Title
The Tracer Data Collection System: Implementation and Operational Experience

Authors
Marca, James E.
Rindt, Craig R.
McNally, Michael G.

Publication Date
2002-07-30

UCI-ITS-AS-WP-02-13

The Tracer Data Collection System: Implementation and Operational Experience

UCI-ITS-AS-WP-02-13

James E. Marca
Craig R. Rindt
Michael G. McNally

Department of Civil and Environmental Engineering and
Institute of Transportation Studies
University of California, Irvine
jmarca@uci.edu, crindt@uci.edu, mmcnally@uci.edu

July 2002

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

The Tracer data collection system: Implementation and operational experience

James E. Marca, Craig R. Rindt, and Michael G. McNally

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697
tel: (949) 824-6571
fax: (949) 824-8385
jmarca@translab.its.uci.edu

July 30, 2002

Abstract. The Tracer system consists of any number of mobile extensible data collection units (EDCU) communicating wirelessly in real time with a base station server. The mobile EDCUs store speed, longitude, latitude, and timestamp readings from a global positioning system (GPS) antenna. They also transmit these readings wirelessly, in real time, over a cellular digital packet data (CDPD) modem. A small pilot study with four vehicles was conducted to test the system. Three vehicles have collected over 100 days worth of data, and the one other vehicle has collected over 70 days worth of data. The operational experience gained from this wireless data collection implementation is documented here. The results demonstrate that long-term wireless data collection is feasible, but that the CDPD link's instability can add errors to the data stream.

Wordcount: 4036 + 3 figures

INTRODUCTION

This paper documents the development and implementation of a wireless data collection system, referred to as the Tracer system. Tracer was developed for a number of end applications. As the name suggests, the Tracer system traces the movement of equipped vehicles through space and time. The hardware component consists of the extensible data collection unit (EDCU) used to collect GPS data and transmit it wirelessly. The software component consists of programs to enable the collection, transmission, storage, processing and use of the GPS data. The implementation of a web-based activity diary using the Tracer position data is documented in another paper *(1)*. Other papers *(2, 3)* describe describe some preliminary analysis of the pilot data, using data analysis tools and techniques that have been integrated into the on-line survey web site.

Collecting GPS data for use in a travel analysis context certainly is not a new idea. The next section documents other approaches that have been reported in the literature. There are undoubtedly many more practical implementations, as this exciting technology has a great many researchers joining the fun. Among travel behavior researchers, our implementation, as described here, appears to be unique in its usage of real-time wireless data transmission technology. The description of the EDCU and the Tracer system includes the client and server programs that enable the actual data collection and storage. This paper concludes with some observations on using this system for extended time periods.

GPS DATA COLLECTION

The Tracer system arose from a concept of an *extensible data collection unit* that could be used for a variety of purposes. A major inspiration was the success of the Federal Highway Administration's Lexington Area Travel Data Collection Test *(4, 5)*. This project demonstrated some of the advantages of using a GPS device to collect data automatically. While GPS antennas have errors associated with them that must be handled appropriately *(6)*, these errors are largely mechanical and quantifiable. This is a tremendous improvement over the variable and intractable human errors related to personal reporting of travel times, routes, and destinations.

Most of the initial implementations of GPS data collection devices have focused on quantifying the errors arising with the use of this new technology, and establishing contexts for their use. For example, the Lexington study compared reported trip rates with GPS measurements, discovering among other things that while reported trip start and end times are rounded off to 5 or 15 minute intervals, the GPS records show an a uniform distribution of departure times *(5)*. Using the same data, Yalamanchili et al. *(7)* provide some descriptive statistics of the differences between GPS and recall data. They also focus extensively on evidence of trip-chaining detection and propensity. They observe that the households that participated in the GPS/handheld computer subsample did not enter trip purpose information for all of the trips that the GPS device recorded, with information provided primarily for "significant" trip purposes.

Wolf et al. *(6)* explore the use of a GPS device combined with a personal digital assistant (PDA) for collecting travel data in the Atlanta metropolitan area. They make several recommendations for reducing and managing the errors related to the use of such a device, with a focus on the integration of GPS points into a geographic information system. Later work by these authors explored using the GPS device as the exclusive means of collecting travel and activity data and doing away with the traditional travel survey in favor of a shorter follow-up telephone interview

(8). Rather than relying on the travelers as the primary explanatory source for the measured GPS points, the approach relies upon geographic information system (GIS) data to provide a first cut at establishing trip purposes, routes and destination. The authors compare traditional travel data collected by respondents in a small feasibility study, to data that can be inferred from entering the GPS measurements into a GIS. They claim that the comparison shows that the purely mechanical approach is often superior to relying on respondent data, primarily because of the human errors and underreporting of the respondents.

A large-scale application of GPS data collection is taking place in the city of Atlanta, in association with the SMARTRAQ project at the Georgia Institute of Technology which produced the above research. This travel and activity survey includes a subsample of approximately 500 individuals who are recording their travel using a PDA with a GPS antenna. Unfortunately, the survey is ongoing as of this writing, and so no reports have been published documenting its methods or results. This project is one of the largest applications to date, and its results should be very informative.

Bachu et al. *(9)* also test relying exclusively upon the GPS unit to collect travel data. Their method uses a commercial, battery-powered GPS recording device to collect travel data for several days. Their data collection instrument does not include a PDA, as they deliberately wanted to avoid the battery-life penalties as well as the potential user interface biases. The GPS recorders are collected from the respondents after a few days, and the data are entered into a GIS to generate printed maps. These maps are then shown to the respondents in a face to face follow-up survey to verify the accuracy of the GPS tracking, and to prompt the respondents for information that cannot be recorded by the GPS devices. The authors explain that their methods try to reduce the reliance upon hand-held computers, devices whose very newness is perhaps a barrier to use by survey respondents.

Draijer et al. *(10)* examine the use of GPS records in multiple travel mode studies, not just in personal vehicles. For this purpose the designed a data logging platform called GeoMate, weighing approximately 2 kg and integrating a battery, an embedded computer, and a GPS antenna into a standard video-camera shoulder bag. The GPS antenna was attached to the shoulder strap of the camera bag to afford it a better view of the sky. While the GeoMates were reportedly somewhat temperamental, losing data completely from 22 of the 151 respondents, in general the device was able to be used by respondents on any type of travel mode, including walking, bicycling, public transit, and car. The study focused both on the technical ability to use the GeoMate on different travel modes, as well as on the willingness of respondents to use the device. On the technical side, respondents indicated that the device was too bulky or otherwise inconvenient for non-motorized travel modes, and also often inconvenient for trip purposes such as shopping and visits. In terms of the willingness of the respondents to use the device, it was found that privacy was an important concern, but did not significantly decrease the response rate.

At the time of the design of the EDCU, we were aware of the Lexington study, and of encouraging reports of the benefits of GPS measurements. We drew up a specification and RFP for a small embedded system, with requirements that it fill a variety of different functions *(11)*. The primary goal was to collect and save GPS trip data, which would allow us to analyze traffic conditions and travel behavior. The secondary goals outlined in the RFP were to transmit data in real-time, and integrate the collected data with more detailed survey tools. This functionality would enable dynamic travel behavior studies which used GPS data both as a memory jogger, and as a means to tailor specific questions. The real-time transmission capability was also vital to the use of this device in a traffic probe function, which was important to the needs of the funding project. Interestingly, no other GPS devices used for travel behavior research seem to be leveraging wireless technology for real-time transmission of the collected data, although there are several examples in fleet management and freight operations. Finally, the specification noted that the ideal implementation should be easily extended, both in hardware and software, so that the device could be outfitted with newer technology rather than becoming obsolete in six months. The actual implementation of this specification is discussed in the next section.

THE EXTENSIBLE DATA COLLECTION UNIT AND SUPPORTING SOFTWARE

Aero Data *(12)* was selected to build the EDCU. They designed and delivered a unit, shown in figure 1, which consists of

- an external GPS antenna;
- 16MB CompactFlash removable storage (easily upgradable);
- 32MB RAM, providing a 16MB ramdisk for the runtime filesystem;
- an internal CDPD wireless data modem with an external antenna;
- an uninterruptible power supply (UPS);
- a standard ethernet connection; and
- PC104 board risers, plus space for 2–3 expansion boards.

[FIGURE 1 about here.]

The EDCU does not have a built-in user interface at this time. This was a deliberate design choice, both to reduce the cost of the device, and because the technology of handheld computers is rapidly evolving. Indeed, at the present time we are exploring a future user interface expansion exploiting local area wireless technology such as 802.11b, which has become common and inexpensive only recently. These kinds of expansions can be done thanks to the unit's expansion board slots as well as its open operating system.

The EDCU has been designed to be used in an automobile. The power is supplied by the 12V cigarette lighter that is standard equipment in cars, or by the so-called utility port which is beginning to replace the cigarette lighter. The difference seems to be that the cigarette lighter is always powered by the car battery, whereas the utility port switches off when the car is turned off. To accommodate use in a car in which the power is always on, the EDCU has a motion sensor switch, and will shut itself down after 30 minutes of inactivity. To accommodate vehicles which kill power when the car is shut off, the EDCU has an uninterruptible power supply (UPS), using a small rechargeable battery to provide power to the unit as it shuts itself down gracefully. In practice, neither solution was foolproof, but they generally worked as intended.

The EDCU software runs on the Linux operating system—specifically an older version of the Linux Embedded Appliance Firewall (LEAF) distribution *(13)*. On top of the operating system, the unit runs an HTTP (web) server, and telnet and FTP servers. The web server and FTP capabilities enable remote control of the EDCU units, in case it should be necessary to change their behavior in the field, as well as the ability to download specific log files. Since Linux is a network aware operating system, it was straightforward to develop wireless client and server programs for the EDCU. These programs automatically log data in a central database while the EDCU is active.

The EDCU can collect and save GPS data in multiple ways, as described in figure 2. An internal GPS monitor listens to the signals from the GPS antenna, and saves the data to a files in a directory on the CompactFlash memory card. This memory card is the primary repository of collected data. The GPS data can be "hand carried" to the database by plugging this CompactFlash memory card into a card reader attached to a laptop or desktop computer. If the cellular modem is up and working, then data are transmitted automatically while the vehicle is moving via the Transmit server, programmed by the author. This server watches the log file for new data. Every few seconds, it transmits the latest set of GPS records to the Base Station, using XML-RPC *(14, 15)* calls over the standard HTTP Internet protocol. Log files can be downloaded via the HTTP and the ftp servers. One can also login directly via the telnet server, which is useful both for developing applications as well as for changing parameters in the field. In practice, the automatic real-time wireless connection via the Transmit server has been used almost exclusively, due to its ease of use.

[FIGURE 2 about here.]

The base station operations are filled by web and database servers running at UCI. The functions filled by the base station computers include receiving and logging real-time data to a database, and providing a web interface server. The web interface allows analysts and/or drivers to get information on deployed EDCUs, and view a sketch graphic of the data collected by each unit. These operations are beyond the scope of this paper.

OPERATIONAL EXPERIENCE

The primary advantages of the EDCU, as built and delivered, are its openness and its flexibility. While the basic function of collecting GPS data will likely remain the same, we can add functionality to the unit to suit our needs. The Linux operating system enabled the easy implementation of network protocols, such as the XML-RPC client and server programs that automatically log data. We were also able to setup the EDCU to serve as a firewall and router for any computers (such as a laptop) connected to its ethernet port. This allows these computers to get wireless Internet access via the EDCU's cellular data connection.

The unique feature of the Tracer system, compared to other uses of GPS units in travel behavior research, is the use of wireless data technology to store travel data to a database as soon as it is recorded by the mobile units. This eliminates the need to physically interact with the units in the field to collect data, which in turn makes long, open-ended surveys possible. A small pilot study with the Tracer system has successfully tracked three EDCUs for over 100 days each without ever directly accessing the units to collect data. The ability to collect data for long durations without the need for analyst intervention (and the associated time and monetary costs) is an important step towards being able to field a full-scale survey designed around a long period of GPS data collection.

The most surprising feature of the EDCU, which may turn out to be its most important contribution to transportation engineering, is that this small, embedded device is a wireless web *server*. While the content is rather limited (collected GPS positions, speeds, and times), the unit is serving original content to the broader Internet. The usual paradigm for automotive telematics is that the wireless device in the vehicle is a client for some centralized product, be it advertising or route guidance information. But in our case, the EDCU almost exclusively performs as a server, not a client.

The Tracer system has been installed without problems in several different kinds of vehicles. The only requirement is that the vehicle have a working cigarette lighter to provide the unit power. If the cigarette lighter does not shut down when the vehicle ignition is turned to off, the EDCU should automatically shut itself down after approximately 30 minutes. According to the EDCU vendor, with normal lead acid batteries the EDCU should never cause a battery to drain within that 30 minute period.

A small pilot study consisting of four volunteers well known to the authors used the Tracer system over an extended period of time in order to gather operational experience. In most cases the devices performed as designed. However, an unexpected problem arose when one of the volunteers placed the CDPD antenna inside of the car, rather than on the roof as recommended. This caused the CDPD modem to periodically lose its connection, which was a problem for the following two reasons.

First, a hiccup in the CDPD modem would cause the data transmission program to block. The operational logic of the wireless EDCU data collection process is shown in figure 3. There are two sources of errors with the way this process is currently implemented. Referring to the numbered messages in figure 3, the transmit server sends message 2 and then waits to receive message 6. If the CDPD link fails for message 2 or 6, then message 6 will not be received by the transmit server, and it will block indefinitely. In practice, the transmit server is restarted every 5 minutes, so the effect of a CDPD transmission drop is to cause a 5 minute gap in the collected data.

[FIGURE 3 about here.]

The second effect of the spotty CDPD connection was related to the detection of stops. The base station has no way of knowing a priori whether or not a mobile unit is powered up, collecting data, but out of transmission range, or turned off. Rather than having the base station continuously probe all of the mobile units (which is only possible if the base station knows the IP addresses of the units), the mobile units were instead programmed to send a "shut down" message to the base station when the units turned off. This strategy turned out to be unsuccessful, as the battery backup uninterruptible power supply (UPS) often did not have enough power to transmit that message. Another problem with this strategy was that about half of the various test vehicles did not power down their cigarette lighters when the ignition was switched off. Thus the EDCUs did not power down at all, but instead continued to transmit stationary points. The base station was reprogrammed to attempt to automatically detect a stop. The algorithm was only partially successful, with a bias towards adding too many stops rather than too few. When confronted with CDPD hiccups and the resulting block, the net result was for the base station logic to write out a potential stop to the database. This caused "stops" to be listed at random points along the volunteer's daily commute, wherever the CDPD modem lost radio contact.

The cumulative result of these two sources of error is that the collection of destinations is much larger than the actual number of destinations. Further, due to the gaps in transmission, sections of travel between stops are often missing. Finally, if the CDPD connection is weak and/or a trip takes less than a few minutes, the system may entirely miss a trip and a stop.

All of these problems have fairly easy solutions. The solution to the transmit server problem is to fork the process with each transmission, so that the parent continues to monitor the log file and spawn new transmission processes. A hiccup in the wireless link will stall only the child process, not the parent or the other transmit processes. After a certain time, the persistent child processes can be rounded up by the transmit server dæmon into a new process that will combine and resend their messages. This solution has not been implemented yet on the EDCU, but a similar forking scheme was devised for the base station XML-RPC dæmon to allow multiple simultaneous EDCU connections.

The algorithm to detect stops on the server side has been improved, as part of the effort to analyze the data. The new algorithm checks for time periods of travel data that have zero speed. The current algorithm also attempted to do this, but the implementation seemed to fail for technical reasons relating to how the persistent data was being stored in the database. Second, the new algorithm examines gaps in travel data for spatial continuity before inserting a stop. If the power down and power up points are separated by an appreciable distance, and if the time difference implies a reasonable speed between those points, then it is assumed that the vehicle has not stopped, but rather simply had a transmission gap for some reason. When the data transmission server is improved on the EDCUs, the need for this check should diminish.

CONCLUSIONS

The Tracer data collection system was developed to monitor vehicles continuously, with the intention of tracking both real time travel conditions, as well as monitoring activity destinations. This paper describes the implementation of the Tracer system, focusing on the flexibility provided by its wireless data connection. The system has been in continuous test in four vehicles, with more than 100 days of data collected for three of the vehicles. Aside from some small problems with CDPD connectivity when the antenna was placed inside the vehicle, the system has no serious design flaws.

Then next step forward is to design a formal evaluation of the consistency of the system. The small pilot study was not successful at having the volunteers review and comment on each recorded destination. Significant error rates (incorrectly recorded stops) were only noticed in one case, although we expect that incorrect stops were recorded for the other three volunteers as well. In order to evaluate the appropriateness of this technology for different applications, we should establish the expected error rates due to cellular coverage gaps, and work to improve the data collection programs to access the stored GPS records to counteract these gaps.

REFERENCES

[1] James E. Marca. The design and implementation of an on-line activity survey. Working paper, May 2002.

[2] James E. Marca. A preliminary model of activity destinations as random point processes. Working paper, May 2002.

[3] James E. Marca. Collecting activity data from gps readings. Working paper, May 2002.

[4] Battelle. Lexington area travel data collection test. Technical report, Federal Highway Administration, 1997.

[5] Elaine Murakami and D. P. Wagner. Can using global positioning system (GPS) improve trip reporting? *Transportation Research C*, 7(2/3):149–165, April/June 1999.

[6] Jean Wolf, Shauna Hallmark, Marcelo Oliveira, Randall Guensler, and Wayne Sarasua. Accuracy issues with route choice data collection by using global positioning system. *Transportation Research Record*, (1660):66–74, 1999.

[7] Lalit Yalamanchili, Ram M. Pendyala, N. Prabaharan, and Pramodh Chakravarthy. Analysis of global positioning system-based data collection methods for capturing multistop trip-chaining behavior. *Transportation Research Record*, (1660):58–65, 1999.

[8] Jean Wolf, Randall Guensler, and William Bachman. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record*, (1768):125–134, 2001.

[9] Prashanth K. Bachu, Trisha Dudala, and Sirisha M. Kothuri. Prompted recall in global positioning system survey: Proof-of-concept study. *Transportation Research Record*, (1768): 106–113, 2001.

[10] Geert Draijer, Nelly Kalfs, and Jan Perdok. Global positioning system as data collection method for travel research. *Transportation Research Record*, (1719):147–153, 2000.

[11] Michael G. M^c^Nally. GPS/GIS technoligies for traffic surveillance and management: A testbed implementation study. Submission for PATH RFP FY 1999-2000, February 1999.

[12] Aero Data. Aero Data home page. published on the Internet, 2002. URL `http://www.aerodata.net`.

[13] Linux Embedded Appliance Firewall (LEAF). Linux Embedded Appliance Firewall (LEAF), 2002. URL `http://leaf.sourceforge.net`. This is the successor to the Linux Router Project.

[14] Dave Warner. XML-RPC: It works both ways. davewarner98@yahoo.com, Jan 2001. URL `http://www.oreillynet.com/pub/a/python/2001/01/17/xmlrpcserver.html`.

[15] UserLand Software. XML-RPC home page, 2001. URL `http://www.xmlrpc.com/`.

LIST OF FIGURES

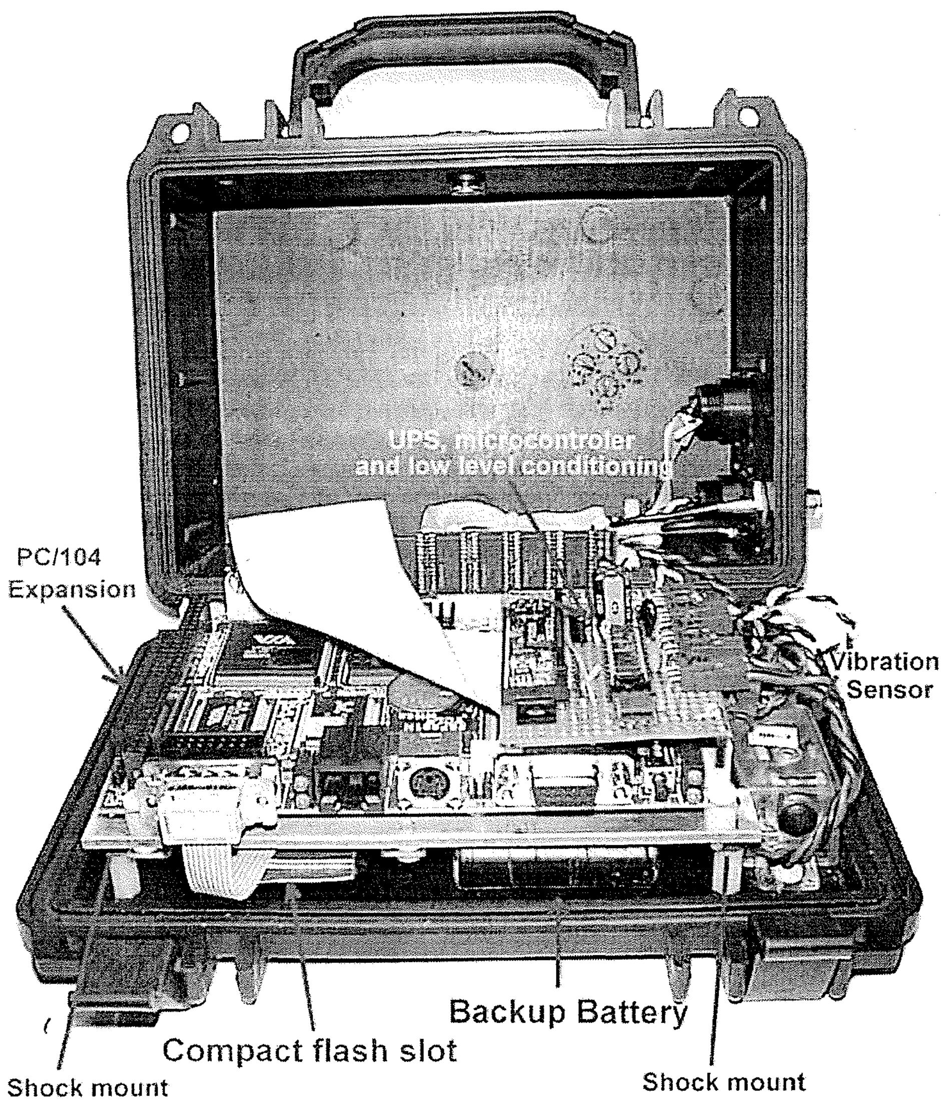

FIGURE 1: The Extensible Data Collection Unit (EDCU)

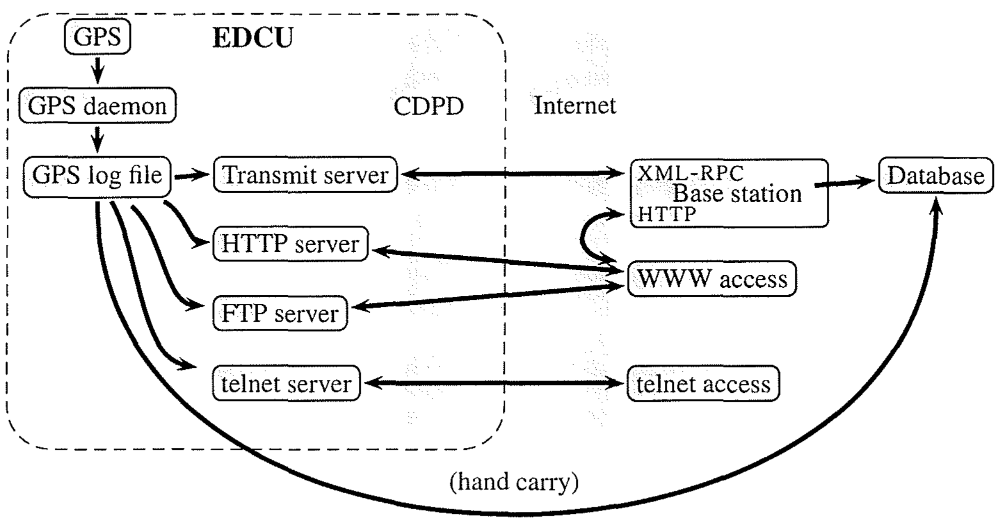

FIGURE 2: Flow of information from the GPS antenna to the database. Data are transmitted automatically while the vehicle is moving via the Transmit server. Alternately GPS log files can be downloaded via the HTTP and FTP servers. One can also login directly via the telnet server. Finally, the GPS data can be "hand carried" to the database by removing the CompactFlash memory card.

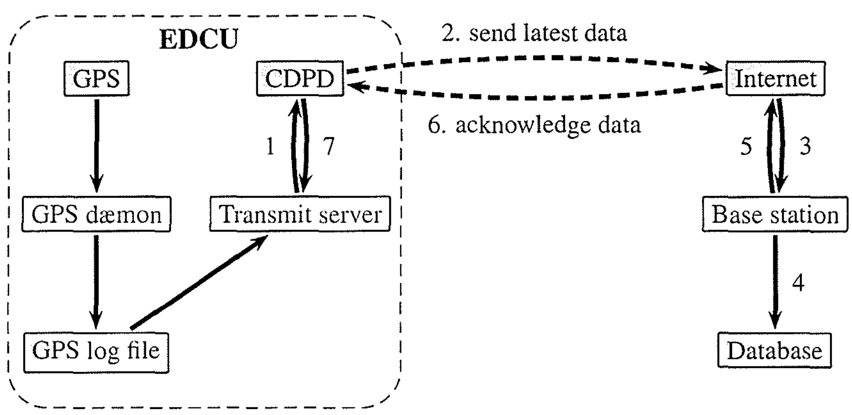

FIGURE 3: Flow of information from the GPS to the database via the transmit server. If the CDPD link fails to deliver message 2 or message 6, then the transmit server does not receive acknowledgment that the data has been stored to the base station database.

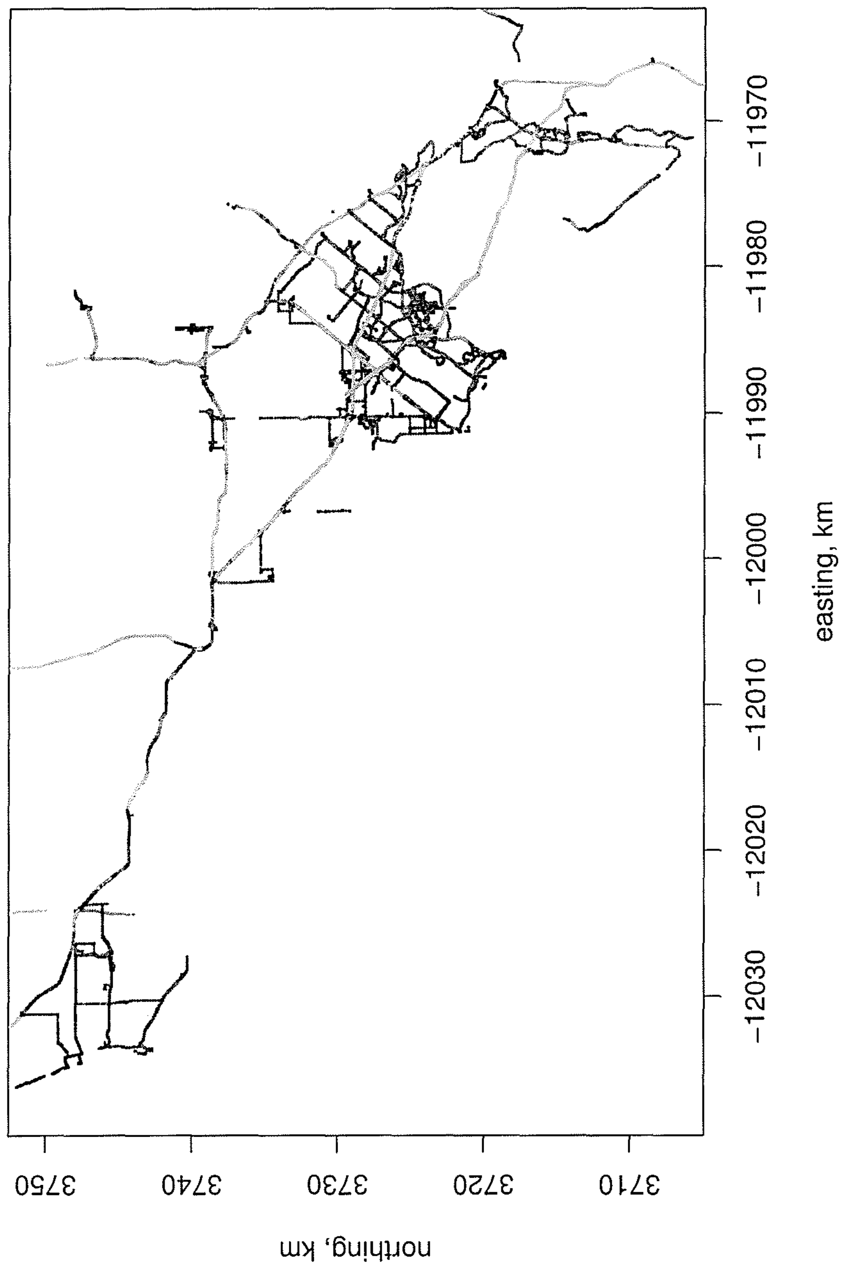

FIGURE 3: All vehicles' observed average travel speeds, within user id 6's travel area. The dark points represent slower average speeds while the lightest points are average speeds greater than 70 mph.