UC Irvine
Recent Work

Title

A GPS Enhanced In-Vehicle Extensible Data Collection Unit

Permalink

https://escholarship.org/uc/item/3867z3s9

Authors

Marca, James E.
Rindt, Craig R.
McNally, Michael G.
et al.

Publication Date

2000-12-01

A GPS Enhanced In-Vehicle
Extensible Data Collection Unit

James E. Marca [1]
Craig R. Rindt [1]
Michael G. McNally [1]
Sean Doherty [2]

[1] Department of Civil & Environmental Engineering and Institute of Transportation Studies
University of California, Irvine; Irvine, CA 92697-3600, U.S.A.
jmarca@uci.edu
crindt@uci.edu
mmcnally@uci.edu

[2] Department of Geography & Environmental Studies
Wilfred Laurier University; Waterloo, Ontario, Canada N2L 3C5
sdoherty@wlu.ca

December 2000

Paper No. 01-3453
A GPS Enhanced In-Vehicle Extensible Data Collection Unit

Duplication for publication or sale is strictly prohibited without prior written permission of the Transportation Research Board

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

Sean Doherty †

November 21, 2000

Abstract

The rapid advancement of technology has created the opportunity for applying new, powerful tools to transportation engineering problems, but often the very speed of technological change hinders the adoption of these tools in a research environment. This paper documents the development of an extensible data collection unit (EDCU). The unit combines a standard GPS unit, a cellular data modem, and an embedded processor running the Linux operating system. Some preliminary uses and applications of the EDCU are presented as well. The EDCU satisfies multiple functional requirements, due to the flexibility of its modular components and its full-powered operating system. The EDCU will serve the in-vehicle data collection needs of travel demand modelers and ITS researchers for the foreseeable future.

*Corresponding author. Institute of Transportation Studies, University of California, Irvine, CA 92697; 949-824-8462 mmcnally@uci.edu

†Dept. of Geography & Environmental Studies, Wilfrid Laurier University, Waterloo, Ontario Canada N2L 3C5

1

1   Introduction

Despite recent advances in Intelligent Transportation System (ITS) technologies, the potential for improving transportation system performance is curtailed by our limited understanding of the relationships between transportation system performance and characteristics of travel demand. While our models of supply tell us what we do under various traffic scenarios, there is little that we may confidently implement due to a fundamental lack of understanding of individual travel demand and route utilization. Any attempt to implement traffic management strategies requires assuming a demand pattern, one that we cannot subsequently modify based on the implemented modification to the supply. Even the real-time application of simple advanced transportation management and information systems (ATMIS) strategies such as modification of signal timings is severely limited by the inability to generate real-time origin/destination (OD) matrices, let alone demand matrices that reflect modifications directly resulting from the implemented strategies.

The difficulty in obtaining reliable and accurate travel behavior data is one of the primary reasons individual travel behavior and route selection is so poorly understood. Travel behavior studies commonly utilize self-report measures, stated-preference surveys, or laboratory simulation to gather data but rarely involve direct measurement of actual individual behavior. When individuals omit aspects of their behavior in surveys or are limited in their behavioral options during a simulation, significant discrepancies can arise between actual behavior and "measured" behavior. These shortcomings introduce significant doubt about the reliability of data collected for model development. As we move forward with the marriage of transportation and information technology, the need to understand the motivations of the individual traveler will become even stronger. The new traveler information technologies open up a new dimension of uncertainty, raising questions about market penetration and user adoption.

To begin to address these shortcomings, we have attempted to merge global positioning system (GPS) technology with in-vehicle data collection in transportation research and operations applications. GPS-enhanced vehicle-based data will provide direct, objective information about traveler behavior and the state of the transportation system provoking this behavior. The goals of this wider effort are as follows.

1. Develop a flexible data recording platform supporting a variety of in-vehicle data recording applications.

2. Implement and evaluate a core set of data recording applications including:

    (a) GPS tracking of vehicle traffic profiles;

    (b) real-time transmission of vehicle position and performance variables; and

    (c) interactive, in-vehicle driver surveys.

3. Investigate use of the technology in such traffic and travel behavior studies as:

    (a) the use of probe vehicles to complement real-time volume/occupancy data collected for network links in the California ATMIS Testbed;

    (b) to investigate route selection behavior in households in response to varying levels of congestion; and

    (c) to increase the level of understanding of trip-making behavior relating to the generation and scheduling of household trips as a function of anticipated travel time and other route information.

This paper discusses in-vehicle data collection for both traffic operations and travel behavior. First, section 2 provides a brief overview of GPS data collection research. In section 3 we provide a synopsis of two ongoing projects that require advanced in-vehicle data. These projects served as our preliminary target applications for design and construction of the extensible data collection unit. With these two projects serving as a guide, we define a set of core applications in section 4 that span the requirements for current and future in-vehicle data collection needs. We then detail in section 5 the design of the extensible data collection unit (EDCU) that is capable of supporting the spectrum of foreseeable data collection needs. Finally, section 6 documents our initial experience using the EDCU.

2 Related work

2.1 The use of GPS in traffic studies

Interest in use of probe vehicles to provide data substituting for or complementary to conventional data sources in transportation networks dates back several years (Westerman, Litjens and Linnartz, 1996; Quiroga, 1997). Standard engineering practice has used measurement devices placed in the medium under observation, despite some interference with that medium. For traffic studies, the marginal increase in volume and other traffic stream interference due to the presence of the probe vehicle was less a constraint than the practicality and costs of deploying probes. A variety of probe surveillance studies have been undertaken using a variety of technologies, including GPS, roadside automatic vehicle identification (AVI), and cellular technology. Quiroga has utilized GPS in a variety of traffic engineering studies with a probe vehicle, and have addressed link speed calculation as well as interoperability with GIS for managing data generated by the GPS (Quiroga and Bullock, 1998; Quiroga, 1997) Barth has used GPS to estimate traffic

stream parameters for emissions assessment (Barth and Johnston, 1995; Young-blood, Johnston and Barth, 1995). In addition to issues relating to the required level of saturation of probe vehicles, there are also issues relating to data fusion (often with conventional loop data) and interoperability (fitting GPS data with GIS-based maps and resolving coordinate systems).

2.2 The use of GPS in travel surveys

The application of global positioning systems (GPS) to the study of individual travel behavior is a relatively recent phenomenon. Applications are divided between "hands-off," passive devices and devices that require some level of user interaction. In the U. S., a completely passive vehicle-based GPS recording device has been developed at the Texas Transportation Institute and applied to 200 households who also completed a standard paper-and-pencil activity diary (Pearson, 1999). Preliminary analysis suggests that the GPS data results in the capturing of more trips than the diary, especially short trips, although more analysis is still to be conducted.

The Lexington area study in the U. S. (Murakami and Wagner, 1999) was one of the first to demonstrate how passively collected GPS data could be combined with hand-held computer to interactively obtain supplemental information trip purposes. Respondents were instructed to turn the unit on within their vehicle at the start of a trip, then specify the trip purposes of each person in the vehicle via prompts on the hand-held computer. These trip purpose, trip start and end times, and origin/destination addresses were then tagged to the resulting GPS point data recorded at 3-second intervals. In this way, the device was able to replicate the same type of information collected via a traditional trip diary. Subsequent analysis of 100 households focused on the improvements in the accuracy of recorded trip start and end time, travel distance calculated on a point-to-point basis. This work also highlighted many problems associated with such units, including loss of position fix and user-induced errors of omission.

As an alternative to using point-to-point distance estimates, Roden (1999) demonstrated a map matching algorithm that could match second-by-second GPS points to a road network derived from Tiger files in the U. S. The algorithm consisted of three mains steps: a smoothing procedure that first removed anomalies in the raw GPS data, followed by a procedure to identify "corners" taken en route, followed by the actual matching of the points to the network and identification of links that make up the traced route. A sample of 300 trips over 58 days from one sample vehicle was used as a basis to demonstrate how a variety of travel distances and times measures could be derived with this approach.

Similar experimentation in Quebec City, using differentially-corrected GPS sample data has shown that even the simplest of GIS-based algorithms that match GPS points to the nearest link using a "buffering" technique could lead to over

90% accuracy without further processing (Doherty, 1999; Ueno, Noîl, Doherty, Lee-Gosselin, Theberge and Sirois, 1999). An algorithm was also developed to identify trip-ends, or "activity" nodes, based on the clustering of GPS points in time and space. All stops of length greater than a given threshold were easily detected by the algorithm. However, it was recognized that short stops, such as when dropping off passengers, would be more difficult to detect automatically, and would likely require some form of user input.

One new GPS project that will soon make a large impact in this field is the work of Wolf et al. at the Georgia Institute of Technology (Wolf, Guensler, Washington and Frank, 1999). This team of research has come up with a multi-instrumented approach that builds upon and extends past approaches. Three separate surveys are planned for the near future, including

1. a person-based hand held electronic travel diary with GPS receiver;

2. a vehicle-based electronic diary with GPS and on board engine monitoring system;

3. a passive GPS receiver and data logger to capture vehicle trips only.

The goal of these surveys is to replicate traditional activity-based diaries (albeit, more accurately), with the addition highly detailed route and link information, vehicle and engine operating conditions affecting emissions, and freeway traffic conditions derived from real-time advanced traffic monitoring system linked to the GPS data. This data is intended to support the development enhanced travel demand models capable of addressing emerging land use, travel behavior, and air quality issues.

These studies have demonstrated that GPS technology has the potential to provide a passive trace of vehicle movements over long periods of time, but can also be combined with more interactive techniques to provide a more comprehensive picture of travel behavior. Post-processing of the data can also be used to identify the links associated with a given route, and the start and end times and locations of activities. These studies have also highlighted several technical issues that require further attention:

- how to keep equipment package as small and durable as possible;

- how to maintain a consistent power supply;

- dealing with the large amounts of incoming data;

- dealing with GPS signal outages; and

- accuracy of GPS data and use of differential correction.

These considerations were all incorporated into the design constraints of the data collection unit described in this paper.

3   Two representative applications

3.1   REACT!

The REACT! survey program provides an efficient means to collect detailed information on observed activity-travel patterns within a household over a week long period. It does so by providing a graphical interface depicting a person's schedule on a computer (similar to a typical day-timer) on which a person may interactively add, modify, and delete activities as they get planned and subsequently executed. A REACT! survey participant logs in at least once a day to keep their schedule up to date over a week long period. REACT! is an Internet-based survey tool, incorporating features such as nightly uploads of survey data, to complete operation of the survey over the Internet when the connection bandwidth is sufficient. REACT! incorporates in its design a spatial GIS interface for interactive location choice (Lee, Sabetiashraf, Doherty, Rindt and McNally, 2000; Lee, Doherty, Sabetiashraf and McNally, 2000).

The next step for REACT! will be to combine its operation with vehicle-based travel data. The GPS data collection device will provide accurate traces of routes and travel choices of survey participants. Since the REACT! program is built around Internet technologies, it should be possible to transmit this travel information from the GPS data collection device over a wireless modem to a base station computer, process that data, and then send it back over the Internet to the REACT! rrespondent's computer. Even without a wireless data connection, in-vehicle travel data could be used to either initialize REACT! and its questions for the respondent, or else to design follow up questions or validate REACT! responses.

3.2   TRICEPS

Testbed Real-time Integrated Control and Evaluation Prototype System (TRICEPS) is a development platform that facilitates the implementation and evaluation of a wide range of algorithms for traffic control and Advanced Transportation Management Systems (ATMS) using both simulated and real world data (McNally, Rindt and Logi, 1999). It is a component of the California ATMS Testbed, a multi-agency transportation operations environment covering two contiguous sub-areas in Orange County, California that include major decision points for freeway travelers in the region as well as a significant portion of the surface street network. TRICEPS supports research activities in the Testbed by providing:

- consistent interfaces for transportation management hardware and software components,

- both simulated and real-world implementations of these components, and

- a set of core transportation management applications, that include automatic incident detection (AID) algorithms based on artificial neural network technology and algorithms for integrated traffic control.

TRICEPS provides a virtual interface to real-world transportation system data collection components. This interface is implemented using the Common Object Request Broker Architecture (CORBA) to manage the communication link with the California Department of Transportation (Caltrans) District 12 (Orange County) data server. This link provides access to data from all data collection and control hardware in the system including loop detectors, ramp meters, Changeable Message Signs (CMS), and video camera data.

A GPS-enhanced vehicle-based data collection device will enable the creation of *probe vehicles*. A probe vehicle travels in the traffic stream with other vehicles, and transmits highly accurate data about the conditions faced by moving vehicles. The addition of probe vehicle data to the TRICEPS platform will provide a unique source of data to transportation management applications operating within TRICEPS. Furthermore, TRICEPS provides real-time loop data from the Orange County freeway system and from arterials in selected cities in the Testbed. This data can be processed along with the GPS-based trajectories to evaluate travel times and congestion levels on the primary traffic routes as well as on alternate routes.

4   Application requirements

While the potential uses of GPS-enhanced vehicle-based data collection discussed in the previous sections vary significantly, we have defined a set of generic applications that span functional requirements for the various uses. We currently envision four data collection applications:

1. basic multi-day survey of vehicle trajectories,

2. obtaining real-time traffic stream conditions via probe vehicle,

3. enhanced multi-day survey of vehicle trajectories with behavioral logging, and

4. routing behavior under real-time route guidance

The most crucial of these for the REACT! and TRICEPS research applications discussed in section 3 are applications 1, 2, and 3. Application 4 is not required for current research, but we anticipate the need for such functionality in the near future. The following sections detail the parameters for each of these data collection applications.

4.1 Basic survey

In the *basic survey* application, the data collection system will record a vehicle's trajectory for later analysis of route utilization. One application of this project will compare *basic survey* data with traffic flow data collected from other sensors in the California ATMIS Testbed (e.g., loop detectors). Parameters for this application include:

- The system must collect sufficient data to enable matching of vehicle trajectories to a transportation network. Such matching may require positional fixes taken as frequent as 1-sec intervals at differential GPS accuracy.

- Drivers will average a maximum of 4-hours per day over a multi-day period which requires storage of positional fixes for a maximum of 28-hours total for a week period.

- The survey participant will have minimal access to the data collection device during the multi-day data collection period. The device should therefore be able to operate without maintenance over this period.

Post-analysis of collected data will vary with different research projects. This requires data be accessible in a flexible format. For example, the *basic survey* data may be compared with traffic flow data collected from other sensors in TRI-CEPS (e.g., loop detectors), or may be converted into map traces for a post-travel interview.

4.2 Probe vehicle study

In the *probe vehicle study* application, the data collection unit will collect data from a vehicle traveling in the traffic stream. The data will be logged in the on-board unit, and transmitted in near real-time to a base station over a wireless communications network. Parameters for this application include:

- The studies will take place in urban and suburban settings.

- Data collected for logging and remote transmission will include the position and speed of the vehicle, point-to-point travel times, and other, as yet unspecified vehicle data, such as acceleration profiles for pollution studies.

- Collection of some types of data may require *transmission* of data via a wireless communications link continuously, even when the vehicle is stopped.

4.3   Enhanced survey

The *enhanced survey* application extends the *basic survey* to include additional logs of respondent activity during the data collection process. For instance, log data might be the results of questions asked of drivers about the purpose, characteristics, expectations, etc. of a trip before or after it is made. Our prototype solution will extend the data collection device designed for the basic survey by connecting a user interface for gathering the additional information. Parameters for this application include:

- The data collection device will be isolated from the survey participants as with the *basic survey*, except for the user interface for interactive survey elements.

- The user interface must be compact enough to not interfere with normal use of the vehicle.

- It is likely that some survey respondents will *not* have extensive experience using computing devices. As a result, the input device for survey data collection must be implemented on a platform allowing the development of user-friendly interfaces.

There are many issues surrounding the inclusion of user interfaces within the vehicular environment. A major concern is the safety of the driver and passengers. Until there is a large body of experience associated with in-vehicle devices, it is unlikely that governmental agencies will allow such devices to be used in a survey. While not necessary for any current research, the design of our prototype unit can be easily extended to incorporate an input device and a small display, so that questions may be asked before and after trips.

4.4   Route guidance study

In the *route guidance study* drivers are provided with descriptive information about network conditions and/or prescriptive guidance about the best routes to take via real-time wireless communication with the vehicle. In addition to the capabilities given for the *probe vehicle* application described in section 4.2, this application requires:

- Two-way communication between a base station and the vehicle for the transmission of descriptive and/or prescriptive guidance from the base station to the vehicle and positional fixes of the vehicle back to the base station.

- A user interface, such as in the *enhanced survey*, that is flexible enough to display route guidance information to the driver, such as current position, desired destination, and suggested routing instructions.

The *route guidance study* application will ultimately require substantial software in the base station to drive the route guidance application. The system must provide support for remote wireless communication and a software Application Programming Interface (API) for two-way data transmission. These features are also required for the *probe vehicle study*. Once again, the user interface is a difficult design element, and one which will be put off until more experience has been gained in this area.

5 Data collection system design

We have designed and procured a prototype extensible data collection unit (EDCU) meeting the requirements outlined in section 4. Thirty in-vehicle EDCUs are available for evaluation and research. These self-contained units are built around a power-efficient x686-class, 133MHz microprocessor running a Linux-based embedded operating system. The unit uses flash-RAM as its primary storage, and controls a GPS receiver, and a cellular packet data (CDPD) modem. The operating system runs programs to control the various EDCU applications outlined in section 3. Figure 1 shows the prototype unit, the GPS antenna, and the CDPD antenna. The diskette gives an idea of the size of the box. Figure 2 shows the inside of the unit, with the various input and output ports, and the flash ram memory card.

Under normal operation, the units are designed to operate by tapping a vehicle's power supply via the cigarette lighter. With the addition of a battery pack, they can also be isolated from the vehicle's power for short duration applications. The units are enclosed in a weather resistant case that is suitable for the relatively harsh environment found in an automobile.

Because the equipment is intended to fill multiple data collection roles with significantly different requirements, versatility is a primary design goal. The data collection system therefore employs a modular design to ensure such flexibility. The device consists of the following modules:

- data collection unit controller (DUC),

- global positioning system module (GPS),

- on-board logging module (LOG),

- communications module (COMM),

- user interface module (UI), and

- base station unit (BSU).

Figure 1: Extensible data collection unit, complete with GPS antenna and cellular data modem antenna.
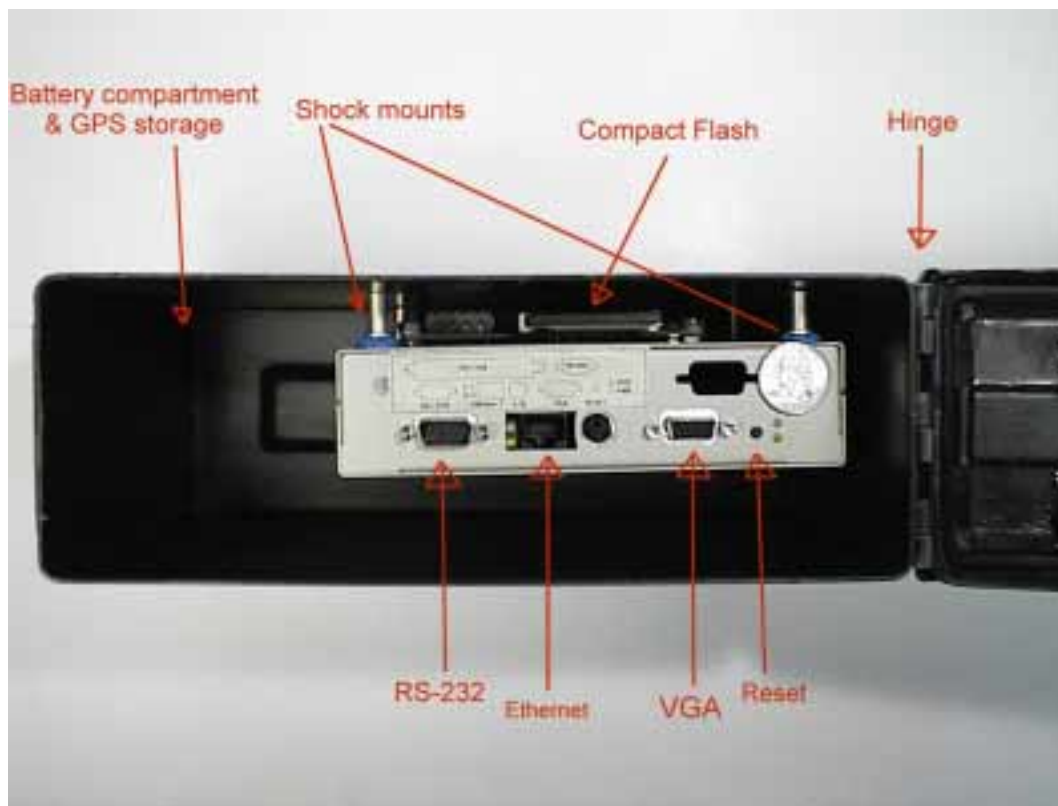
Figure 2: Extensible data collection unit, interior, showing the hardware layout and the different data ports.
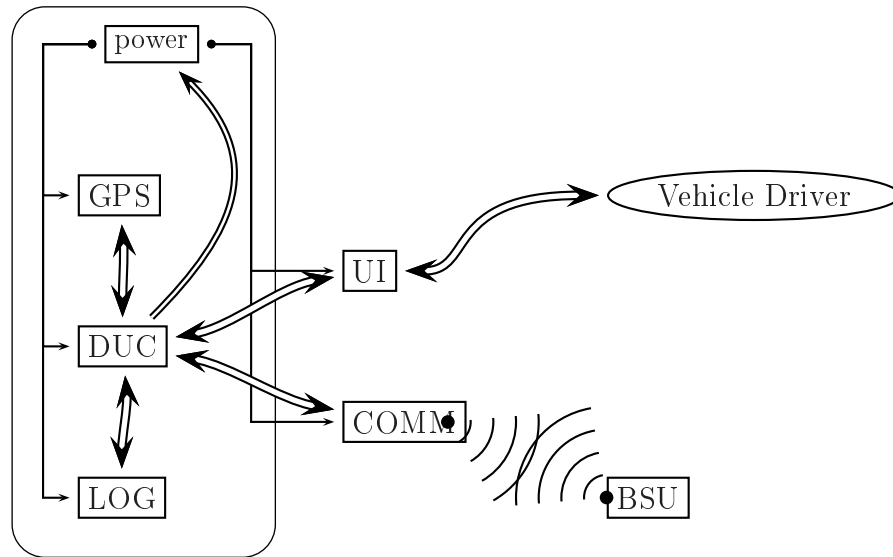
Figure 3: Block diagram of the different data collection system modules

Figure 3 shows a block model of how the six modules would interact to perform an advanced travel behavior survey. The prototype unit only includes the DUC, GPS, COMM, and LOG modules. The DUC and the LOG modules form the heart of the extensible data collection unit (EDCU). The DUC is responsible for the operation of the data collection unit, and the LOG is responsible for saving the collected data. In the prototype design, the DUC is implemented as a computer on a board, with a 16 MB flash RAM module to use as system memory. The flash RAM has been partitioned such that there is adequate space for data logging. To the software applications, the LOG unit exists as a distinct mount point in the Linux file system. Saving data to the LOG is as simple as writing to a file.

The COMM unit is handled as a communications device implementing a TCP/IP stack—in other words, it is not much different at the applications level from a modem connected to the Internet. Sending data to a base station unit (BSU) is accomplished by writing data to a data socket connected to the BSU. Receiving incoming instructions from the BSU is accomplished by listening to an open socket. In practice, getting data out of the unit is accomplished through the usual programs such as FTP, telnet, or through a web browser.

Since we chose to use the Linux operating system, and to use standard CDPD modems, the BSU module effectively becomes any computer connected to the Internet. In practice, the BSU role will be filled by a dedicated server computer, in order to ensure the integrity and security of the collected data.

The system is readily extensible to include the UI module. One of the delivered prototypes included "live" keyboard and VGA ports so that we could experiment with different input and output devices. However, the ethernet connection and the CDPD modem both offer significant advantages over more traditional interfaces. An ethernet connection to a laptop allows access to the power and speed of a full power computer to run the user interface, while reserving the memory and CPU cycles of the EDCU for the task of gathering and storing GPS measurements.

The wireless connection serves a similar function, allowing us to program any sort of communication that we desire. For example, a post-travel survey could be created on the fly by a CDPD modem connected to a survey respondent's home computer. Or a pre-trip planning stage could be transmitted to the EDCU, and the route guidance application programmed to dial the survey respondent's cell phone with recommended deviations from the planned route.

6   Evaluating EDCU functionality

The prototype extensible data collection units (EDCU) were received from the sub-contractor, AeroData, in July, 2000. Understanding the capabilities and limitations of the prototype EDCU is crucial to determining its range of application to ongoing research, and how future versions of the EDCU can be improved. A future paper will document a rigorous battery of functional tests planned for the EDCU. The remainder of this section presents our early experiences with the EDCU, and how its extensible nature has already allowed us to explore unexpected research directions.

6.1   Easy Internet access

An important design element for the EDCU is the ability to communicate with the unit over the Internet via its wireless CDPD modem. While operational tests with the modem have not taken place due to the previously mentioned procurement delays, the EDCU operates flawlessly when hooked up to the Internet via its ethernet port. The decision to use a version of Linux designed for embedded systems was instrumental in making the Internet connection easy. The EDCU is in effect a miniature web-server, providing a simple user interface over hypertext transfer protocol (HTTP) links, access to its flash ram files over file transfer protocol (FTP) links, and access to the raw NMEA output from the GPS unit via Telnet. At any time researchers can gain access to the EDCU to download data or upload data or instructions.

To answer some questions about logging and memory requirements, during development of the EDCU AeroData placed an early prototype on the Internet, allowing access to the unit's data. The initial logging script, written in perl, polls the GPS unit every 10 seconds, and stores the date, time, latitude, longitude, and

speed. That daily file, uncompressed, takes up approximately 440KB. At the end of the day, the files are compressed using gzip, resulting in a file size of about 55KB. If an EDCU was to be used in an extended travel survey function, a day's data could be zipped and uploaded in a very short amount of time, even if the connection with the cellular data modem was poor. Further, a two-week survey would only require about a megabyte of memory, if 13 days are compressed and the final day is left in its original, uncompressed state. If the EDCU were to be placed in a traffic probe vehicle uploading data in a near continuous fashion, the data stream of 1 second measurements would only require a total of 4400KB in 24 hours, which can easily be handled by even the spottiest modem connection. Under actual operational conditions, we may choose to transmit more information than just the time, position, and speed. For example, information about the satellites is important if one wishes to apply differential correction to the received signals, which we will do. Also, the small footprint of the positional data means that there will be some space on the CDPD connection to transmit other information over the link, such as new driving instructions.

6.2 Integration with GIS programs

Another area of interest for GPS-enabled travel survey data is the integration of that data with existing geographic information systems (GIS). The power of knowing where a vehicle is during a survey period is increased when one can merge that information quickly with a GIS program. To test the functionality of the EDCU in this regard, we searched the available offerings of Internet-enabled mapping software. Most existing programs carried hefty licensing fees, which is undesirable in an academic setting. The exception to this is BBN Technologies' OpenMap (http://openmap.bbn.com). The license is an open source variant that provides full access to the source code, and free use for non-commercial purposes. The OpenMap GIS is written in Java, and therefore is ideally suited to Internet-based mapping problems.

With a small amount of programming, we were able to build a map layer that probed a prototype EDCU hooked up to the Internet and posted that information on a map. While the ability to paint a dot in the middle of a map in Colorado is not in itself very exciting, the fact that we didn't *know* that the EDCU was located in Colorado prior to running the program *was* exciting. Functionally equivalent cases include mounting the EDCU in a probe vehicle and communicating via a CDPD modem to an OpenMap server running in a traffic operations center, or an EDCU in a travel survey volunteer's vehicle sending tracking data back to that volunteer's personal computer. Figure 4 shows the output of a test drive around Orange County, CA, placed with no map-matching algorithms on top of a map generated from TIGER-Line files. The data in this figure were processed using a laptop connected to the EDCU.
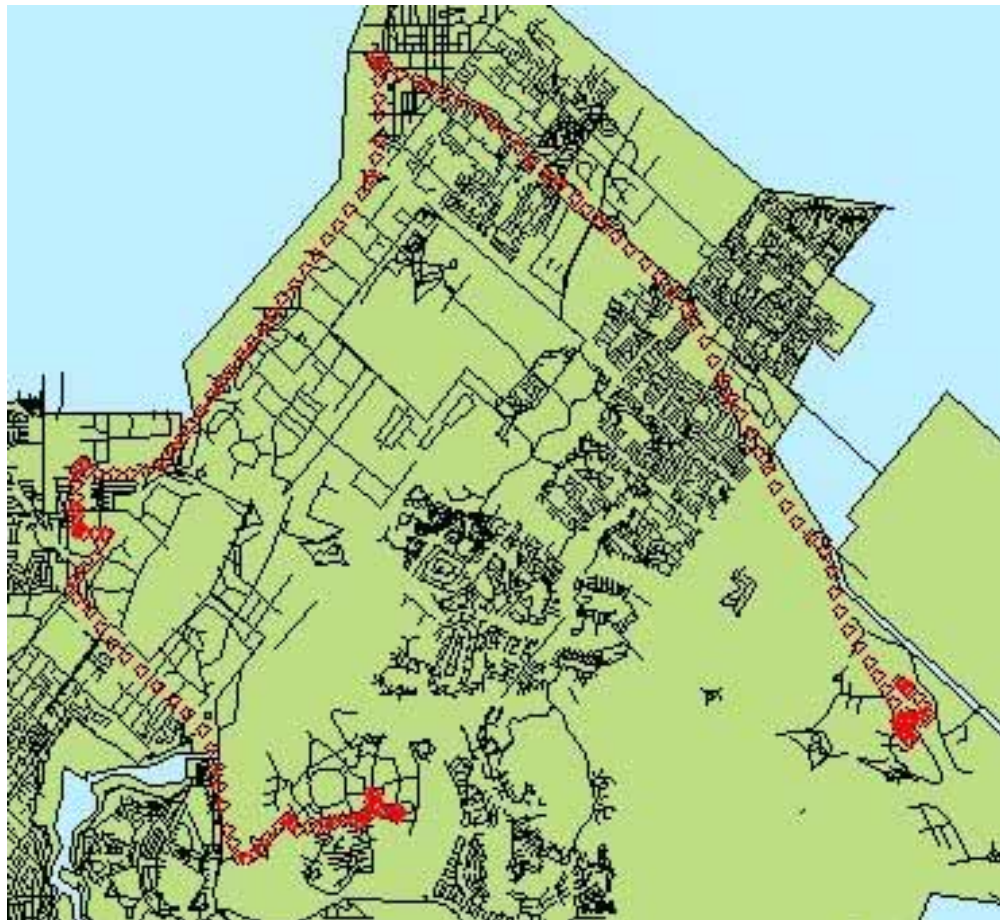
Figure 4: Trace of GPS points produced every 10 seconds. Map trace generated using custom written layer in BBN Technologies' OpenMap that can link to the EDCU and gather data dynamically. Orange County, CA boundaries and streets generated from Tiger-Line files.

FWY	Cross Street	Loop Detector Station	Time (pm)	Speed (mph)	
				Loop	GPS Avg
I5N	SR 133	1204808	8:58:20	52.1	57.6
	Jeffrey	1204924	8:59:06	52.6	58.9
	Culver	1204982	9:00:36	61.4	57.8
	Culver	1205012	9:01:21	59.2	60.3
SR55S	McFadden	1203239	9:04:20	56.0	54.4
	Dyer	1203161	9:07:30	59.7	53.4
I405N	Bristol	1201453	9:09:20	48.0	50.1

Table 1: Comparison of loop detector estimated speeds with GPS recorded speeds

6.3 Integration with TRICEPS

Using the same data points presented in figure 4, we performed a quick test of how the recorded GPS travel speeds compared to the speed estimates generated by the freeway loop detectors. The travel occurred on a week night, between 8pm and 10pm, and conditions were quite good. The Testbed was used to collect 30 second loop detector data on all of the area freeways. The position information was then compared to the loop detectors in front of and behind the recorded positions, both in time and space. The results, a portion of which are presented in table 1, showed good correspondence between the loop data and the vehicle data, which is not surprising given the light traffic conditions. The comparison did point out that many of the loop detectors were not working properly. A surprisingly high number of positions were matched to just one working loop detector.

6.4 Software platform extensibility

The choice of an embedded system running the Linux operating system, while important for enabling easy Internet access, is perhaps more important due to the ability to change the EDCU's programmed behavior. While the operating environment is fairly spartan, with initially 16MB and eventually 32MB of total system operating and storage space, the programs that control the behavior are small themselves.

To illustrate this point, we decided to explore generating XML documents containing the output of the GPS unit's National Marine Electronics Association (NMEA) sentences. Extensible Markup Language (XML) gets a lot of press for its ability to transform itself into different kinds of web pages (tailored for computers or cellular telephones, for example). However, for the transportation engineering community, XML is likely to be much more important for its ability to mark up *data* of any kind and make it highly portable. The XML alphabet soup is even

```
Date       Time      Lat        Lon          Speed
08/01/2000 03:33:08  33.660705  -117.744402  0.000000
08/01/2000 03:33:18  33.660697  -117.744408  0.000000
08/01/2000 03:33:28  33.660692  -117.744417  0.000000
08/01/2000 03:33:38  33.660688  -117.744423  0.000000
08/01/2000 03:33:48  33.660688  -117.744425  0.000000
08/01/2000 03:33:58  33.660692  -117.744420  0.000000
08/01/2000 03:34:08  33.660693  -117.744410  0.000000
08/01/2000 03:34:18  33.660690  -117.744400  0.000000
08/01/2000 03:34:28  33.660682  -117.744395  0.000000
08/01/2000 03:34:38  33.660680  -117.744400  0.000000
08/01/2000 03:34:48  33.660682  -117.744407  0.000000
08/01/2000 03:34:58  33.660683  -117.744412  0.000000
08/01/2000 03:35:10  33.660687  -117.744415  0.000000

                ...

08/01/2000 03:51:00  33.660655  -117.744425  0.000000
08/01/2000 03:51:10  33.660650  -117.744353  0.000000
08/01/2000 03:51:20  33.660480  -117.744358  6.900000
08/01/2000 03:51:30  33.660357  -117.743775  9.600000

                ...
```

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE GPS:Trace SYSTEM
  "http://www.its.uci.edu/ jmarca/DTD/trace.dtd">
<GPS:Trace xmlns:GPS="http://www.its.uci.edu/ jmarca/gpsxml/">
    <GPS:Record>
      <GPS:Time>
      <GPS:UTC>08/01/2000 03:33:08</GPS:UTC>
    </GPS:Time>
      <GPS:Pos>
      <GPS:Lat>33.660706</GPS:Lat>
      <GPS:Lon>-117.7444</GPS:Lon>
    <GPS:Speed>0.0</GPS:Speed>
  </GPS:Pos>
  </GPS:Record>
  <GPS:Record>
    <GPS:Time>
      <GPS:UTC>08/01/2000 03:51:10</GPS:UTC>
    </GPS:Time>
    <GPS:Pos>
      <GPS:Lat>33.66065</GPS:Lat>
      <GPS:Lon>-117.744354</GPS:Lon>
      <GPS:Speed>0.0</GPS:Speed>
    </GPS:Pos>
  </GPS:Record>
  <GPS:Record>
    <GPS:Time>
      <GPS:UTC>08/01/2000 03:51:20</GPS:UTC>
    </GPS:Time>
    <GPS:Pos>
      <GPS:Lat>33.66048</GPS:Lat>
      <GPS:Lon>-117.744354</GPS:Lon>
      <GPS:Speed>6.9</GPS:Speed>
    </GPS:Pos>
  </GPS:Record>
  <GPS:Record>
    <GPS:Time>
      <GPS:UTC>08/01/2000 03:51:30</GPS:UTC>
      ...
  </GPS:Record>
</GPS:Trace>
```

Figure 5: Transformation of raw GPS position log into an XML Trace document

thicker than that of the Intelligent Transportation Systems (ITS) world, and so a full description of our work would take up too much space. In sum, we developed a description of how to mark up GPS NMEA sentence data, and then added a small routine to Linux's GPS controller daemon to output using this format when requested to do so.

The XML markup codes embed in the data words that describe what each element is. For example, the latitude might be represented by `<lat>37.106712</lat>`. Readers familiar with other markup languages, such as HTML, might recognize the format, but not the tag. That is because we define the `<lat>` tag in our definition document. The extra characters required by the XML markup increase tremendously the amount of space required to store the GPS position data. For example, a 24 hour period of 10 second observations generates 440KB of data. Adding the extra characters required to include the XML tags inflates the file size up to 1,400KB! However, XML is important for communication, not for storage. It is simple enough to store the data without the tags, and then to add them on when sending the information to a requesting client.

To illustrate this concept, we developed a Java program which loaded up the

raw EDCU log files, processed them, and then generated XML output. Similar functionality was programmed in C into the actual GPS Linux driver, but has not yet been uploaded to the EDCU operating system image. The processing step, in addition to adding the XML tags, also checked whether the GPS unit was stationary (according to the GPS speed). A snippet of the output of the program applied to the same trace data used for figure 4 is shown in figure 5. This information can be passed on to any other program that understands GPS Trace XML documents, such as a mapping program or a Web-based survey tool like our own REACT! program, making collected travel data transparent to other developers.

7   Conclusions

This paper has documented the design and construction of a prototype version of an extensible data collection unit (EDCU). The motivation behind building an *extensible* data collection unit was primarily the fact that there were many different projects needing mobile GPS data. Two projects served as design points for the EDCU: the REACT! Internet-based activity survey, and the TRICEPS ATMS system. Using these two applications, and after reviewing the literature and various hardware vendors for other solutions, we designed and specified the EDCU documented in this paper.

The preliminary experiences with the unit are quite promising. The mere fact that we have a robust, extensible GPS data collection tool has prompted many new project ideas, and has led to collaborations with other researchers. The most promising aspect of the device is its wireless connection to the Internet. We expect that this aspect will prove invaluable for designing and testing the advanced traveler information systems of the future.

References

Barth, M. and Johnston, E. (1995). Using gps technology to obtain accurate speed, acceleration, and grade information for on-road emissions measurement, *Fifth CRC On-Road Vehicle Emissions Workshop*, San Diego, CA.

Doherty, S. (1999). Moving beyond observed outcomes: Integrating gps spatial traces with interactive computer-based surveys for advanced analysis of spatial-temporal patterns and underlying decision processes, *Proceedings, Transportation Research Board Data Committee's Mid-year Meeting*, Washington, DC,.

Lee, M.-S., Doherty, S. T., Sabetiashraf, R. and McNally, M. G. (2000). ichase: An internet computerized household activity, scheduling elicitor survey, *Transportation Research Board Annual Meeting, 2000*, TRB Publications.

Lee, M.-S., Sabetiashraf, R., Doherty, S. T., Rindt, C. R. and McNally, M. G. (2000). Conducting an interactive survey of household weekly activities via the internet: Preliminary results from a pilot study, *Proceedings, IATBR 2000*, IATBR, Sydney, Australia.

McNally, M. G., Rindt, C. and Logi, F. (1999). Triceps: An atmis field implementation for control and evaluation, *Technical Report Technical Report PATH PRR-99-02*, PATH, Berkeley, CA.

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

Pearson, D. (1999). Presentation, *Transportation Research Board Data Committee's Mid-year Meeting*, Washington, DC,.

Quiroga, C. (1997). *n Integrated GPS-GIS Methodology for Conducting Travel Time Studies*, PhD thesis, Louisiana State University.

Quiroga, C. and Bullock, D. (1998). A data model for managing gps travel time data,, *in* C. T. Hendrickson and S. G. Ritchie (eds), *Applications of Advanced Technologies in Transportation: Proceedings of the 5th International Conference,*, American Society of Civil Engineers,, pp. 403–410.

Roden, D. (1999). Presentation, *Transportation Research Board Data Committee's Mid-year Meeting*, Washington, DC,.

Ueno, M., Noîl, N., Doherty, S., Lee-Gosselin, M., Theberge, F. and Sirois, C. (1999). Presentation: Extending the scope of travel surveys using differential gps, *ION-GPS '99—The 12th International Technical Meeting of Satellite Division of the U. S. Institute of Navigation*, U. S. Institute of Navigation, Nashville, Tennessee.

Westerman, M., Litjens, R. and Linnartz, J.-P. (1996). Integration of probe vehicle and induction loop data: Estimation of travel times and automatic incident detection, *Technical Report Technical Report PATH PRR-96-13*, PATH, UC Berkeley.

Wolf, J., Guensler, R., Washington, S. and Frank, L. (1999). The use of electronic travel diaries and vehicle instrumentation packages in the year 2000 atlanta regional household travel survey, *Proceedings of the Transportation*

Research Board Conference on Personal Travel: The Long and Short of It, Transportation Research Board, Washington, DC.

Youngblood, T., Johnston, E. and Barth, M. (1995). Dgps technology for vehicle parameter estimation, *International Workshop on Driving Cycles and Vehicle Data-logging*, Ottawa, Canada.