This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

Detecting air travel to survey passengers on a worldwide scale

Fabien Girardin[a,c], Pierre Dillenbourg[b] and Nicolas Nova[a]
[a]Simpliquity, Bassins, Switzerland; [b] School of Computer and Communication Sciences, Swiss Federal Institute of Technology Lausanne, Switzerland; [c]Universitat Pompeu Fabra, Barcelona, Spain

Fabien Girardin
Simpliquity Sàrl
rue du Pelaz 21
CH-1269 Bassins
Switzerland
Fabien.Girardin@simpliquity.com

Market research in the transportation sector is often based on traditional surveys, such as travel diaries, which have well documented shortcomings and biases. The advent of mobile and wireless technologies enables new methods of investigation of passengers behaviour that can eventually provide original insights into mobility studies. Because these technologies can capture travellers' experience in context and real time, they pave the road for new surveys methods. In this paper, we demonstrate that mobile phones can recognize air travel with a light algorithm that scans their connectivity to cellular networks. The originality of our method is that it does not rely on any GPS-like location information and runs on a large variety of mobile phones. It detects flights on a worldwide scale and asks travellers to report on their travel experiences as they occur, eliminating the recall bias of traditional solutions. Once the system detects a journey, it triggers a flight satisfaction questionnaire that sends answers to a centralized server. This approach respects the traveller's privacy and proved a 97% success rate in detecting flights in a 12-months study involving 6 travellers who boarded on 76 planes.

Keywords*:* Sensing and activity recognition, mobility detection, transportation study

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

Introduction

One of the greatest challenges of any survey is to produce high quality and useful data with limited budget and resources. In the past, travel surveys were mostly conducted by mail, face-to-face interviews or methods based on self-declaration such as travel diaries. With the advent of mobile and wireless technologies, combining spatial information systems and computation capacities, travel surveys may assist many types of transportation studies. Research over the last ten years suggests that locative technologies can capture trips that are often missed by traditional methods (Wolf, 2004). Other benefits include: improved accuracy of travel distance and time, routing and experience data previously unobtainable, the ability to capture longer periods of travel and capture more timely data in an economical manner (see Committee on Travel Surveys Methods, Griffiths and al., 2000). With the drop of the costs of incorporating these technologies as hardware prices have declined, multiple approaches for integrating these technologies into travel surveys have been developed. For instance, mobile and wireless technologies afford methods that capture travellers' experience in context and real time, either implicitly, with sensors embarked on a device (e.g. GPS, accelerometers), or, explicitly, by tools for capturing one's own experience (e.g. cameras, sound recorder, in situ questionnaires).

Our work further explores these benefits with a system that allows collecting longitudinal air travel data[1] at a key moment of the travel experience, just after landing. It relies on mobile phones, a device most travellers carry with them when they fly and precisely turn on once they land. Besides the demonstration of a novel methodology for

---

[1] In our case longitudinal means repeated observations over long periods of time as opposed to along the major axis (latitude, longitude).

gathering mobility data at a world-wide scale, our contribution focuses on a flight detection algorithm independent from phone network operators and taking into consideration the otherwise intractable privacy concerns. The solution includes a standard digital questionnaire that is activated when the algorithm detects a flight. The triggered survey collects time-dependent air travellers' behavioural data on their travel choices and experience including willingness-to-pay, itinerary choices and travel behaviour over time. Beyond the survey per se, this collected information is very valuable to develop and calibrate travel behaviour models such as for instance behavioural models of passenger choices. In this paper, we discuss the architecture, design, acceptance and reliability of the developed system evaluated during a worldwide test of the solution. Like any travel survey, our choice of an appropriate solution also considered the trade-offs in cost, data quality, and statistical reliability.

Background in electronic travel surveys

Regardless of the transport mean being investigated, fine-grained research on individual travel behaviours and patterns has been rare. The main reason for this lack of research is that traditional techniques such as paper diaries or direct observation are costly and require a lot of human resources. At the same time, the advances in location-sensing and mobile technologies have begun to dramatically change the methodology employed in travel survey methods (Schöfelder, 2003, Shoval and Issacso, 2006). Current approaches to automate the capture of travel behaviours can be categorized into different approaches that rely on individual automatic logging and aggregated tracing—each offering different perspective into human's travel behaviour.

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

Automatic logging typically records usage information passively without any explicit user intervention. Most research in that direction has been led in the domain of traffic and vehicular data analysis. For example, Wolf et al. (2001) demonstrated a proof of concept study to obtain trip purposes out of Global Positioning System (GPS) data to replace traditional travel diaries (7% of false trip assignment). Similarly, Schönfelder et al. (2002) presented an approach to capture longitudinal travel behaviour by means of GPS. Beside problems inherent to longitudinal surveys (e.g. limited pool of respondents, fatigue effects), the authors identify potential technical drawbacks such as: transmission problems, warm-up times before getting a fix, cost of post-processing of the GPS data. Furthermore, these experiences suggest that the fluctuating quality of traces depend on GPS errors (e.g. premature end of data stream, GPS unit warm-up time), on the GIS database inaccuracies (e.g. inaccurate coding of the land use in the parcel database) and on hardware issues (e.g. high power consumption of GPS devices) that prevent their use in context-based longitudinal surveys. Finally, there is often an extra piece of hardware that needs to be carried especially for the purpose of the study. To partially solve these issues, the authors highlight the importance of taking the user into account: "*The level of user interaction is believed to be an important issue for the development of future survey design incorporating GPS data collection elements*" (Schönfelder et al, 2002). This implies, for instance, designing the survey to enable individuals to state trip purposes to obtain a broader picture of the detected trips.

Other mobility studies rely on the Global System for Mobile communications (GSM) network as sensors of people's mobility. In a first approach, mobile devices calculate and report their position to a centralized service. The TeleTravel System (TTS)

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

project (Wermuth, 2001) combined a mobile device GSM tracking technology and an electronic travel diary to determine the travel behaviour of the respondents. In addition, some scholars work in the combination of GPS and GSM approaches to produce more complete spatial data (origin, route, destination) when tracking individual's routes (Kracht, 2004). However, they rely on the database knowledge of the cellular network topology and base stations location information to position the mobile devices, limiting the study to a specific area.

Another promising approach relies on aggregated cellular network traffic data as traces of people's movements. The measurement by mobile network operators of traffic intensity and migration of each cell in a GSM network enables capturing movement patterns from mobile phone users. Through this technique Ratti et al. (2006) measure the evolution through space and time of the density of people in a city. Based on similar network data, Aha et al. (2007) link the digital trace of visitors with visited events and locations. The use of aggregated statistics does not present traces of the individual, like her identity or trajectory: in effect the study only estimates the number of mobile phones in a given area of the city at a given time, thus avoiding privacy issues raised by other methodologies (González et al., 2008). Indeed, a major concern here lies in the potential privacy intrusion related to collecting data without individual's consent (Gutman and Stern, 2007). In studies that do not request people's permissions, there is an increased risk of identifying people or organizations, especially when the data spatial have a high accuracy. Other issues with this approach include the scalability of the tracking system, for instance, when dealing with a variety of cellular network standards and operators.

Table 1 summarizes the issues and constraints for the main approaches for electronic travel surveys.

Table 1. Summary of the current electronic travel surveys techniques with their issues and constraints.

Approach	**Issues and constraints**
Individual automatic logging (GPS)	Power consuming, fluctuating data quality, participant fatigue effect to carry extra hardware, limited pool of respondents.
Individual automatic logging (GSM)	Knowledge of the topology of the wireless network topology and base station locations that limits the study to certain regions. Our approach does rely on this information to detect mobility.
Aggregated tracing (GSM)	Agreement with GSM network operator that limits the study to certain regions.

Unlike these different ways to capture travel information, our approach relies on the mobile phone to generate "automatic passive" GSM fingerprints and trigger an in-situ questionnaire. It is an hybrid solution of implicit motion detection with the air traveller's consent and explicit disclosure of the travel experience. The constraints that led to this solution are detailed below.

Design considerations

We developed an all-terrain solution for which more rustic technologies such GSM fingerprinting for detecting mobility and Short Message Service (SMS) for survey data communication were preferred. The originality of our approach was to consider these practical constraints not as peripheral concerns to shape the delivery of technology but as the definition of the problem per se. Particularly, the system had to detect flights "*accurately*" worldwide in respect to travellers' "*privacy*". "Worldwide" means "*any country*", "*anywhere*" in the country and "*any user*". We develop these dimensions below.

Any country

"Any country" means "any operator" and "any law". "Any operator" excludes any advanced data communication; this led us to send data through SMS, the most widely used data application on the planet, with 2.4 billion active users. In addition, it also excludes any solution that requires an agreement with a cellular network operator. "Any law" implies that our solution has to be compatible to the most restrictive rules in term of tracking. This excludes any solution that does not rely uniquely on standard phone components to detect its own mobility. In addition, it is not sufficient to ask travellers once for all if they agree to reveal their location. In contrast, for each flight detection, it is crucial to ask explicitly whether the participant agrees to reveal her location and to abort the questionnaire if she answers negatively.

Anywhere

Anywhere excludes any GPS-based system since GPS does not work well in covered area such as airports. In contrast, GSM is the most widespread cellular telephony standard in the world, with deployments in more than 200 countries. As of January 2009, the GSM family of technologies has 3 billion subscribers and 80% of the world market (GSM Association, 2009). For setting up this project, we interviewed 228 UK residents departing from London Heathrow and Gatwick in early July (128 for leisure and 100 for business). The aim of this survey was to gather information on mobile cell phone usage to make sure that an optimal amount of of travellers would accept our solution. While this survey does not pretend to be representative of the population that travels by air, it gave us the necessary information to support the methodological and technical choices of this proposal. We found that 96% of respondents use a GSM mobile phone, 88% had it with them at the airport and 71% were sure to use it at their destination.

Any user

"Any user" means "any contract" and "any phone". "Any contract": the solution has to work with any type of contract with a mobile operator, which includes prepaid cards that often have roaming restrictions. This was another argument for using SMS. SMS is part of the GSM series of standards since 1985. Since then, support for the service has expanded to include alternative mobile standards such as ANSI CDMA networks. "Any phone" means that it would run on any selected participant's mobile phone. We chose Symbian[2] as the most promising standard at that time within the complex "jungle" of operating system for mobile devices[3]. "Any user" also means undisciplined air passengers who do not switch their phone off in the plane.

Accurately

"Accurately" means the algorithm has to minimize two types of errors: under-detection of flights (i.e. the traveller flew but the algorithm does not detect it) and over-detection (i.e. the algorithm infers the traveller flew while it is not the case). Under-detection reduces the quality of recorded data but over-detection is only annoying for the participants if it occurs frequently. In fact, some over-detection can be beneficial, if not too frequent, since it reminds travellers they participate to a survey. In our solution, the survey asks the participants to confirm that they indeed have taken the plane. When they have not, the

---

[2] http://www.symbian.org/

[3] Symbian counts for 46.6% of market share sales for operating systems used in smartphones in Q3 2008. Source Canalys report: http://www.canalys.com/pr/2008/r2008112.htm

system records the information to avoid repeating the mistake in the same cellular network configuration.

Privacy

A true passive logging system will not only need the approval of the participants but, for some countries, the approval of the Government. The term tracing can be interpreted as electronically following the movements of a human being without him noticing it. In contrast, a system will not be considered as tracing and will fall into normal mobility and transportation research activity if it can: a) detect travel only b) tells the traveller that this fact is to be communicated c) and asks for confirmation with a question such as "please confirm that this information can be communicated".

In the remainder of this paper, we first describe the solution we implemented and its architecture. Particularly, we present the resulting rule-based algorithm to detect air travel from GSM fingerprints. Subsequently, we describe the results of the worldwide field experiment performed over a 12 months period. We compare this work with similar approaches and conclude with a discussion on the contributions of this study and their implications for future works.

Implementation

A GSM base station is typically equipped with a number of directional antennas that define sectors of coverage, or cells. In our study, we built a software mobility sensor based on wireless signals received by the mobile phone. We wrote a custom application for Symbian mobile phones to measure and record the surrounding GSM radio environment at a constant interval of 15 minutes. We will see later on that the choice of this interval is very important. Each reading generates a GSM fingerprint called a

Location Area Identity (LAI) that uniquely identifies a location area within any mobile network. More specifically LAI comprises the Mobile Country Code (MCC), Mobile Network Code (MNC) and Location Area Code (LAC) (Figure 1).

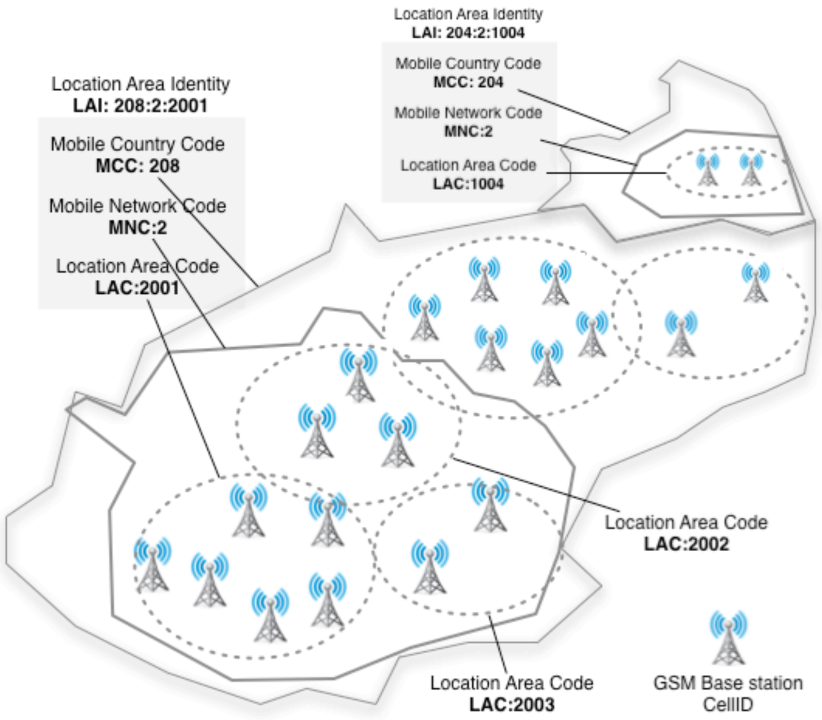


Figure 1. A GSM network in a country assigned with a Mobile Country Code (MCC). Within this country each mobile operator is assigned a unique Mobile Network Code (MNC). The operator deploys its GSM base stations and aggregate them in specific areas indentified with a unique Location Area Code (LAC). Our implementation exploits these unique 3 codes that form a LAI (Location Area Identity), discarding the CellID of the base stations, to detect the mobility of air travellers.

With an analysis of the sequence of LAIs, phones can detect their mobility without interaction with a central service. We refer to 'triggers' as the criteria that the phone software uses to detect that the traveller has taken the plane.

System architecture

The architecture of the software is divided into five modules (Figure 2). When the mobile phone is switched on, an autostart module (module 1) waits for the user to unlock the device (i.e. enter her PIN code) and launches the GSM Tracker module (module 2) after a fixed delay of 45 seconds. This delay has been set-up to avoid interferences with other applications and to leave the necessary time for the mobile phone to establish a connection with a local base station. The tracker module scans the GSM network fingerprints in a scheduled interval of 15 minutes. The travel trigger module (module 3) uses an algorithm to analyze the retrieved LAI. In case the algorithm does not detect a travel, the last 2 retrieved LAIs are kept as a pair in a non-travel list. If a travel is detected, a questionnaire module (module 4) starts with a graphical interface that appears on the device's display. Questions are loaded from a text file and answers are sent via SMS by the communication interface module (module 5). The SMS mechanism we developed uses a fail-over technique that handles the message to be sent. In case of lack of network connectivity, the message stays in a queue that empties when the mobile phone return under GSM coverage or roaming accessibility. The system keeps on trying to send the data every 30 minutes until successful delivery.

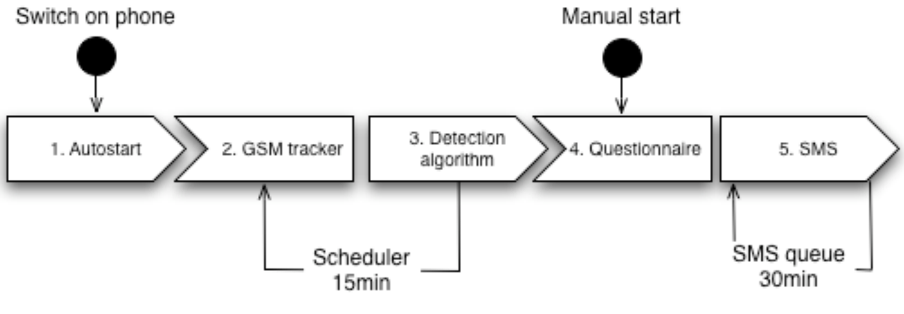


Figure 2. The five modules of the system architecture.

For privacy concerns, we let users the possibility to exit from the system, if they do not want their mobility to be known in any way, even by their own mobile phone. Therefore we also give users the opportunity to manually launch the questionnaire.

Software implementation

We decided to implement the system for the Symbian OS operating system because, at the time of development, it was the unique platform that allowed to 1) programmatically start an application on a mobile phone boot mechanism (for the Autostart module) easily call the execution of other modules (to start the GSM tracker, the questionnaire and SMS modules when necessary), 2) use a "scheduler" to systematically call a module (the GSM tracker every 15min) and finally 3) get access to the device network interface for the GSM tracker to retrieve the current LAI. In contrast, these requirements were impossible and complicated to develop on JavaME[4], a platform generally easy to program on.

The walled garden context in the mobile industry challenged the implementation of the system. Indeed, we found out that the developed code would run appropriately on the diversity of platforms. For instance, some mobile phone manufacturers did not properly implement the Symbian operating system or some mobile operators disabled some features (e.g. programmatic access to the network interface). Therefore, we had to develop different codebases according to some mobile phones or operators unpublished limitations. Similarly, we discarded any solutions that relied on proprietary software (such as Google MyLocation) that does not guarantee an open API.

The strict management of power consumption of the system was another constrain of the implementation. It prevented us in considering opportunistic scenarios to implicitly communicate the travel data. It is another reason why SMS was preferred over other communication interfaces such as Bluetooth or WiFi that are very energy demanding when scanning for their environment. Moreover, the implementation of these modules would have added unnecessary complexity (e.g. with additional code bases) to our simple

---

[4] http://java.sun.com/javame/

solution. In our longitudinal evaluation, the fractional use of the mobile phone processor to launch the GSM tracker every 15 minutes remained unnoticed.

Air travel detection algorithm

The algorithm detects air travel with the same principle as fingerprint-based location systems (Hightower et al., 2005): the radio signals observed from fixed sources are consistent in time, but variable in space. Thus, given a series of wireless scans with similar fingerprints, we conclude that the phone is not moving. Similarly, we interpret changes in the sequence of fingerprints as indicative of motion. The tracker module scans the GSM network at in interval of 15 minutes. The retrieved LAI uniquely identifies a LA (Location Area) within any PLMN (Public Land Mobile Network). Note that the LAI does not contain the identifier of the base stations (CellId), but that is not necessary since there we are not looking for finer network information. Indeed, depending on the urban density of an area, a LAC can extend from few kilometres to around 100 kilometres.

Our algorithm is based on the principle that ground transportation enables continuous connection between the cellular network base stations and the traveller's phone, while a flight leads to an interruption of connection (Figure 3).

After the collection of weeks of LAIs, we were able to simulate and refine a very simple algorithm that replaces complex GPS approaches by common sense balance of two types of mistakes, the false positive and the false negative. The algorithm minimizes the risk of false negatives (i.e. the user has indeed taken a flight that has not been detected) but augments the risk of false positive as explained below. During ground transportation, the user may loose connection for a few minutes, for instance in a road tunnel. This will not be considered as a travel as long as the time without connection is below 30 minutes. In cases where the time without connection is longer and associated with a change of LAI, for instance a long metro ride, we face a false positive. In this case, the system prompts a question on the mobile phone "*It seems that you have taken a plane, is that correct?*". In response, the participant answers negatively and the system adds the LAI pair to the non-travel list. Hence, it learns not to prompt the questionnaire anymore for that specific sequence. False negative cases will hence quickly disappear from any frequent travel pattern.

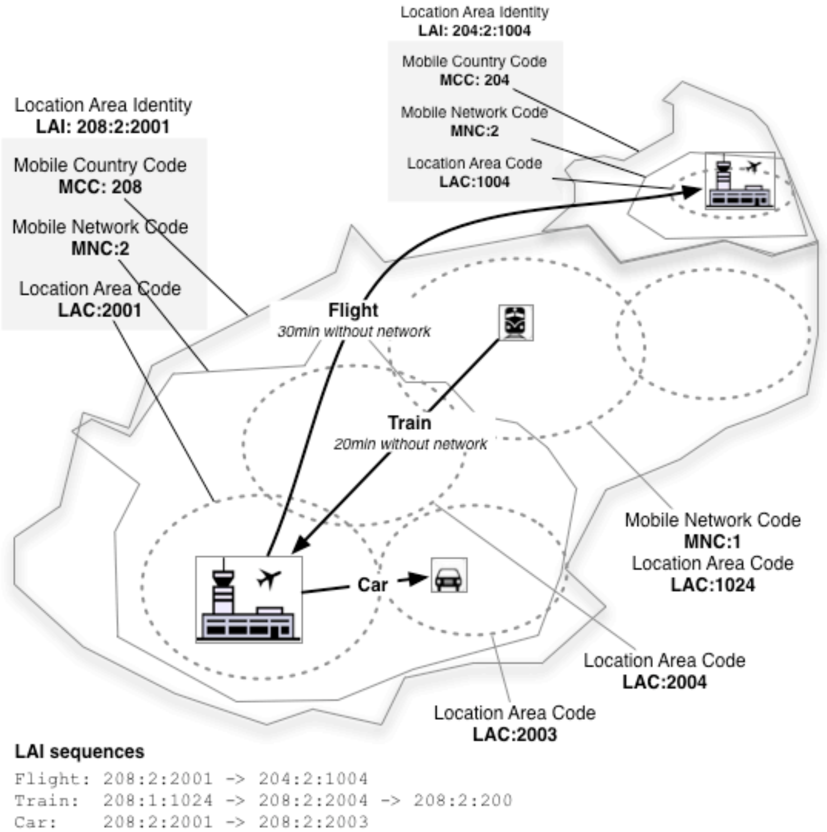


Figure 3. The algorithm analyses the sequences of LAIs. This figure shows 3 examples: In flight, the mobile phone roams from one country code to another. If our software does not retrieve any LAI within 30 minutes, it detects an air travel. In a train, the mobile phone moves within different network providers and area codes. No survey appears if the disconnection periods do not exceed 30 minutes. Similar scenario takes place for a car that moves within different location areas.

We first describe how this principle has been implemented and then what are its exceptions (Figure 4). The algorithm obviously decides that if the fingerprint hasn't changed (LAIn = LAIn-1), the phone user has not taken a plane since a LAI does not cover two connected airports. Second, the algorithm checks if the pair of fingerprints [LAIn-1, LAIn] is present in non-travel lists. This list is central to our algorithm. Third, according to the previously mentioned principle, the algorithm decides that if the phone did not succeed connecting a base station during 30 minutes or more, the user has probably taken a flight.

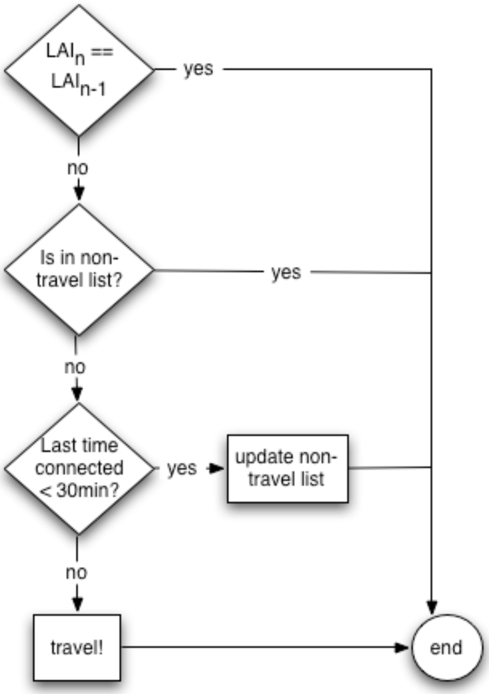
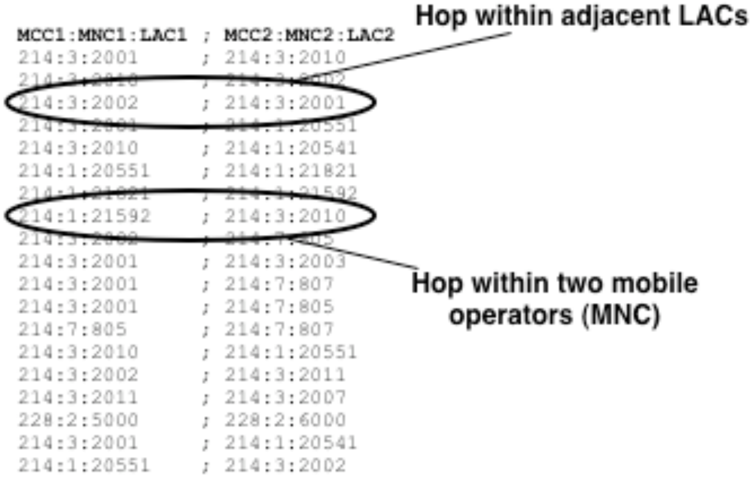


Figure 4. Travel Trigger algorithm

If the last successful reading took place within less than 30 minutes, the system records a list of LAI pairs, called non-travel list that contains all consecutive locations between which no flight occurred. Figure 5 is an example of a non-travel list generated after 3 weeks of travel from an individual living in Spain and sometimes flying to Switzerland. It illustrates the fact that most of the non-travel events are due to a hop to a proximate LAC and transition of MNC handling connectivity within the same country (MCC).



```
MCC1:MNC1:LAC1  ; MCC2:MNC2:LAC2
214:3:2001      ; 214:3:2010
214:3:2010      ; 214:3:2002
214:3:2002      ; 214:3:2001
214:3:2001      ; 214:1:20551
214:3:2010      ; 214:1:20541
214:1:20551     ; 214:1:21821
214:1:21821     ; 214:1:21592
214:1:21592     ; 214:3:2010
214:3:2002      ; 214:7:805
214:3:2001      ; 214:3:2003
214:3:2001      ; 214:7:807
214:3:2001      ; 214:7:805
214:7:805       ; 214:7:807
214:3:2010      ; 214:1:20551
214:3:2002      ; 214:3:2011
214:3:2011      ; 214:3:2007
228:2:5000      ; 228:2:6000
214:3:2001      ; 214:1:20541
214:1:20551     ; 214:3:2002
```

Figure 5. Example of a non-travel list. Each line constitutes a pair of "proximate" LAIs with the format MCC1:MNC1:LAC1;MCC2:MNC2:LAC2. This example show hops within adjacent LAC and mobile operators. For privacy concerns, we do not associate any time to these records. In addition, it must be stressed that the mobile phone does not communicate this data.

In-situ questionnaire

The survey component of the project consists in a short questionnaire that starts immediately after travel detection to collect passengers' travel experience. More specifically, two questions are first prompted to the user. First, the system asks her to confirm the fact that she has taken the plane. Second, the system asks the traveller to confirm explicitly the fact that she accepts her trip information to be communicated to the centralized survey server. It may be the case that the participant is busy with passport or custom controls or looking for his or her luggage. In other situations, she may start answering questions and then be interrupted by a phone call or any external event. In both case, the application regularly prompt the traveller to complete the questionnaire.

The questionnaire supports several forms of questions, such as single choice, multiple choice or text entries. An example of the text file would contain a 1-minute questionnaire on the a) acceptance of record of location; b) flight number; c) overall opinion on the flight; d) overall feeling after the flight.

Data communication

Once the questionnaire has been completed, the phone compiles the answers and the LAI information within a text message and sends them via SMS to a centralized server, as for example:

```
Survey:65535 FlightNo:LX1953 LAI1:{30-04-
2006,19:14:29,228:2:5000} LAI2:{30-04-
2006,20:14:32,214:1:20541} Answers{1,0,3}
```

This data is then stored in the survey database. The advantage of SMS over any other data communication solution is first that they generate very low costs (for this type of study that collects a limited amount of data), and second that they work in limited connectivity situations with a basic GSM networks. Finally, if there is no connection, the SMS will be sent later when the connection is restored. For instance, one traveller flew to the West Coast of the USA with a Swiss pre-paid SIM card. All his flights were correctly detected on his mobile phone and the participant answered the surveys. However, since his operator did not have roaming agreement with any American operator in that region, the survey data were received at his return to Switzerland. The evaluation period served us in testing and collecting further practical experiences.

Evaluation

The algorithm and data collection process have been tested and improved for 12 months between February 2006 and February 2007. GSM logs have been collected from multiple persons and types of phones. 6 frequent travellers, familiar with the project, were involved in the project as participants. 3 used their own mobile phone and 3 others used a

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

phone we gave them and on which we had pre-installed the software. They were not genuine users, although we noticed that when doing professional or family trips, they completely forget about the travel detection application and tend to behave normally until the phone vibrates and rings to ask travel questions. We benefited from the familiarity of the participants to the project to request some of them to not keep their phone switched on during the flights where it matters, such as in short continental flights.

Participants used Subscriber Identity Module (SIM) cards from Switzerland (from where the 3 major operators have been tested), Spain (1 operator) and the UK (1 operator) on Symbian series 60 phones. They boarded 76 planes with our software, connecting 27 airports in Europe and North/Central America, Asia and Australia as represented on Figure 6.

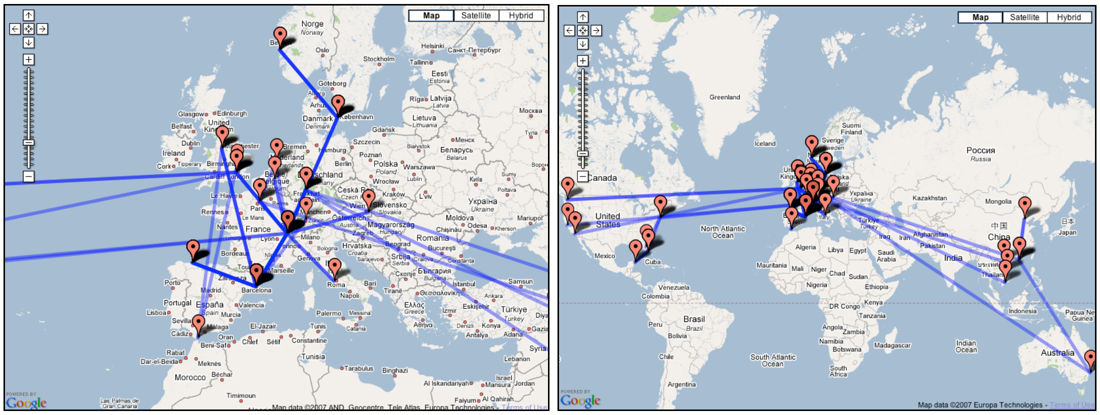


Figure 6: European (left) and Worldwide (right) detected travels. Maps generated from the SMS received by the centralized server. The departure and arrival airport are retrieved from the flight number disclosed by the traveller.

First, the evaluation of the system showed that false positives did not annoy nor was portraying the system as unreliable technology. Nevertheless, further empirical evidence would be necessary to fully confirm that behaviour.

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

Second, the algorithm has been successful in detecting flights and avoiding "false negative" in real-world settings. Among the 76 flights, 74 air travels were properly detected (97%) and data correctly reported to the centralized survey server. We had the case of two undetected flights. One because, the traveller did not switch on his phone on a stop-over therefore missing the data on the first flight. The other undetected flight took place on a short flight in which the tester did not switch off his phone. This suggests that the main limitation of the algorithm lies in the possible behaviour of some passengers, which correspond to these two cases.

Stop-overs

We noticed that stop-overs are potentially problematic to detect with passengers not switching on their phones. If the traveller has a short stop-over between two flights, it might be the case that she does not turn on his mobile phone. In this case, she will receive the questionnaire at the end of the second flight and responds the survey on the two segments of his journey.

Short flights

Similarly phones not switched off in the plane are prone to not detect flights, particularly for short flights. When a plane stays at a low altitude over powerful cellular networks, it may be the case that the phone continues to connect to base stations during the flight. This scenario happened on a test flight from Geneva, Switzerland to Munich, Germany.

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

Related works

Works related to our system cover three themes; motion detection with wireless fingerprints; activity and place recognition from motion detection and in situ surveys triggered with activities and places recognition algorithms.

Mobility sensing and motion detection with wireless networks

Similar to our approach, research groups have been investigating mobility pattern detection without the use of spatial coordinates as a location indicator, but instead sets of consistently heard radio transmitters. Asakura and Iryo (2007) determined it by interpreting changes in the set of nearby base station and signal strengths of the phones as indicative of position and motion of their owners. Sohn et al. (2005) gathered GSM network trace data over a period of one month, logging a total of 249 walking events and 171 driving events. Their methods show that GSM-based sensing may provide enough activity information for some applications, without the overhead of requiring additional sensors. Our work further proves that motion detection using GSM is feasible and works well throughout people's long distance travels.

In addition, this research avenue has been investigated in the field of lifelogging services in the last five years. The underlying idea of these applications lies in the recording of every digital interaction that can be captured by cell phones or mobile devices. It then allows users to access, share and search personal memories. From all the data that can be collected, spatial movements are of course of interest to researchers. For instance, an application such as LifeTag (Rekimoto et al., 2007), based on WiFi tracking, can also be seen as a solution for mobility sensing. As argued by the authors, the advantage of such method is that WiFi allows indoor-tracking (as opposed to GPS). We

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

can also wonder whether WiFi, if the infrastructure is properly set, can enable finer-grained positioning than GSM.

Place recognition

In addition to mobility surveys, investigation has been performed in the field of place detection. The general problem of recognizing significant places from location data has received much attention. Hightower et al. (2005)'s BeaconPrint algorithm finds repeatable sets of GSM and Wi-Fi fingerprints to automatically learn the places someone go and then detect when they return to those places. Similarly Zhou et al. (2005), looked at the relationship between place discovery and importance and found that those places judged meaningful by the subject were much easier to detect. Liao et al. (2003) made an attempt to automatically determine which of the important places is the user's home. They used machine learning on labelled place data to achieve 100% classification accuracy in finding locations of their five subjects' home and work places. Nurmi and Bhattacharya (2008) went a step further with an algorithm that can accurately identify places without temporal information.

In-situ questionnaire

The ability to detect motion and recognize place enabled the development "context-triggered sampling" (Froehlich et al. 2006). This technique, pioneered by MIT's Context-Aware Experience Sampling tool for the PDA (Intille et al., 2003), uses sensors to infer context to trigger a brief survey and capture data on participants' thoughts, feelings, and behaviours as they are experienced. It has several advantages when compared with traditional sampling methods, such as random or time-based triggering. For example, context-triggered surveys are much more likely to occur during events that are of interest to capture offering a distinct methodological advantage since they do not rely on the reconstruction of information from memory or logs therefore minimizing recall bias of self-reporting methods. Additionally, the computer can continuously save context data; allowing the researcher to cross-check answers with sensor data and uncover behavioural patterns not initially considered.

However, these solutions use extra sensors and do not capture the travel experience because restricted to very defined areas. To our knowledge, the method we propose is original because it runs worldwide, on any GSM network, it collects and reports survey data on air travel experience – with respect to survey participants' privacy - at a key moment of the travel experience, just after a landing.

Conclusion and future work

Detecting travel is difficult due to the challenge of building informative, yet unobtrusive algorithms that respect privacy. GPS sensing is available a low percentage of time of a typical person's day, as it needs a wide swath of clear sky to sense enough geostationary satellites (LaMarca et al., 2005). Moreover, it is very power consuming that makes it an unpractical solution for longitudinal surveys. Several tests have been conducted to log personal movements using GSM mobile phones. Some advantages of mobile phones relative to a GPS system is that they function underground and inside buildings, and the density of cellular base stations is higher in the most dense urban areas. Also, the market penetration of cellular phones is very high, so the cost of equipment is low. Recently, several studies tested the use for tracking personal travel, especially as many mobile phones already have the capability of recording and storing the position over time. They show that the position accuracy afforded by cellular phone is lower than GPS. Hence, while gross measurement of travel is achievable, specific routes or travel modes are less likely to be determined without greater participant interaction.

Our contribution demonstrates a novel methodology for gathering mobility data at a world-wide scale, contrasting with traditional travel survey methods. We have demonstrated the feasibility of using a basic, unmodified, GSM phone, owned by 3 billion subscribers worldwide, to recognize air flights. Our solution applies to all areas with GSM coverage (860 networks in 220 countries and regions in early 2009[5]). Our algorithm and data collection process has been evaluated over 12 months with a 97% successful flight detection rate (74 out of 76 flights) When extended to a larger scale, the collected information could become very valuable to develop and calibrate travel behaviour models such as for instance behavioural models of passenger choice. However, more empirical evidences are necessary to test the robustness of the system and collect more observation of participants' perception of false positives. Nevertheless, our contribution shows that:

1) Simple behavioural patterns, captured in our case as the LAIs pairs, constitute a powerful alternative to complex technological developments for longitudinal travel survey without modifying the habits of the participants. This approach could be extended for detecting other types of transport (e.g. train, boats, ski) on different mobile devices such as watches and cameras, not by using with the same algorithm, but by applying the same principles as in our light algorithm.
2) National regulations regarding tracing people could be an obstacle to any travel detection system. At the opposite, a system that warns the traveller that this fact is to

---

[5] Source GSM Association http://www.gsmworld.com/roaming/gsminfo/index.shtml

be communicated through as survey is not considered as "tracing" and falls under normal travel and transportation research activity.

On a final note, while many computer scientists and engineers consider practical constraints as detrimental to the elegance of technological solutions, we instead view them as opportunities to rethink solutions which eventually have to change over time. The airlines rules will change the social practices and require a revision of our algorithm, with respect to the use of mobile phone on board, the cost of communications, and cultural habits. For instance, lately, there is a new movement to allow passengers to use their own mobile phones among several airlines launching their own in-flight services (e.g. using OnAir[6]). In consequence, our system will certainly need to adapt to this specifically in a near future. Therefore, a major advance would be that the data collected by survey participants allow to develop a permanently updated database of GSM base stations for base stations installed in planes and in all airports worldwide. This extremely valuable database would be progressively enable simpler flight detection techniques for future transportation studies.

References

Ahas, R., Aasa, A., Silm, S., and Tiru, M. (2007). Mobile positioning data in tourism studies and monitoring: Case study in tartu, estonia. In Marianna Sigala, L. M. and Murphy, J., editors, *Information and Communication Technologies in Tourism 2007*, Ljubljana, Slovenia. Springer Vienna.

Asakura, Y. and Iryob, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice*, 41(7):684–690.

Froehlich, J., Chen, M. Y., Smith, I. E., and Potter, F. (2006). Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In Dourish, P. and Friday, A., editors, Ubicomp, *Lecture Notes in Computer Science*, pages 333–350. Springer Berlin / Heidelberg.

---

[6] http://www.onair.aero/

González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, (453):779–782.

Griffiths, Robert, Richardson, A.J., Lee-Gosselin, Martin E.H. (2000). Travel Surveys, Transportation in the New Millennium, Transportation Research Board (http://gulliver.trb.org/publications/millennium/00135.pdf)

Gutman, M. and Stern, P., editors (2007). Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data. National Academies Pres

GSM Association (2009), Market Data Statistics, http://www.gsmworld.com/newsroom/market-data/market_data_summary.htm (last visited, January 28 2009)

Hightower, J., Consolvo, S., LaMarca, A., Smith, I., and Hughes, J. (2005). Learning and recognizing the places we go. *In Proceedings of the Seventh International Conference on Ubiquitous Computing (Ubicomp 2005)*, Lecture Notes in Computer Science, pages 159–176. Springer-Verlag.

Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., and Bao, L. (2003). A context-aware experience sampling tool. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 972–973, New York, NY, USA. ACM.

Kracht, M. (2004). Tracking and interviewing individuals with gps and gsm technology on mobile electronic devices. In *Seventh International Conference on Travel Survey Methods*.

Liao, L., Fox, D., and Kautz, H. (2005). Location-based activity recognition using relational markov networks. *In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*

LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I. E., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., and Schilit, B. N. (2005). Place lab: Device positioning using radio beacons in the wild. In Gellersen, H.-W., Want, R., and Schmidt, A., editors, *Pervasive*, Lecture Notes in Computer Science, pages 116–133. Springer.

Nurmi, P. and Bhattacharya, S. (2008). Identifying meaningful places: The non-parametric way. In *Pervasive*, pages 111–127.

Ratti, C., Pulselli, R. M., Williams, S., and Frenchman, D. (2006). Mobile landscapes: Using location data from cell-phones for urban analysis. *Environment and Planning B: Planning and Design*, 33(5):727 – 748.

Rekimoto, J., Miyaki, T. and Ishizawa, T. (2007). LifeTag: WiFi-based Continuous Location Logging for Life Pattern Analysis. *3rd International Symposium on Location- and Context-Awareness (LOCA2007)*, pp.35-49

Schöenfelder, S., Axhausen, K., Antille, N., and Bierlaire, M. (2002). Exploring the potentials of automatically collected gps data for travel behaviour analysis - a swedish data source. In Möltgen, J. and Wytzisk, A., editors, *GI-Technologien für Verkehr und Logistik*, number 13 in IfGIprints, pages 155–179. Institut für Geoinformatik, Universität Münster.

Schönfelder, S. (2003). Between routines and variety seeking: The characteristics of locational choice in daily travel. In *10th Intl Conf on Travel Behaviour*.

Shoval, N. and Isaacson, M. (2006). Tracking tourists in the digital age. In *The Annals of Tourism Research*.

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in 'Journal of Location Based Services' following peer review.

Sohn, T., Li, K. A., Lee, G., Smith, I. E., Scott, J., and Griswold, W. G. (2005). Place-its: A study of location-based reminders on mobile phones. In Beigl, M., Intille, S. S., Rekimoto, J., and Tokuda, H., editors, *Ubicomp*, volume 3660 of Lecture Notes in Computer Science, pages 232–250. Springer.

Wermuth, M., Sommer, C., and Kreitz, M. (2003). Impact of New Technologies in Travel Surveys, chapter 27, pages 455–482. Pergamon, Oxford.

Wolf, J. (2004). Applications of new technologies in travel surveys. In *7th International Conference on Travel Survey Methods* in Los Suenos, Costa Rica.

Wolf, J., Guensler, R., and Bachman, W. (2001). Elimination of the travel diary: an experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record*, (1768):125–134.

Zhou, C., Ludford, P. J., Frankowski, D., and Terveen, L. G. (2005). An experiment in discovering personally meaningful places from location data. In *CHI Extended Abstracts*, pages 2029–2032.