GPS Use by Households: Early Indicators of Privacy Preferences Regarding Ubiquitous Mobility Information Access

Caitlin Cottrill
University of Illinois at Chicago
412 S. Peoria St., Suite 340 CUPPA Hall,
Chicago, Illinois 60607 USA
+1-312-996-4820, ccottr2@uic.edu

Piyushimita Thakuriah (Vonu)
University of Illinois at Chicago
412 S. Peoria St., Suite 340 CUPPA Hall,
Chicago, Illinois 60607 USA
+1-312-996-4820, vonu-pt@uic.edu

ABSTRACT

Privacy is a major policy factor in the deployment of Vehicle Infrastructure Integration and other ubiquitous location-based mobility access systems, yet the full benefits of many such systems will be obtained only after significant deployment of the necessary technologies. In order to address privacy concerns and build solutions that reflect users' privacy preferences, it is necessary to identify factors that affect user's privacy preferences. Using a household travel survey that also instruments a sub-sample of households with vehicle-based or portable GPS devices, we develop econometric models of privacy propensities of households as a function of their sociodemographics, trip-making behavior, location and other factors. The results provide insights into factors that might serve as barriers into the early adoption of these technologies.

Keywords

Privacy, Vehicle Infrastructure Integration, ubiquitous location-based mobility access systems, GPS, household travel survey

1. INTRODUCTION

Information regarding location is becoming pervasive. A variety of infrastructure-based systems as well as portable, wearable and vehicle-based devices can now generate location information instantly. The increased collection of data and information has naturally given rise to concerns regarding privacy. In response to these concerns, efforts have been made to address data privacy and anonymity while ensuring legitimacy of messages and data.

The effectiveness of many of these instant information access programs may depend on the level of market penetration of the technologies. For example, the core elements of the Vehicle Infrastructure Integration (VII) initiative are vehicle-to-vehicle and vehicle-to-infrastructure communication. These capabilities are to be achieved by means of intelligent on-board processing systems with wireless communications for real-time transmission and processing of relevant information. The information that is transmitted might relate to vehicle location, direction of travel, speeds and vehicle kinematics.

MobiQuitous 2008, July 21-25, 2008, Dublin, Ireland.

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3826
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3826

Such information will enable applications such as traveler information, signal timing optimization, winter maintenance and road weather prediction. Safety-related information on hazards, which can assist in collision avoidance, may also be enabled. With a larger proportion of vehicles instrumented to provide such functions, more seamless VII information will be generated. Hence it is important to examine those issues that might serve as barriers to deployment. Privacy concerns have been noted to be potentially strong barriers to seamless deployment of such vehicular and automotive technologies.

Traveler information services have been deployed over time in various media. A Seattle-area study (1) showed that broadcast traffic reports on radio and television have awareness levels close to 100%, with awareness levels over 90% among respondents of both sexes, all age groups, and at all levels of income, education, and technology usage. Many metro areas across the U.S. now also have web-based traffic information services for both highways and transit, though awareness levels of such services are low. A range of traffic services is now also available through in-vehicle navigation systems, satellite radio, HD radio and other media, which are sold to subscribers.

Technically, many of these technologies need to "know" both where the user wishes to travel and when they wish the travel to take place, such that privacy violations become a potential. However, VII, Vehicular Ad-Hoc Networks, Mobile Ad-Hoc Networks and the next generation of instant locational access technologies take privacy concerns a step further. These technologies depend on the user's location, speed and direction being transmitted in real-time either to other vehicles, Road Side Equipment (RSE) or other mobile devices. Because of potential security threats to highly personal information by hackers and others with malicious intent and potentially having to trust the individuals or organizations in charge of collecting such information, users may experience feelings of lack of control over accidental loss of information. The public could also have the perception that these digital footprints can be reconstituted, recreated and saved indefinitely, which, in the wrong hands, can be damaging.

2. RESEARCH OBJECTIVE

The purpose of this paper is to understand if variations exist among the public's acceptance and potential adoption of technologies such as VII. Our approach is the following: we conduct an empirical analysis of an ongoing household travel

survey conducted by the Chicago Metropolitan Area for Planning (CMAP). CMAP is the Metropolitan Planning Organization (MPO) for the six-county Northeastern Illinois region in the U.S., which includes the City of Chicago. The survey (the CMAP Travel Tracker Survey), was administered by Computer Assisted Telephone Interview (CATI), to 11,700 households in this area. A sub-sample of the respondents were queried as to whether they would be willing to participate in further survey activities, including using vehicle-based or portable GPS devices. The purpose of the dual CATI interviews on trip-making and GPS recording of trips is to reduce trips that are unreported, which is a problem with recall in survey-based efforts.

Using this data source, our goal is to model propensities of households to refuse to participate in the GPS portion of the survey. Refusal to use the GPS devices may represent individuals and households who perceive ubiquitous locational tracking to be problematic, and thus is indicative of locational privacy preferences. We analyze the role of sociodemographics, trip making activity and location on the propensity to refuse to use the GPS devices.

3. BACKGROUND REVIEW OF POTENTIAL SURVEY ISSUES

As noted by Haggerty and Gazso, those persons who are most concerned about privacy will generally elect to simply not participate in data gathering practices such as surveys. (2) In a research report conducted for the Census Bureau, however, Thomas Mayer noted that, "Whereas there is a dearth of quantitative research that examines the relationship between privacy and confidentiality attitudes and item nonresponse, conclusions from qualitative research indicate that item nonresponse and misinformation are potential hazards resulting from respondents who have privacy and/or confidentiality concerns." (3) Additional studies have indicated that item non-response of sensitive issues such as income and wealth may be related at least in part to concerns of privacy (4). While literature on item non-response is limited, studies related to unit non-response in travel surveys have indicated that there tend to be lower response rates from African Americans, large households, low-income households, and recent immigrants. (5,6)

- Lower response rates by African Americans (Sharp and Murakami, 2004),
- Lower response rates by large households (Sharp and Murakami, 2004),
- Lower response rates from low-income households (Contrino and Liss, 2004), and
- Lower response rates from recent immigrants (Sharp and Murakami, 2004).

4. CMAP TRAVEL TRACKER SURVEY

Household members were asked to record their travel behavior over a one- or two-day period, and these trip records will be utilized within the planning process to help evaluate the potential benefits of planned or proposed projects.

At the launch of the project, both a pilot survey and a series of community meetings took place to identify potential areas of community concern and to ensure a smooth survey process. As part of this effort, the following participation concerns were identified:

- Apathy,
- Language barriers/literacy,
- Fear, and
- Special populations (transit users). (7)

Of these identified concerns, fear is perhaps the most relevant to the issue of privacy. As noted in the white paper on maximizing participation, "In some cases…fear stems from apprehension about providing private information to strangers. This is a very real concern for travel surveys where respondents are asked for specific times and locations, not only about their own travel, but also that of their children. Other sources of fear relate to fear of government and authority." (7) Of particular concern in the CMAP travel survey effort is the recruitment of Latinos to participate, an effort of some difficulty given the current political atmosphere in relation to immigration. Community meetings identified a number of actions that may be taken to mitigate these concerns, including:

- Use of community meetings to solicit participation,
- Use of targeted information packets to increase response rates from traditionally underrepresented communities,
- Gathering endorsements from community leaders and community groups, and
- Use of targeted and general press releases and transit notices.

In addition, the CMAP Travel Tracker Survey website (http://www.nustats.com/chicago/) includes a statement relevant to privacy and confidentiality concerns, which indicates to potential participants the privacy protecting measures being taken by the survey's conductors. (8) It is hoped that the explicit inclusion of this statement will result in greater comfort with the survey and a greater willingness to participate.

5. DATA OVERVIEW

The analysis that follows is concentrated on those persons who were asked to take part in the GPS survey, as the most pertinent analysis will be based on their responses. At this point in the data collection process, 821 households that had taken part in the Travel Tracker survey had been asked to participate in a secondary GPS survey. Roughly 60% of these households reside in Cook County, with an additional 21% residing in DuPage or Lake County. The majority of households asked to participate had the following characteristics:

- Do not have children,
- Speak English as the primary language,
- Do not use transit, walk or bike,
- Own their own home, and
- Are white, non-Hispanic.

Of the households asked to participate in the GPS survey, 291 (or 36.4%) refused.

6. RESEARCH QUESTIONS

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3826
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3826

The following evaluation measures are designed to answer the following questions:

1. Are there similarities between households or individuals that refused to answer specific questions?
2. What factors affect the propensity of households to refuse to use the GPS device?

The CMAP data is divided into the following five categories:

- Household data
- Person data
- Vehicle data
- Trip data
- Location data

The number of questions for which answers were refused was calculated for each household. Table 1 provides data on the number of refusals per household. It is interesting to note that nearly 20% of refusals are generated by less than one percent of households in the survey.

Table 1: Refusals by Household

# of Refusals	# of Households	% of Refusals	% of Households
60+	6	2.91%	0.08%
50-59	9	3.68%	0.13%
40-49	13	3.97%	0.18%
30-39	38	9.32%	0.53%
20-29	133	22.79%	1.86%
10-19	293	28.47%	4.09%
1-9	1578	28.85%	22.04%
0	5090	0.00%	71.09%
Total	**7160**		**100.00%**

From the data, it appears evident that there are certain "levels" of privacy seeking behavior, with those households generating the greatest number of refusals having the greatest desire for privacy. A primary goal of this paper will be to determine if there are any statistical similarities between household demographics, travel behavior, and response refusals.

Table 2 outlines variables where the subject of the interview refused to answer the question or responded that the answer was unknown. It is evident from this table, and from the number of questions not listed, that there is a fairly clear distinction between those demographic or other data that persons are willing to share and those that they are not. For example, it is clear that at the person level, the questions of age and education level are considered more proprietary than questions pertaining to student status. Additionally, it is notable that several questions pertaining to a person's vehicle (such as the make, year and where the vehicle is parked) had a relatively large number of response refusals. The following two items of interest, however, should be noted when analyzing the data:

1. Several of the variables noted above, including whether the respondent paid to park and whether a portion of the trip took place on an expressway or tollway, had response refusals and "don't know" responses collapsed into one category. This may result in an overstatement of response refusals in these categories.
2. Because there are a greater number of records for such data sets as trips and vehicles, it is likely that there will be more response refusals in variables from these sets than from households.

Label	# of refusals	Variable	Conditions	Data Set	DK/R*
rschol	1	Grade level of school if student	If FT or PT student	Person	X
rwabik	3	Bike or walk to school		Household	X
rwhome	4	How often work at home (over 16)	If 16+, employed & telecommutes to work	Person	
rvolun	5	Volunteer employment	If 16+ & not employed	Person	X
RRibus	6	Use transit		Household	X
rwtfr	6	Fare Type	If mode=transit	Transit	
rpayf	7	Method of fare payment	If mode=transit	Transit	
remply	10	Employed (if 16+)	If 16+	Person	X
rhlive	11	Length of stay at current location		Household	
rresty	14	Description of home		Household	
rgend	14	Gender		Person	
rdtype	17	What type of disability, if disabled		Person	
rwmode	17	Work mode	If 16+ & Employed	Person	
rtelew	18	Telecommute to work	If 16+ & Employed	Person	
rhptol	18	Method of toll payment		Trip	
rsmode	21	School mode, if student	If FT or PT student	Person	
rbikes	25	Number of bicycles		Household	
rwkstat	26	Employment status	If 16+ & not employed	Person	
rlic	29	Valid license, if 16+	If 16+	Person	X
rdslic1	31	Disability plates, if disabled		Person	X
rtwext	35	Condition of disability	If disabled	Person	X
rbfhom	46	Description of home		Household	
rmake	53	Vehicle make		Vehicle	
rhisp	74	Hispanic or Latino	Only asked of main HH respondent	Person	
rphlns	79	Number of phone lines		Household	
rpervh	89	Vehicle available for work	If 16+ & Employed	Person	X
ryear	107	Vehicle year		Vehicle	
rdisab	109	Disabled		Person	
rsched	132	work schedule, flexibility	If 16+ & Employed	Person	X
rprkty	138	Park (if got out of vehicle at location other than home)	if mode=auto driver or auto passenger	Trip	
rcplns	146	Number of cell phone lines		Household	
rown	153	Own home		Household	
rrace	153	Race	Only asked of main HH respondent	Person	
rbody	155	Vehicle body type		Vehicle	
reduca	280	Level of education completed		Person	X
rage	286	Age		Person	X
rparkd	327	Where vehicle is parked		Vehicle	
rincom	665	Household income		Household	
rpaypk	3342	Paid to park (if got out of vehicle at location other than home)	if mode=auto driver or auto passenger	Trip	
rexptl	7138	Trip made on an expressway or tollway	if mode=auto driver or auto passenger	Trip	X
Total Refusals	**13790**				

Table 2: Question Response Refusals

*An 'X' in this column indicates that no distinction was made in the data reporting between a refusal to answer the question and a response of "Don't know."
** HH refers to household; FT refers to full-time; PT refers to part-time.

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3826
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3826

7. MODEL ESTIMATION

This section discusses the results of two groups of models: Model I of propensity of households to refuse to answer specific questions as a function of sociodemographic, trip-making, locational and other variables and Model II of propensity of households to refuse to use GPS devices also as a function of sociodemographic, trip-making, locational and other variables.

7.1 Model of Propensity to Refuse to Answer Specific Questions

A binary logit model was developed to analyze the probability of a person refusing to respond to one or more questions regarding the household, self, or trip. The model utilized 49,392 of a possible 110,650 records due to missing data. Of these records, 10,725 (or approximately 22%) had one or more response refusals. Measures such as the Akaike Information Criteria (AIC), Schwartz Criterion (SC) and-2 Log L indicated a reasonable fit of the tested model. According to the model, households with the following characteristics are significantly more likely to have response refusals

- Non-English speaking households
- Large household size
- Households with a large number of bicycles
- Households with a large number of cell phone lines
- Hispanic households
- Households with one or more students
- Households with a higher income

These results are, to some extent, unsurprising. As noted above, other studies have found that several of these characteristics may be associated with higher non-response rates. In short, it is evident that there are a number of household characteristics that tend to encourage protection of proprietary information. What impacts these characteristics may have on the likelihood that households will be willing to take part in GPS travel surveys will also be an area of concern, as information from both traditional and GPS surveys will be critical in making efficient transportation planning choices.

7.2 Model of Propensity to Use GPS

A second analysis was conducted to determine, first, if there is a correlation between a household's refusal to respond to survey questions and an unwillingness to participate in the GPS portion of the travel survey, and second, if there are similarities between those households that initially agreed to take part in the GPS survey. Because of data limitations, however, the following issues must be taken into consideration:

1. While the majority of the data being used for the CMAP travel tracker survey analysis has been collected through October 2007, data on those households that initially agreed to take part in the GPS portion of the survey is available only through August 2007. It is anticipated that the initial analysis will be updated as more recent data becomes available.
2. Initially, households were asked to participate in the GPS portion of the survey based upon certain characteristics (number of trips taken, presence of a car cigarette lighter to power the GPS unit, etc.). Over time, these requirements were dropped due to response rates and technology changes, but information regarding at what point these changes were made are not yet available. As a result, it is possible that results will be skewed due to the limited population initially qualified to take part in the survey.
3. The data as provided indicated only a) if the household had been asked if it would be interested in taking part in a GPS survey, and b) if the household responded yes or no. No information is provided on whether the household ultimately took part in the survey.

Keeping these limitations in mind, it is evident that the results presented here will be preliminary only. As more data become available the results will be modified and tested to ensure correct interpretation.

To conduct the initial analysis, data on those households that were asked of their willingness to participate in a GPS survey were joined to data regarding number of refusals per household. The following ordinary least squares (OLS) regression model was run to test if there is a relationship between refusal to respond to questions in the general survey and potential willingness to participate in the GPS portion of the survey:

$$NoDoGPS = f(RefsTtl)$$

where NoDoGPS indicates a household's refusal to participate in the GPS survey, and RefsTtl indicates the total number of questions that a household refused to answer.

Initial results indicate that response refusal on survey items has very little relationship to a household's indication of willingness to take part in the GPS portion of the survey. With an adjusted r-square of 0.0029 and a Pr > |t| value of 0.0656, it seems evident that a household's unwillingness to respond to demographic and behavioral questions has little bearing on the willingness to potentially take part in a survey of actual behavior. While it will be necessary to conduct additional analyses as data becomes available, the following may potentially influence the initial willingness of persons to participate:

1. Question phrasing: When asked if they would be willing to participate in the GPS survey, participants were not given specific information regarding how data would be collected. This information omission may have influenced households that would have otherwise refused to participate to indicate a willingness to take part.
2. Interest in technology: It is possible that some households indicated a willingness to take part in the follow-up survey due to an interest in the technology that will be used to gather the data.

A second analysis was conducted to determine if there are similarities in those households that indicated a willingness to take part in the GPS part of the survey. The following binary logit model was tested for its fit:

$$NoDoGPS = f(ChildDum, STRATA, EngLangDum, Cook, DuPage, Grundy, Kane, Kendal, Lake, McHenry, RiBusY,$$

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3826
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3826

WaBikY, HHVEH, HHSIZ, BIKES, OwnY, HLIVE, CPLNS, PHLNS, INCOM, HHWRK, HTRIPS1, HHLIC, White, BlackAfAm, AmIndAk, Asian, Other, HispY)

In this case, NoDoGPS is a dummy variable where 1 indicates that the household was asked to participate in the follow-up survey and responded "no" where 0 indicates that the household responded yes. From the logit model, the following variables are shown as having a significant positive impact on the likelihood that a household will refuse to take part in the GPS survey:

- Living in Lake County (a suburban county north of the City of Chicago)
- Use of transit
- Length of tenure at the current household location

The following variables may be regarded as significantly impacting the likelihood that the household will agree to take part in the GPS survey:

- Walking or biking to school
- Number of bicycles in household
- Number of workers in the household
- Number of trips taken by the household
- Being of American Indian or Alaskan Native background

In addition, an analysis of the marginal effects indicates that households tend to be less likely to take part in the GPS survey given the following conditions:

- Presence of children under the age of 16
- English as the primary language
- Household resides in Cook, DuPage, Grundy, Kane, Lake or McHenry county
- Household uses transit
- Length of tenure at the present household location
- Number of cell phones
- Higher income level
- Increased number of household driver's licenses
- Household is White, American Indian or Alaska Native, Asian or Hispanic

In contrast, households with the following characteristics tended to be less likely to refuse to participate:

- Household resides in Kendall County
- Household member(s) walk or bike to school
- Household has a large number of vehicles or bikes
- Household owns residence
- Increased number of land lines
- Large number of household members
- Household takes a large number of trips
- Household is of Black/African American or "other" descent.

While it is evident that there are some similarities in those households that indicated a willingness to participate in the GPS survey, it is difficult at this stage in the data collection and dissemination process to draw any concrete conclusions. Initial results indicate a need to delve deeper into the problem to ascertain relationships between household characteristics and willingness to share detailed and non-falsifiable travel information.

8. CONCLUSIONS

It is evident from this study that the desire for privacy will have myriad impacts on the ability of transportation planners to ensure accurate data for modeling purposes and also on the large-scale deployment of systems that depend on significant market penetration of static or mobile devices (handheld or car-based) that generate location information. Because of the significant relationships between specific household characteristics and the likelihood of refusing to reveal certain information, it is possible that systematic biases will be present in any model that utilizes data gathered from survey participants. ITS technologies have the potential to overcome these biases via more wide-ranging applications, such as combined roadside and on-vehicle sensors that will be standard equipment on new vehicles as part of the VII program, but it will still be necessary to address the preference for privacy-seeking behavior as part of any modeling framework. The results of this study are intended to provide an initial step towards determining what characteristics tend to indicate high privacy-seeking preferences and using these determinations to identify beneficial ways of implementing ITS applications on the roadway.

Future research should build upon the results identified here. Because many ITS applications will be dependent upon the collection of real-time information and route-based travel behavior, how and to what extent this data will be collected will need to be addressed within the context of the potential willingness of the user to provide data. The lack of correlation between refusal to respond to survey questions and willingness to take part in the GPS portion of the survey is, in this case, a useful result, as it indicates that participants view travel behavior as less proprietary than household demographic information. This points to the potential for widespread use of traveler-based ITS applications as long as certain privacy-preserving methods are implemented.

Balancing the need for increasing amounts of data to fully take advantage of the benefits of ITS with the desire for members of the traveling public to retain some sense of privacy in public will require a great deal of attention in the forthcoming years. It is hoped that the results of this study will begin to provide a framework by which a good equilibrium may be reached.

9. REFERENCES

1. Petrella, M and J. Lappin. "Los Angeles and Seattle: A comparative analysis of customer response to traffic information." Transportation Research Record No. 1886, Intelligent Transportation Systems and Vehicle-Highway Automation 2004.
2. Haggerty, K.D. and A. Gazso1. "The Public Politics of Opinion Research on Surveillance and Privacy." Surveillance & Society 3(2/3): 173-180.
3. Mayer, Thomas S. "Privacy and Confidentiality Research and the US Census Bureau: Recommendations Based on a Review of the Literature". Research Report Series (Survey Methodology #2002-01) Statistical Research Division, US Bureau of the Census, Feb. 7, 2002, Statistical Review of public and interviewer perceptions.

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3826
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3826

4. Riphahn, R. and O. Serfling. "Item non-response on income and wealth questions," Empirical Economics, Springer, vol. 30(2), pages 521-538. 2005.
5. Contrino, H. and S. Liss. “Sources and Impacts of Nonresponse in Household Travel Surveys: Three Case Studies” Paper presented at 7th International Conference on Travel Survey Methods. Playa Herradurra, Costa Rica, August 2004.
6. Sharp, J. and E. Murakami. 2004. Travel Survey Methods and Technologies Resource Paper. Paper Presented at the National Household Travel Survey Conference: Understanding Our Nation’s Travel, Transportation Research Board, Washington, DC, 1-2 November 2004.
7. Bricka, S. and C.R. Bhat. “A Comparative Analysis of GPS-Based and Travel Survey-Based Data.” TRB 2006. For Presentation and Consideration for Publication. Paper #06-0459.
8. Bricka, S. and C.R. Bhat. “Using GPS Data to Inform Travel Survey Methods.” Innovations in Travel Demand Modeling Conference, Transportation Research Board, Austin, TX, May 2006.
9. Khalil, A. and K. Connelly. “Context-aware Telephony: Privacy Preferences and Sharing Patterns.” Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work. Banff, Alberta, Canada. P. 469 – 478. 2006.
10. Nissenbaum, H. “Protecting Privacy in an Information Age: The Problem of Privacy in Public.” Law and Philosophy. Vol. 17, No. 5/6. Nov., 1998. Pp. 559-596.
11. Christopher, M.K., C. Simek, J. Hercules, and D. England. “Chicago Regional Household Travel Inventory White Paper: Maximizing Participation.”
12. Zimowski, M., R. Tourangeau, R. Ghadialy and S. Pedlow. “Nonresponse in Household Travel Surveys.” Prepared for the Federal Highway Administration. October 1997.
13. Christopher, M.K., C. Simek, J. Hercules, and D. England. “Chicago Regional Household Travel Inventory White Paper: Maximizing Participation.” 2007.
14. CMAP and NuStats. “Privacy Statement.” http://www.nustats.com/chicago/.

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3826
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3826