

**Towards Distributed Data Collection  
and Peer-to-Peer Data Sharing**

UCI-ITS-AS-WP-02-15

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, mmcally@uci.edu

August 2002

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

## **Towards distributed data collection and peer-to-peer data sharing**

James E. Marca, Craig R. Rindt, and Michael G. M<sup>c</sup>Nally

Institute of Transportation Studies

University of California, Irvine

Irvine, CA 92697

tel: (949) 824-6571

fax: (949) 824-8385

jmarca@translab.its.uci.edu

August 1, 2002

**Abstract.** Two recent technological events should be examined more closely for their applicability to ATMIS and other ITS systems. First, the recent rise of wireless LAN protocols such as 802.11b should be studied for use in a vehicle context. Second, the rise and fall of Napster, followed by increased development on more decentralized protocols such as Gnutella and Freenet are leading to an increasingly robust peer-to-peer platforms. This paper argues that peer-to-peer communication and local area wireless networks should play a central role in any system to distribute traveler information services. Instead, the National ITS Architecture makes no mention of peer-to-peer information exchange, relegating local area wireless connections to vehicle control.

**Wordcount:** 4417 + 5 figures

## INTRODUCTION

The authors have recently completed the initial phase of development on a wireless data collection system, called Tracer (1), coupled with a dynamically generated activity survey called ANNE (2). A preliminary analysis of the data collected in a pilot study of the ANNE/Tracer suggested that the spatial distribution of activity destinations may be clustered and highly repeatable when examined longitudinally. An effort was made to develop cross-sectional activity analysis tools, but it quickly became evident that the activity names and the other free text entries elicited by the survey were unique to each respondent. In order to compare data across individuals, the individuals themselves would have to describe the meaning behind their entries.

This restriction did not apply to the more quantitative measurements recorded by the GPS antennas. Speed, position, and time-stamp are uniformly applicable across individuals. This paper starts with that observation, and explores ways in which the data collected by the Tracer system can be shared between the system participants. This decentralized data sharing progresses naturally towards more local and immediate technologies such as wireless LANs. Since 802.11b cards and home-wireless networks are extremely popular, we feel that an exploration of these ideas is appropriate before the devices begin to show up in vehicles.

## MOVING PAST WIRELESS DATA COLLECTION

As was stated in the introduction, a pilot test of the Tracer system gathered multiple months of travel data on four volunteers. This paper examines all of the four pilot survey participants together. The approach is to observe areas where information might be usefully shared between the individuals, without appealing to a central Information Service Provider (ISP).

One of the paradigm shifts that the wireless data collection made possible by the Tracer system is to view telematics technology as creating mobile content producers, not consumers. Too often the terms of the debate over traveler information systems and in-vehicle telematics are framed by commercial content providers who hope to monetize the traveler as a consumer. The streams of GPS data flowing out of EDCU-equipped vehicles are unique, original sources of information. The EDCUs are collecting information on real traffic conditions, preferred routes, and other travel information, as well as the destination point patterns that are more applicable to activity analyses. Collecting and aggregating this information could be useful to the traveler, allowing her or him to study accurate measurements of real travel times and speeds rather than relying on often erroneous estimates of the same. In addition, this information could be useful to *other* travelers. The next section documents the destinations and travel paths of the four pilot survey volunteers, and discusses the benefits of information exchange.

Mechanisms for exchanging data between individuals are explored in the subsequent section. The easiest solution is to maintain a centralized database, but travel destinations are highly personal, and should be kept private and under the travelers' control. How the Tracer system can be decentralized is also discussed, as well some of the modifications that will be necessary to make the wireless data collection cheaper to the individual, and to make the sharing of information over the Internet practical. Finally, the current National ITS Architecture (3) is reviewed in the penultimate section for evidence of and ability to integrate peer-to-peer information exchange.

## OVERLAPPING DESTINATIONS AND PATHS

There isn't much point to sharing information if each traveler already has perfect information. As there are no accepted sources for acquiring perfect information, one would have to suppose that the mechanism for becoming informed is experience—knowledge gathered from past trips and destinations. Figure 1 shows the various destinations recorded for each of the four pilot survey volunteers. The different point sets have areas of overlap—most notably the area surrounding UC Irvine, since all are UCI employees—as well as areas of unique destinations. Each person has a different but overlapping set of travel experiences, and therefore a different empirical of knowledge about the travel and activity space.

[FIGURE 1 about here.]

While the pilot survey is by no means an exhaustive record of all destinations and all paths taken by the four vehicles, it does capture a great deal of recent history. If one presumes that the degree and applicability of travel and activity knowledge are directly proportional to the frequency and recentness of a visit to an area, then the non-overlapping portions of figure 1 represent areas where one traveler has certainly more recent and possibly more complete information than the others. Further, the diagram shows only the physical destinations, not the activities performed at the destinations. If one were to include the activity performed in some portable, XML-based document format, then even the areas of overlap could become sources of new information for each traveler.

A more convincing diagram comes from examining the travel segments. Figure 2 shows the average travel speeds at all points recorded and transmitted to the base station for the volunteer labeled user id 6. The average was calculated very simply by summing up the speed observations at all repeated longitude and latitudes and dividing by the number of observations at that point. More complicated calculations of average speed are possible, for example those that take into account the speeds of each trip rather than viewing the GPS records as separate measurements. Nonetheless, it demonstrates that even with this coarse computation, different average speeds can be observed over time (and visualized in a plot) for different routes.

[FIGURE 2 about here.]

[FIGURE 3 about here.]

In the same travel area, figure 3 plots the average speeds observed by all four Tracer vehicles. Obviously a great many new links are added to the plot. On the one hand, one would expect that a person would have a feel for the expected speeds on various links by knowing what kind of facility it is (freeway, arterial, and so on). However, one can never be sure about the expected volume on a link, until one had driven on it a number of times and has built up a body of experience about the conditions on the link at different times of day.

What figure 3 shows is that user id 6 could learn quite a bit from the information collected from the other 3 vehicles. The speeds shown are the averages over all observations, but if each user had access to the full set of travel observations, queries could be set up by the time of day and the day of the week, and along an expected path or set of paths. Obviously this is not as good as full information, but it is the next best thing—broad practical experience. Further, the information can be gathered by contacting peers directly, rather than appealing to some central ISP.

## USING TRACER TO SHARE TRAVELER-COLLECTED INFORMATION

As it is implemented, there are no provisions within the Tracer system or the ANNE on-line survey to allow individuals to examine other travelers' destinations or travel patterns. Within the context of an activity survey such functionality is not necessary, and in fact should be discouraged. However, moving beyond the survey, it is possible to implement a "publish" feature, in which participants in the system could tag either individual trips or trips that fall into broad categories as publishable. While preserving each user's privacy is especially important when travel monitoring is involved, releasing time and speed data on common arterials and freeways would perhaps be acceptable, as long as the exact criteria could be controlled by the person sharing the information.

The publish feature is easily programmed because the survey itself has been implemented using the libraries of Slashcode, the software that powers the popular website `slashdot.org`. The map of travel and the quantitative analysis of the GPS records would form the basis of the published data, along with any other textual notations the user might supply. Instead of being stored in the private and restricted survey tables, the published data would be copied to a publicly accessible set of tables containing historical observations of network conditions.

Once the data is published by the traveler, then it becomes available to the other travelers in the system for queries. The current mechanism for queries within the survey web site is a very simple plotting tool which builds a plot based on checkboxes corresponding to the previous entries of the respondent. So for example, one click on "work," "meal," and the "or" checkbox to see a plot of all trips to and from activities that are labeled "work" or "meal." If other travelers' link speeds in and around these destinations are included in this plot, then one might see a more complete picture of route choices.

## FROM CENTRALIZED TRACER TOWARDS DECENTRALIZED INFORMATION EXCHANGE

The main barrier to implementing data sharing within the on-line web site is that the data are collected centrally. Privacy will be an increasing concern as wireless devices become more common. Travelers should be suspicious of other people looking over their electronic shoulders, and they should have ultimate control over what is done with information that they themselves have collected. The best way to do this is to decentralize the Tracer system and its on-line web site component. While this is not yet an option for most of the traveling public, the steady rise in cable and DSL modems means that more and more people have the ability to save and retrieve their own data on their own web server. This section comments on some of the work that will be necessary to allow this decentralization to occur while still maintaining the ability to query all travelers' published experiences.

As far as the hardware cost is concerned, the natural growth rate of computer technology should make the required hardware essentially free of charge. The current incarnation of Tracer is hosted on computers that were cutting edge two or more years ago. The hardware was available for free to the project in that it was obsolete. A Pentium III-400 is incredibly slow compared to today's \$200 beige box computers, but with sufficient RAM and disk space it is perfectly adequate for running a small website. The EDCU is currently rather expensive to obtain, but the rising popularity of GPS-enabled devices means that the prices should fall to under \$100 soon. The generic GPS logger probably will not have the ability to transmit data wirelessly, but given the

terrible price to speed ratio of CDPD service, it is probably best to expect that individual Tracer users would prefer to hand carry their data to their PC, or rely on local area wireless options.

Once the travel data is captured by the in-home Tracer unit, it is ready to be displayed with the on-line web site. The user would add notations as with the current version, and then choose what data to publish. Unlike the current system, publishing data is not as simple as setting a "publish" flag in a common database. The published data must really be publicized to the broader Internet. Similarly, the traveler must be able to find other Tracer users and their published data.

The best way to accomplish this is to rely on web standards, specifically XML and related technologies. A document type definition (DTD) will need to be designed which will allow users to know what to expect when they request data. Similarly, a technology call resource description format (RDF) has been developed which allows websites to succinctly describe their latest headlines and content. These technologies are being actively used and developed by the weblogging community, people who I view as the early adopters of the Tracer system in its web publishing incarnation.

There is little benefit to be gained by the initial adopters and publishers of travel information. As more and more people set up their own Tracer systems, of course jumping on the bandwagon is likely to provide much more information than the user will ever contribute. But at the beginning there are likely to be very few network effects. This is why I see today's on-line diarists and webloggers as the natural early adopters of the system. These individuals have already demonstrated that they are willing to publish content regardless of how many people are reading their material. The Tracer system will bring the real world to the fingertips of the weblogger. Instead of typing text and hyperlinks, the user can add longitude and latitude coordinates, and links to maps. And of course some individuals might even compulsively try to offer the most complete sets of trip data, in the hopes of becoming the launch point for all travel-related queries.

Finally if and when Tracer and similar systems become widely used, the most likely mechanism for transferring data from the vehicle to the home computer is likely to be a wireless local area connection, such as the 802.11b standard or its successors. This means that many vehicles will be traveling with local area wireless devices and GPS data collectors, with perhaps a local cache of historical information relating to the day's planned travel. A clear area for future research is to explore how one would get vehicles to share this information with each other as they are traveling, rather than waiting for the trip to be completed.

## REVIEW OF PEER-TO-PEER COMPONENTS IN NATIONAL ITS ARCHITECTURE

The National ITS Architecture (3) describes a centralized system, in which travelers and their vehicles are largely viewed as end consumers of information. The overall architecture is shown in figure 4 with the poor image quality a result of using the official architecture image files. Although it appears that the architecture encourages vehicles to talk to each other over vehicle to vehicle wireless connections, in fact the architecture specifies that these connections are to be used for automated highway systems and automatic vehicle control. The official summary descriptions of the three relevant subsystems are excerpted below from (3).

[FIGURE 4 about here.]

**Vehicle subsystem** This subsystem resides in an automobile and provides the sensory, processing, storage, and communications functions necessary to support efficient, safe, and con-

venient travel by personal automobile. Information services provide the driver with current travel conditions and the availability of services along the route and at the destination. Both one-way and two-way communications options support a spectrum of information services from low-cost broadcast services to advanced, pay for use personalized information services... Ultimately, this subsystem supports completely automated vehicle operation through advanced communications with other vehicles in the vicinity and in coordination with supporting infrastructure subsystems...

**Personal information access subsystem** This subsystem provides the capability for travelers to receive formatted traffic advisories from their homes, place of work, major trip generation sites, personal portable devices, and over multiple types of electronic media. These capabilities shall also provide basic routing information and allow users to select those transportation modes that allow them to avoid congestion, or more advanced capabilities to allow users to specify those transportation parameters that are unique to their individual needs and receive travel information. This subsystem shall provide capabilities to receive route planning from the infrastructure at fixed locations such as in their homes, their place of work, and at mobile locations such as from personal portable devices and in the vehicle or perform the route planning process at a mobile information access location. This subsystem shall also provide the capability to initiate a distress signal and cancel a prior issued manual request for help.

**Remote traveler support subsystem** This subsystem provides access to traveler information at transit stations, transit stops, other fixed sites along travel routes, and at major trip generation locations such as special event centers, hotels, office complexes, amusement parks, and theaters. Traveler information access points include kiosks and informational displays supporting varied levels of interaction and information access...

The diagrams associated with the above systems support the conclusion that the ITS architecture does not acknowledge peer-to-peer information sharing. Specifically, figure 5 shows personal information requests routing to map update providers, information service providers, emergency management, and transit management. Similarly, vehicle requests are routed to those resources, as well as to the roadway subsystem, and to other vehicles. However, the other vehicles link content only contains vehicle to vehicle coordination, and the vehicle to roadside link contains only messages pertaining to automated highway systems and vehicle probe data.

[FIGURE 5 about here.]

The National ITS Architecture may still be transformed by market forces. Peer-to-peer communication and distributed data collection and storage will be driven from the bottom up, as travelers adopt these new technologies for other reasons. The main impediment is the focus on vehicle to vehicle control messages, rather than traveler to traveler information passing. In the short term there will be a limit to the available bandwidth for vehicle to vehicle communications via local area wireless networking. The ITS focus on vehicle to vehicle control is likely to fill the channel without leaving any room for other, consumer-driven applications. This is especially true when there is a centralized information service provider (ISP) function described, linked to a rational revenue model based on premium subscription services.

The danger from an institutional perspective of taking this radical approach is that the technology is as yet unproven. Local-area wireless modems are only now being widely deployed. There is no real evidence concerning their practical bandwidth when being used on a crowded highway environment. This in turn limits the theoretical analysis and design of a distributed database. The safer option for ITS is to design around wire lines collecting data from infrastructure (road detectors, video devices, etc.), and then rely on private industry to solve transmitting that data back to the travelers, packaged as answers to travel queries.

As an early adopter of the wireless data technology, however, the Tracer system is exposed to the biggest problem with the current cellular technology approach to sending information (whether the technology is the slow but working CDPD modems, or the fast but not yet available 3G technologies). Specifically, cellular technology requires that users pay a license fee to a corporation in order to send and receive data. In the case of CDPD, the fee is astronomically high compared to the actual speed of data transfers. This fee is a significant barrier to adoption.

In contrast, the spectrum allocated to the 802.11 family of protocols, to Bluetooth, and to other less developed wireless networking technologies is unlicensed. As long as individuals adhere to the low power broadcast requirements, the FCC allows the spectrum to be used free of charge. Free is not a barrier to use, and in fact will probably encourage use to the detriment of the system as a whole. The vision offered here is a free advanced traveler information system. Collecting one's own data incurs only hardware costs, and then sharing that travel information with others incurs no resource depletion costs. Being able to share the information via free local area wireless spectrum is especially attractive.

Finally, a centralized ISP vision is much more difficult to scale than a peer-to-peer vision. Imagine a sudden accident on the freeway. Under the ISP model everybody will be dialing up to ask their own unique questions on how this accident affects them, and will be getting variations on the same answer. The central server will have to handle a sudden onslaught of queries, and the wireless transmission capacity will have to be designed to handle huge surges in traffic. The more popular the client-server model gets, the worse its performance will be under peak loads, and the more over-designed the network in general will have to be to handle these peak loads. One should expect nothing less in a system designed by traffic engineers.

In contrast, a peer-to-peer information sharing approach improves with increased adoption. While there are plenty of issues that still need to be solved, in principle the accident in the above example will cause a flood of information to propagate outward from the accident, with data hopping from one vehicle to the next, and from vehicles to infrastructure nodes where it can travel even faster. Rather than having to frame a specific query and wait for an answer, one can instead listen for new information and pass that information along, reacting appropriately if the information is relevant to the planned travel. More peers means more eyes and sensors gathering information, more aggregate storage capability, more total processing power, more geographically continuous reach, and more independently developed algorithms and ideas.

Finally, if there is one thing that the Napster phenomenon should demonstrate to those who would sell digital content, it's that information (digital content) will be shared readily by people if it is easy to do so and represents no loss to themselves. I claim that information service providers will provide services that will not be significantly differentiable from what will be common knowledge once ITS systems are pervasive. That is, there are only so many unique traffic situations that might arise, and only so many donut shops that have never been visited before. If information is freely shared between vehicles, then once one vehicle knows about a good routing strategy or a local



donut shop, that information will propagate outward to every other vehicle who might ask for that information. Similarly, Napster demonstrated that people like sharing data, and like broadening their minds with new information. The vast majority of music available for download on Napster at its height was not pop, but rather interesting odds and ends that gained a new audience. It did not diminish the quality of the music to share the music files. In fact, sharing files helped everybody on the system since there were more servers to choose from.

Similarly, it will not diminish the quality of the collected traffic information—whether real-time or historic—to share that data with others. While one might argue that a selfish hoarding of knowledge of the best route or system instabilities will allow an individual to switch unilaterally and be better off, the plots at the beginning of this chapter for just the four pilot survey participants made it clear that for every route that a traveler knows well, there are many more that are completely unknown. Sharing an “expert” route might make a commuter slightly worse off in the one case, but as an expert on the route, the commuter can probably adjust with little effort. In contrast, the occasional foray into unknown territory is likely to be fraught with bad choices and no information of local dynamics. Having access to a pool of “expert” routes on all possible paths would be a huge benefit.

## CONCLUSION

This chapter presented some early notions of how to move from the Tracer survey system to a decentralized, peer-to-peer advanced traveler information system. The core argument for the usefulness of such a system is that many of the routes and destinations of the four individuals who participated in the Tracer pilot study did not overlap. It was argued that individuals are likely to have a great deal of recent experiential knowledge about certain routes and destinations, but little to no expertise for the vast majority of potential paths and activity sites.

A discussion of the adaptations needed to convert the Tracer system into a peer-to-peer information system came next. The main conclusion is that a data type definition (DTD) or XML Schema is needed to codify what information might be shared between independent users of the Tracer system. Further, a short discussion of the likelihood of adopting the equipment ensued, focusing on personal Tracer implementations. The biggest change to the system will have to be removing its reliance upon licensed wireless data services, such as CDPD, which are likely to be too expensive for the average consumer. Instead, a more likely option is for the user to have a local area wireless hub located in the garage, say, for uploading past data, and possibly downloading the day’s travel plans and related data.

This chapter closed with a review of the National ITS Architecture, looking for peer-to-peer information exchange. No such provision was found, aside from vehicle to vehicle automatic control messages. It was argued again that peer-to-peer information dissemination is likely to scale better and be more useful in the long run than a centrally controlled traveler information service.

## REFERENCES

- [1] James E. Marca, Craig R. Rindt, and Michael G. M<sup>c</sup>Nally. The tracer data collection system: Implementation and operational experience. Working paper, July 2002.

- [2] James E. Marca. The design and implementation of an on-line activity survey. Working paper, May 2002.
- [3] ITERIS. National ITS architecture, version 4.01. on the Internet, July 2002. URL <http://itsarch.iteris.com/itsarch/index.htm>.

**LIST OF FIGURES**

1 All destination points for all four vehicle participating in the ANNE/Tracer pilot study. All vehicles have many destinations in the vicinity of UC Irvine, as well as unique information about other destinations and routes. . . . . 10

2 User id 6's observed average travel speeds at all travel points. The dark points represent slower average speeds while the lightest points are average speeds greater than 70 mph. GPS records with speed=0 are not included. . . . . 11

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. . . . . 12

4 The National ITS Architecture usage diagram. From (3). . . . . 13

5 Detail from the personal information access subsystem. From (3). . . . . 14

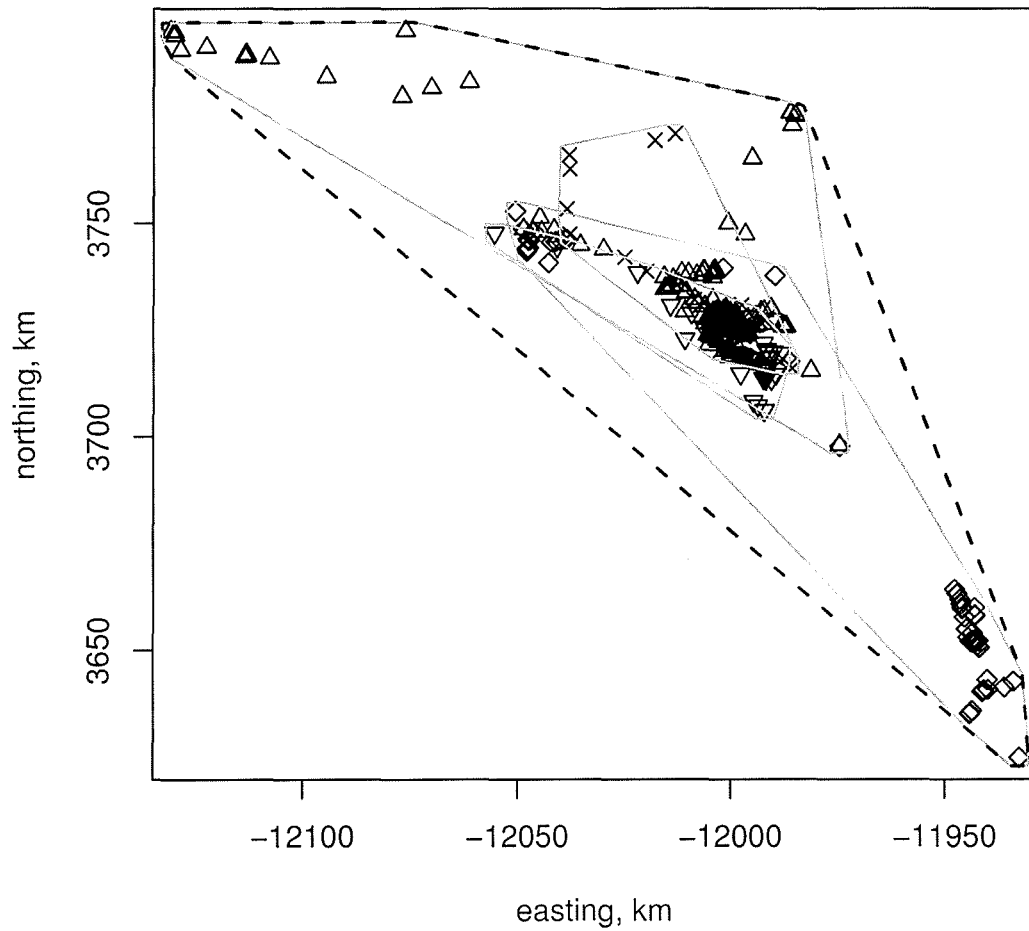


FIGURE 1: All destination points for all four vehicle participating in the ANNE/Tracer pilot study. All vehicles have many destinations in the vicinity of UC Irvine, as well as unique information about other destinations and routes.

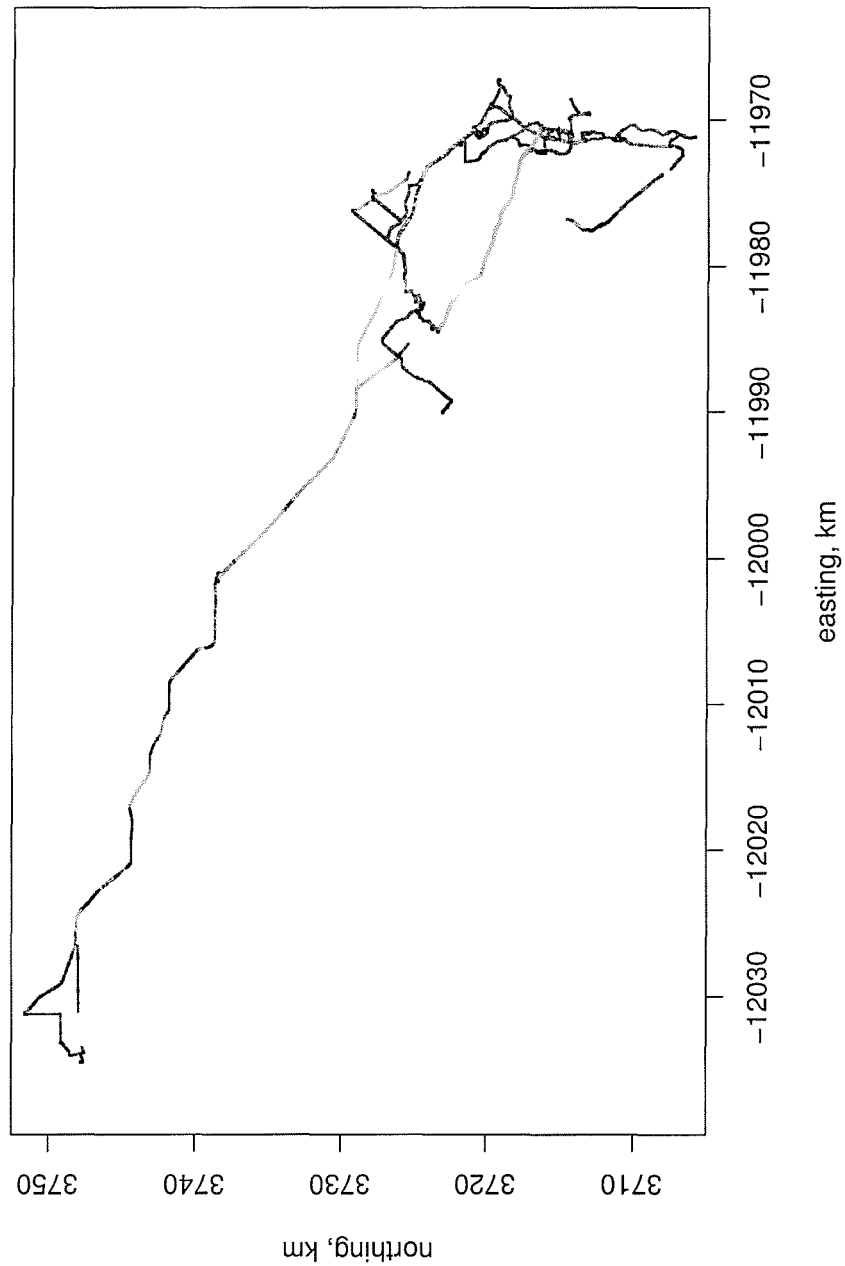


FIGURE 2: User id 6's observed average travel speeds at all travel points. The dark points represent slower average speeds while the lightest points are average speeds greater than 70 mph. GPS records with speed=0 are not included.

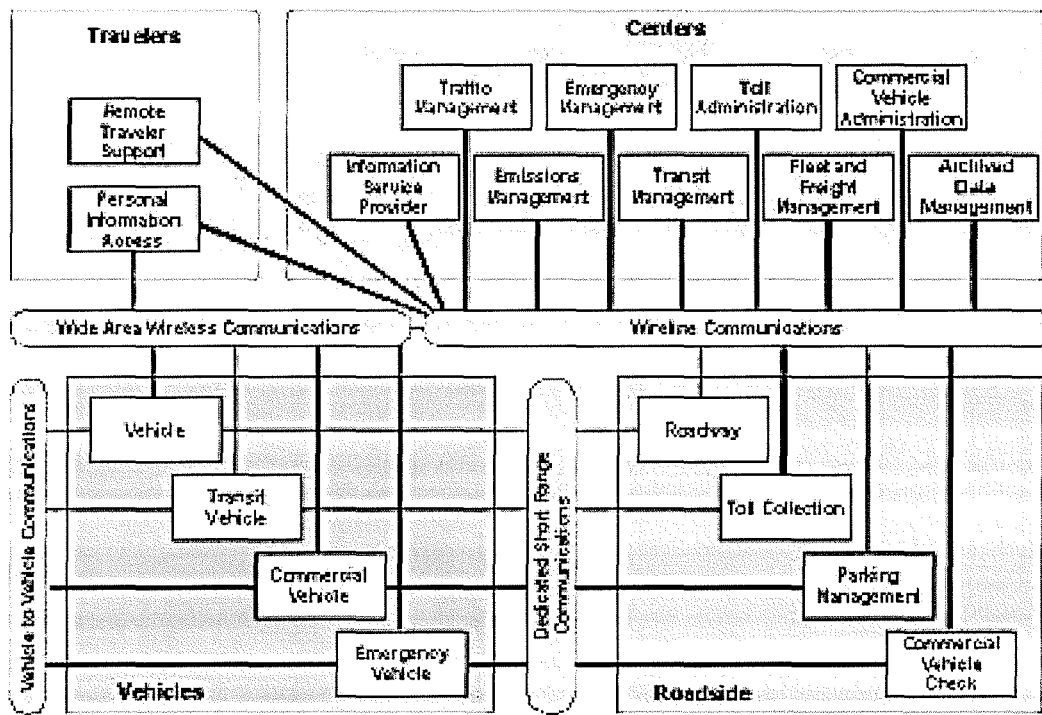


FIGURE 4: The National ITS Architecture usage diagram. From (3).

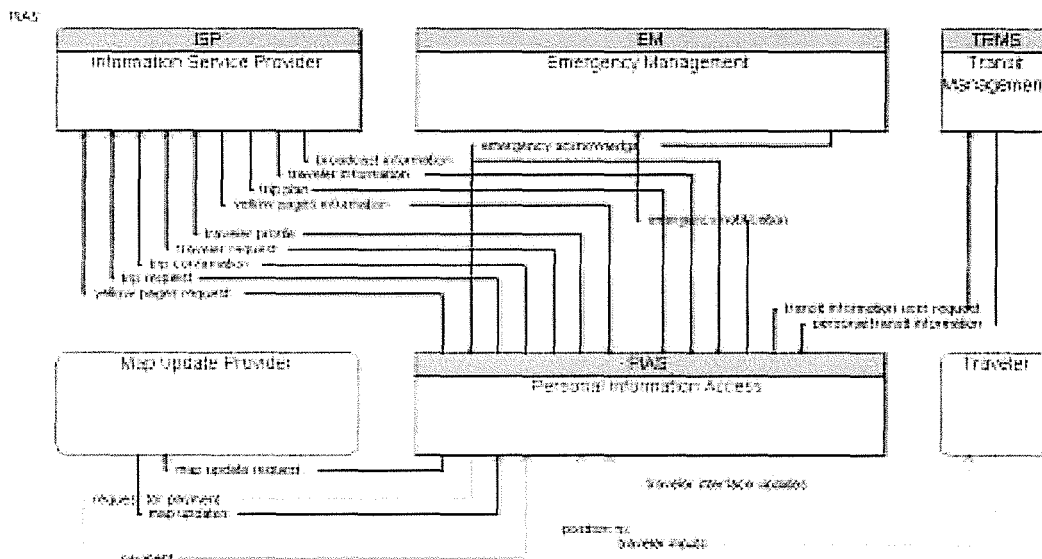


FIGURE 5: Detail from the personal information access subsystem. From (3).

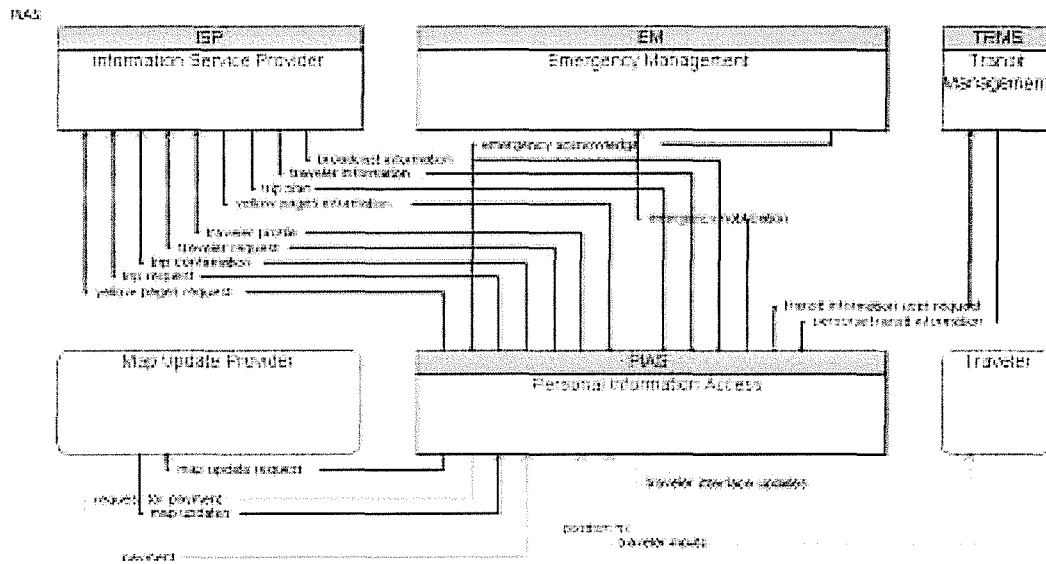


FIGURE 5: Detail from the personal information access subsystem. From (3).