Rethinking Connected Vehicles for Spectrum Scarcity

Data Collection

What data will you collect or create?

(A) This project will make use of existing large datasets of empirical data collected in the citywide trial of a vehicular network in Porto, Portugal.

(B) This project will make use of existing large datasets made available by the U.S. DOT that it obtains through technology trials and through simulations. This data sheds light on the effectiveness of vehicle-to-vehicle and vehicle-to-infrastructure applications for road safety, and on driver behavior with and without the technology.

(C) This project will make use of existing datasets that describe the locations of Wi-Fi hotspots gathered through crowd sourcing.

(D) This project will create new datasets that show performance metrics for connected vehicle systems and for wireless systems that share spectrum with connected vehicles.

How will the data be collected or created?

Dataset A on connected vehicles in Portugal was collected in the PI's previous research projects through a collaboration between Carnegie Mellon University and the University of Porto.

Dataset B on connected vehicles in the US is collected from U.S. Department of Transportation websites.

Dataset C on wi-fi is collected from FON, and possibly similar sets of crowd-sourced information about wi-fi deployments

Dataset D on connected vehicle performance is created through use of simulation software.

Documentation and Metadata

What documentation and metadata will accompany the data?

Datasets A, B and C were collected from external sources and/or already-completed studies, which establish their own metadata standards and produce their documentation.

For Dataset D, which we produce through simulations, we develop simple schemas, some of which define precisely how inputs to simulations are defined in "configuration files" that simulation software uses to obtain all model inputs, and "output files" that the simulation software uses to store the results of a simulation run. We produce documentation of both input and output files that is accessible to current participants in this project, and archived for future use.

Ethics and Legal Compliance

How will you manage any ethical issues?

We foresee no significant ethical issues with these datasets.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The dataset collected from Portugal contains proprietary data which we use for research purposes as input to our simulations, but we must protect raw data from third parties. Researchers with access to proprietary data in any form must agree not to share it, or store it anywhere beyond protected sites associated with this project.

Storage and Backup

How will the data be stored and backed up during the research?

All data generated as a result of this project will be stored on either a university server or a computer assigned specifically to one of the researchers. In either case, project data will be backed up to a remote server in encrypted form at least once per day to protect from loss of data from hardware failures, fire, theft, and other such catastrophic events.

How will you manage access and security?

We store all proprietary data and most non-proprietary data on secure remote servers provided by Carnegie Mellon University. Information on these servers is password-protected. We sometimes store especially large datasets in external hard drives with password protection, so that researchers can work with them offline while protecting security. Passwords will be available only to researchers, and more specifically only to those researchers with a need to know that particular information.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Data created in this project will be used to produce papers for publication. We do not currently expect these datasets to be of long-term value after papers are published.   Publications will be available from publishers, and on the PI's Carnegie Mellon University website to the extent possible given any copyright restrictions.

What is the long-term preservation plan for the dataset?

All data, software and documentation of possible long-term value is retained on servers provided by Carnegie Mellon University, in storage under control of the PI of this project. Other individual researchers on the project periodically transfer important data, software and documentation to the PI for preservation.

Data Sharing

How will you share the data?

We do not expect this project to produce new datasets that would be shared outside the research project.

Are any restrictions on data sharing required?

There are no restrictions on data created in this project. There are restrictions on data collected from past research in Portugal, i.e. Dataset A, as discussed above in the section on intellectual property.

Responsibilities and Resources

Who will be responsible for data management?

The PI will be responsible for data management.

What resources will you require to deliver your plan?

We require access to Carnegie Mellon University servers, and a small number of portable external hard drives.