automatically.

To further this objective, a number of field tests were conducted to collect data under realistic conditions. Among several others, a Safety Pilot Model Deployment (SPMD) was conducted by University of Michigan Transportation Research Institute. The Safety Pilot consisted of 2,800 cars, trucks, bicycles, buses and motorcycles instrumented with communication devices. There were 27 RSE’s, 21 of which were placed at signalized intersections, 3 at curves and 3 along freeways. The data consist of 28 field of basic safety messages (BSM). We use nine of these that correspond to vehicle dynamics: latitude, longitude, elevation, speed, heading, estimated longitudinal (front-to-back) acceleration, estimated lateral acceleration, estimated vertical acceleration and yaw rate. Other fields of interest beyond this study include steering wheel angle, brake system status, vehicle width, vehicle length, throttle position, state of exterior lights, and state of transmission. These fields were not populated in our data set, likely because the data were collected using after-market equipment installed on vehicles of volunteers. Each RSE has a unique identifier. When a vehicle comes into view of an RSE, it is assigned a unique, temporary identifier. The identifiers are randomized to prevent vehicle tracking. The data are broadcast at a nominal 10 hz, but in practice, the rate is variable owing to dropouts. The high sampling rate coupled with the low latency 5.9GHz communication channel is intended to provide ample communication, command and control to in crash-imminent situations given vehicle dynamics.

Data quality is often viewed as having several aspects. First is the incomplete, in that the data are missing or not measuring the system it is expected to describe. Under this would be errors like “stuck at zero” data, which is all too common coming from sensors. We see this and significant gaps in the V2I data. Second is the improbable, data values that are highly unlikely to occur, yet cannot be logically ruled out. Finally, the impossible – data values that are simply unrealistic, e.g. speeds in excess of 200 mph, GPS coordinates in the middle of a lake, and so on.

It is not sufficient to identify these errors – if possible they should be corrected. Missing data issues are often corrected through imputation; that is to infer based on historical or concomitant data. Our approach is to use the physical properties of a vehicle, specifically the continuity of position and velocity acceleration in combination with a data augmentation technique (expectation maximization) from statistics to reconstruct a plausible trajectory.

The goal of this work is to reconstruct historical or archival data, not real-time filtering or prediction. This gives us the freedom to use computationally expensive, non-deterministic algorithms to yield the best estimates. These in turn, could be used to create models of intersection and curve dynamics.

**2. ???**

**3. ???**

**4. ???**

**5. ???**

**REFERENCES**