Location based services - a bridge too far for data linkage and behaviour under security and privacy concerns?

Marcus R Wigan, *Oxford Systematics*
Location based services - a bridge too far for data linkage privacy concerns?

Marcus Wigan
Oxford Systematics¹
Heidelberg, Victoria 3084, Australia
Telephone: +61 3 9459 9671
Emails: oxsys@optusnet.com.au; mwigan@unimelb.edu.au; MarcusW@demis.nl

ABSTRACT

Location-based services (LBS) are dependent on a knowledge of a real time location, knowledge of the environment, and integrated with communications. An ideal specification for travel data collection. LBS has become pervasive very swiftly, but the implications are not yet widely recognised. The addition of realtime information, response and service providers to the now familiar combination of GPS, and data recording is the focus of the present paper. The business development path to LBS is outlined, and the implications for data gathering, matching and response considered. The privacy and surveillance aspects are of varying sensitivity in different cultures, even within a single country, but the addition of intelligence methods of data gathering add a further layer to existing concerns. The effectiveness of even limited geospatial tagging to make de-identified data identifiable goes well beyond the methods already emergent for reducing multiple identities in health and other fields to full identification. The substantial potential of LBS to enable improved understanding, monitoring and management of transport provision and movements are clear, but barriers to its wide adoption are outlined in terms of the cultures of authorities collecting data and those of the subjects of that collection.

Introduction

The concept of location based services has been developing rapidly, especially since internet enabled cell phones with GPS facilities (such as the Apple iPhone, the Nokia N95 and now many others) have become widely available. Initially transport data acquisition with GPS (or cell tower triangulation for traffic flows) was done by recording the tracks of individuals or vehicles and doing a post analysis. Later additions were made to customised devices including GPS chips. These were constructed to record user information as specialised equipment for transport data gathering. These activities were in general of restricted application, and constrained to individual data collection studies.

¹ Partner: Volvo Centre of Excellence: the Centre for Governance and Management of Urban Transport, Faculty of Architecture at the University of Melbourne, and Senior Consultant, Demis BV, Delft, the Netherlands.
Commercial developments were also proceeding on the same lines, but at a far higher level of cost and service. Tracing stolen vehicles, monitoring detailed movements of trucks and triggers to detect doors being opened (and other actions such as a vehicle crash) were developed and marketed to freight transport operations and the owners of expensive cars. These methods are now also used to track key staff in real time, spatial and continuous tracing.

The availability of many LBS systems in huge volumes is still expanding, and the issues involved have yet to be worked out with the broader community involved. We have a great incentive to get this right, as it can and will vastly improve our ability to collect data and manage transport and access facilities. The implications of the enhanced surveillance aspects are not fully appreciated.

**How did we get here?**

The US FCC ruling that location be implemented in all cell phones to support emergency location probably accelerated a trend already apparent, ie. to add GPS capacities to existing cell phones. The growth of the mobile internet (GPRS, G3 etc) as a basic cell phone offering added bidirectional communications to this mix-and the infrastructure for LBS become widespread. This was triggered not by the first systems in the field (Nokia’s 90 series internet enabled smart phones) but by the launch of a full data and communications ecosystem by Apple with their iPhone and iTunes applications store concept.

The first paper by the present author on this subject was in late 2008, and only a few examples of services predicted on such a foundation were emerging (Clarke and Wigan, 2008). In the short space of time since then they have become numerous, and in the hands of millions of people.

The acceptability of such tracking depends on many factors, and differs in different cultures. Freight operations is one culture where this has been negotiated many times, first as part of tachograph recording, then for port access and airport security and other specialised situations where the culture implies a requirement to accept locational and action accountability in various concrete forms. Freight supply chains are also increasingly heavy users of minimalist embedded identity systems such as RFID(Radio Frequency Identity) chips attached to devices, palettes, locations – and even people, and used to for location based services in ways that increasingly overlap mobile phone based services and capacities.

In the next section will examine the two major different transport domains:

- Freight and supply chain; and

---


• Person driven.

**LBS Systems**

Location based services are services that are enabled or enhanced by knowledge of the specific location of the enquirer (Schiller and Voisnard, 2004). Inevitably such data based services work in two directions, enabling:

• Active LBS: The LBS user to secure information specific to his or her current location; and:

• Passive LBS: Third party tracing of the locations, times and tracks of the party carrying the LBS device.

The former is an enhancement of transport activities, while the latter is essentially a surveillance mode. Unfortunately the two cannot in most implementations be separated from each other, although some organisations have made real efforts to do so.

Consequently privacy remains a key issue, and may modify the uses made of LBS both by the carrier of the LBS enabled device and third parties wishing to secure the information that is generated by its operation and use.

The present paper addresses some of the issues involved and their interactions with travel behaviour, monitoring and responses.

Privacy has always been an issue, both from the commercial concerns of the LBS carriers and suppliers (see for example Vodafone (2003) guidelines for active and passive services meeting the US FTC’s formal requirements for fair information of notice, choice and access) It is critical to recognize that passive services require considerably more care in their privacy treatment than active, as the user may be unaware of the monitoring involved in passive services.

Examples of active LBS are user-initiated friend finders, restaurant, toilet and points of interest in close proximity, all of which work based on a knowledge of the actual real time location of the user. Passive systems include automatic number plate recognition, road tolling records, CCTV cameras, fleet management, child tracking, stolen car monitoring and recovery. Emergency support services and selective information provision may be location (point or inside a predefined area) or environmentally triggered and are a mix of the active and passive modes.

Almost all of these modes appear in transport operations, movement, monitoring or emergency domains, and location based services are now a reality for the community at large given the advent of the GPS aware application oriented smart phones (Apples
iPhone, RIMs Blackberry etc) The impact of the Apple series of iPhones has been substantial, and added magnetometers (with the recent G3S), GPS and local computing power to make the current generation of smart phones handheld intelligent LBS systems. There are now a very large number of LBS iPhone applications, many of which require data to be shared with the LBS provider, and thus link social networks with marketing databases at very low cost to the LBS provider.

The potential application to transport data collection is obvious, as these phones include GPS, G3 and significant memory and computing power and do not require the special equipment needed for earlier generation GPS based transport data acquisition. However amongst the earliest applications to appear were public transport navigation systems, and GPS position sharing and advertising, with the location awareness proximity information systems on commercials services such a petrol stations and restaurants. Most recently the full turn by turn navigation systems have been added. The TomTom for the iPhone being a good example of this migration of single purpose stand alone special purpose device to the smart phone computing platform.

The back end of such systems is far from simple, and requires a full scale telecommunications operation middleware layer (Jacobsen, 2004) to handle the scale and nature of the spatial queries and their frequency and volume. This has been done by several vendors, and is transparent to the end users – but does potentially allow excellent privacy protocols to form part of the middleware layers. Vodafone (2003) is just one example.

LBS as a whole is considerably larger than the mobile telephone space. Radio Frequency Identity systems (RFID) are rapidly growing as well. These systems rely upon an active or passive interrogation of a specialised computer chip which radiates or responds to a unique identity.

**Common Issues in LBS**

The use of unique identifiers is long established as an effective and useful way of managing large numbers of objects. Bar coding is perhaps the most widely and immediately recognised, but LBS capacities became possible when communications were added to the identifier. Once this is done then large scale databases and communication systems can make full use of the real time (and historical) location data to connect individuals, suppliers, customers and the backgrounds to all of these.

The two very different domains in which LBS operates are:

- Tracking (as part of a supply chain or for people monitoring)
- Communication of facilities near a specific location, activated by presence near it.

Within these two categories a number of systems are immediately recognisable.
In tracking, ANPR (Automated Number Plate Recognition) is indistinguishable from electronic toll charging systems as both collect time, identity and location data which can then be integrated with accounting and enforcement systems for actions to occur. These are charging a specific account, adding the datum of time and location to a traffic or transport database, or even for an immediate action based on a profile of ‘persons or items of interest’.

These are just as much ‘services’ to the community as a whole as provision of a list of restaurants nearby when a mobile phone is used. However the culture is very different. In tracking applications the cultures associated with these two examples are enforcement backed identity detection and recording. These are sensitive issues, but the organisations using these two methods are usually large, and subject to well established rules of operation and accountability. Generally traffic management is perceived to be undertaken for the general good, and such tools are one of the means of enforcing appropriate behaviour on specific individuals or vehicles.

The sensitivities rise when such data and capacities are applied outside the normally expected ranges of application. Mass surveillance is materially assisted by such systems, and the integration of ANPR and electronic tolling into policing and broader uses is not as widely realised or accepted.

The LBS systems emerging from intelligent GPS equipped mobile phones are pitched as a valuable service for individuals. Finding out when the next train will arrive, the nearest Indian restaurant, or the closest toilet (or lavatory, washroom or comfort station depending on the version of English used) are all helpful pieces of selective information that adds value to ones ability to act in a specific location. They are services to individuals, not vehicles or pallets of freight.

This type of LBS is extremely attractive to marketers, as being able to tailor advertisements for services or goods to individuals in the proximity of an outlet exactly at the time that people are adjacent to them is very valuable. The ability to do this does not depend on the individual making a query about the relevant service, it can also be (and is) pushed onto their mobile device without them making a specific request.

This clarifies that there are two very different LBS models for mobile devices carried by individuals.

• Where the individual initiates a request for information; and

• Where a third party identifies the presence of a particular individual and pushes a message or other information to their device unasked

The former is fully in the control of the individual (or at least appears to be, the back up systems providing the information may well retain and use individual data from databanks or other areas), while the latter draws upon all the available data available to the marketing body tailoring the LBS ‘push provision’ as we will now term it.
The reactions of individuals to these very different models of control of information are unsurprisingly rather different – once it is understood what is going on in push provision marketing.

A very useful large scale survey tapping these issues covers US adults of all ages, and the summary below illustrates the nature of the concerns, and the need for greater transparency (at least), and regulation (at best). The cultures relevant here are frequently asserted to be age related, and marketing bodies also assert that customers appreciate tailored messages and advertising. This has recently been tested (Turow et al, 2009).

Contrary to what many marketers claim, most adult Americans (66%) do not want marketers to tailor advertisements to their interests. Moreover, when Americans are informed of three common ways that marketers gather data about people in order to tailor ads, even higher percentages - between 73% and 86% - say they would not want such advertising. Even among young adults, whom advertisers often portray as caring little about information privacy, more than half (55%) of 18-24 years-old do not want tailored advertising. And contrary to consistent assertions of marketers, young adults have as strong an aversion to being followed across websites and offline (for example, in stores) (Turow et al 2009).

The key datum here is the big shift in attitude once information about the actual situation was provided. This applied across all age groups, suggesting that transparency was equally important to all when LBS is involved... and that consumer at all ages knowledge was in very limited about the uses made of LBS information, or the regulations that actually apply.

As transport, pedestrian movement etc are all basic to these LBS services, these findings are important.

Transport and traffic data collection issues

Some of the existing transport and traffic applications are described by Wigan and Clarke (2008), but it is clear that smart mobile phone methods of collection of transport and traffic survey data are extremely valuable, and offer coverage that simply cannot be achieved in any other way. This is a variation of the Tracking mode, of LBS raised earlier, but this time it is a research body that is the controlling party for the information. This is critical, as formal agreements can be made between subject and data collector, so that the subject has full information and control of participation. The latter point is not always true for some research (or rather intelligence) methods, but these are covered later.

Research using GPS data collection process undertaken with the permission of the device carrier usually requires a confidentiality and participation agreement between the carrier and the research body. This is normal practice for transportation and household surveys, and equally appropriate for GPS based work – and for more generalised LBS dependent
Wigan, M.R.  

Location Based Services

7

studies. This research data collection process is transparent and well moderated, and widely acceptable to end users.

The difference is that many LBS approaches used to collect data are indirect, and so the participation of the subject cannot always be secured. It is rarely sought. This data is just as valuable as openly collected transport survey data, but as the parties involved are in many cases marketers the same processes of transparency do not always apply. Traffic data is privacy vulnerable, as much of the existing infrastructure is built and operated by formal traffic monitoring and enforcement bodies. The emergence of mobile phones into this space now raises the privacy issues to a new level, as individuals (rather than vehicles) are the specific objective of phones. Methods have been developed to address this, and automatic traffic monitoring using mobile phones has been addressed by UC Berkeley and Nokia Palo Alto (Hoh et al, 2008).

Hoh et al address the concerns raised by RFID, automated toll collection systems and ANPR systems. Hoh et al recognise that mobile telephone monitoring as a cheaper technique ‘raises significant privacy issues’. Their solution is to defuse the locational sensitivity by creating virtual trip lines, which do not require an assumption that GPS (or mobile equipped) vehicles broadcast their location continuously. The statement:

“We consider sensitive information any information from which the precise location of an individual at a given time can be inferred” (Hoh et al 2009)

One must add that this is not limited to real time, but to any epoch where suitable operational historical data is available. Hoh offers this concise and appropriate statement of the concern and vulnerability that they are addressing. It addresses both real time, anticipatory and ex-post exploitation of such data capture. At a higher level, their approach is typical of the culture of research/university responsibility for data and subjects, somewhat in contrast to the marketing ethos.

The key issue in Hoh et al’s treatment of privacy in transportation data collection is to ensure that the system is designed from the start such that the privacy sensitive data simply is not acquired by the system, and, one must add, can be shown not to (meeting the criteria first defined by Webber et al (1990).

The methods of identifying precise location using mobile phones do not depend solely on GPS, or indeed just on mobile phone cell tower triangulation. Combinations of these methods and dead reckoning and extended Kalman filtering can provide excellent results especially when combined with map matching to constrain the location space within which a vehicle might be moving (see, for example, Ochieng et al, 2004; Taylor and Blewitt, 2006).

The impact of cumulative anonymised records

Next generation LBS Systems and Augmented Reality
The next generation of LBS systems will include augmented reality, where LBS systems will provide data attached to locations and views in real time. Such systems are already being developed for organising image capture. The Apple iTunes Geolocation feature enabled by the GPS in iPhones is just one of many large scale examples of such large scale data acquisition systems in broad consumer use.

Adding other attributes as one views (or approaches on foot or other means) buildings and other locations is Augmented Reality\(^4\), and examples are beginning to emerge (just one example is given by Hollere et al (2007).

One version of Augmented reality (from Georgia Tech: Collerton (2009)), has already surfaced and drawn privacy concerns by combining micro-monitoring of human movements and virtual maps in a Google context. There will be more.

Progressive developments in LBS and marketing will lead to further merging of locational and individual data bases. This will enhance both what can be delivered to consumers, and the downsides of such universal monitoring.

Once again the price of having such highly integrated data available makes the GIS location issue the more important to address. Few if any concerns have yet been expressed about Augmented reality as have begun to arise in LBS plans.

**Individual Identification from multiple de-identified sources**

Locational movement and timing data is so valuable for both ITS and transport and activity surveys that there is pressure to collect, acquire or secure large scale databases of this type for their value in travel and activity analysis as well as the real time contributions possible with this type of data. However there are now rapidly improving techniques for combining different types of larger scale aggregate anonymised data to identity individuals (eg. Sweeney (2001a).

In the health and genetic area this has major implications if achieved at any time, and concerted efforts\(^5\) are under way to create functional access management regimes to protect identity of the individuals concerned for the benefits that such large scale health data can provide.

The ability of even very limited geospatial data added to otherwise anonymous records has been documented for some time. The general methods for protection have been well reviewed by Sweeney (2001b) who demonstrated that simply adding US Zip Code she could identify a substantial fraction of the populations covered in other databases. Similarly, as cited in Emam (2009):

---

\(^4\) A good source of background is at [http://www.augmented-reality.org/ismar/](http://www.augmented-reality.org/ismar/)

\(^5\) For example the efforts at the University of Texas in this field for Google: see [http://www.utexas.edu/features/2009/10/12/cybersecurity](http://www.utexas.edu/features/2009/10/12/cybersecurity)
“An expert witness was able to re-identify with certainty 18 out of 20 individuals in a neuroblastoma dataset from the Illinois cancer registry, and was able to suggest one of two alternative names for the remaining two individuals”

This is drawn from legal reports on cases in Illinois, and shows that the issues of anonymity and de-identification are now contestable and becoming the subject of case law in the US\(^6\).

The reported results that none were identified incorrectly is a highly significant aspect of this work. Usually Type 1 and Type 2 errors lead to high personal costs due to inaccurate identification, and the signs that such errors are not frequent when using this group of techniques that make them peculiarly suitable for forensic or enforcement applications.

This is likely to become a major issue as locational data and transport and activity data build up, albeit in separate silos as enforcement is a major issue for ITS operators of all kinds, from toll road operators to car manufacturers (where records are kept in the engine management chips and increasingly other measures for security, tracing, recovery and ITS operations will expand this range – and registration and warranty records ensure that de-identification and protection of anonymity will soon become difficult indeed.

Transport is implicitly a locational based service \textit{par excellance}, and so these accumulations of aggregate data – in good faith de-identified by the creators and holders before permitting wider use - require at least as careful attention as the health records already being considered for higher levels of protection in anonymisation and access techniques, and arguably at least as sensitive once in use by wider groups.

Transport data users deploying or using ITS, LBS and household interview records may not yet be fully aware of these developments.

\textbf{Legal frameworks}

An alternate responsible culture is demonstrated by the European Commission (2002), which focuses on user empowerment and transparency for ‘push’ messages:

\begin{quote}
\textit{“Automated calling is only allowed in respect of subscribers who have given their prior consent”} (extracted from Directive 58, article 13 of EC (2002)), and even more specifically
\end{quote}

\begin{quote}
\textit{“if the operator wants to do direct marketing, then the user must be given the opportunity to object, free of charge and in an easy manner, to the use of his or her contact data. This opportunity must be given at each message”}.
\end{quote}

---

This explicitly rejects the presumption of agreement, and in Article 9 extends to cover the critical hidden background issues of exploitation of location based data as follows “Location based data may only be processed when it is made anonymous or with the consent of the user for the duration necessary for the provision of a service”. However the Directive does not cover the use of prior information to target individuals arriving at a specific location or on a specific transport mode. Such targeting may be done using specific or Bayesian or associative methods of identification, and these more subtle aspects of the EC approach are shared by all the previous approaches.

Efforts to anticipate the privacy issues raised by LBS were also made in the US, with an unsuccessful bill introduced to and debated in Congress in 2001, which would have required the FCC to regulate as follows:

(A) require providers of location-based services and applications to inform customers, with clear and conspicuous notice, about their policies on the collection, use, disclosure of, retention of, and access to customer location information;
(B) require providers of location-based services and applications to obtain a customer's express authorization before--
(i) collecting, using, or retaining the customer's location information; or
(ii) disclosing or permitting access to the customer's location information to any person who is not a party to, or who is not necessary to the performance of, the service contract between the customer and such provider;
(C) require that all providers of location-based services or applications--
(i) restrict any collection, use, disclosure of, retention of, and access to customer location information to the specific purpose that is the subject of the express authorization of the customer concerned; and
(ii) not subsequently release a customer's location information for any purpose beyond the purpose for which the customer provided express authorization;
(D) ensure the security and integrity of location data, and give customers reasonable access to their location data for purposes of verifying the accuracy of, or deleting, such data;
(E) be technology neutral to ensure uniform privacy rules and expectations and provide the framework for fair competition among similar services;
(F) require that aggregated location information not be disaggregated through any means into individual location information for any commercial purpose; and
(G) not impede customers from readily utilizing location-based services or applications.

(2) PERMITTED USES.--The rules prescribed under subsection (a) may permit the collection, use, retention, disclosure of, or access to a customer's location information without prior notice or consent to the extent necessary to--
(A) provide the service from which such information is derived, or to provide the location based service that the customer is accessing;
(B) initiate, render, bill, and collect for the location-based service or application;
(C) protect the rights or property of the provider of the location-based service or application, or protect customers of the service or application from fraudulent, abusive, or unlawful use of, or subscription to, the service or application;
(D) produce aggregate location information; and
(E) comply with an appropriate court order.

Perhaps the most significant aspect of the draft Bill reads as follows:

(3) ADDITIONAL REQUIREMENT.--Under the rules prescribed under subsection (a), any third party receiving, or receiving access to, a customer's location information from a provider of location services or applications pursuant to the express authorization of the customer, shall not disclose or permit access to such information to any other person without the express authorization of the customer.

(4) EXPRESS AUTHORIZATION.--
(A) FORM.--For purposes of the rules .......... the Commission shall specify the appropriate methods, whether technological or otherwise, by which a customer may provide express prior authorization. Such methods may include a written or electronically signed service agreement or other contractual instrument.

The overall effect of this Bill, had it been passed (which it was not) would have been to provide a clear basis for locational privacy, and subject control over the further aggregation and utilisation of the LBS data with other data, or along a data sales chain.

Interestingly it would not have limited ITS services significantly, simply constrained the use of the data outside the space in which it had been explicitly accepted that it be used. This would have enhanced LBS services, at the cost of the further aggregation of real time and historical locational data being progressively added to mass data surveillance and marketing data banks and data mining.

The EU Directive is not and will not be as effective in addressing the privacy and data aggregation issues, most of which remain in most countries.

The behavioural effects of such cumulative data monitoring are not yet clear, but the first indications are negative in terms of acceptance of the outcomes. The very valuable aspects of LBS systems (which one must note include many ITS systems) may be prejudiced if not soon addressed.

Harbingers include the progressive series of unexpected negative outcomes from the large amounts of private data posted on Facebook and MySpace, but the far more powerful association of detailed locational data (often in real time) has yet to be experienced by the wider publics.

The large gap between expectation and reality in terms of data use and legal backing that emerged in the Turow (2009) indicates that such warnings are likely to be of limited value – until the negative events start to arise.

It would clearly be desirable for the development and maximum proper use of LBS that a better legal framework and data accountability framework be put in place soon, preferably before these events start to reach general public awareness.
Intelligence methods applied to LBS provision and data collection

Individuals can readily be re-identified in anonymised data when GIS data is included. This is well understood in the GIS world (Spiekerman, 2004). Only a minimal amount of additional data is required to re-identify individuals, and this does (or should) constrains the wider use of household travel survey data, for example. However this only addresses concrete links between the individuals and the data related to them.

The different professional and regulatory cultures addressed do not and cannot deal with the intrusion of formal investigative intelligence styles of targeting of objects (people or vehicles) of interest (Wigan and Clarke, 2008). The issue here is that spatial and temporal proximity are key features in narrowing the range of persons of interest in intelligence gathering. The fact that particular vehicles are observed in the same locations and same times is already in use in the Uk exploiting mass collection of ANPR records.

The collection of LBS data not only allows surveillance by means of identification of people habituating either or both of the same locations and times, but this form of purely data analysis driven surveillance is not subject to detection - or even participation - of the parties involved.

The unsuccessful US Bill addressed these issues in a way that the EU Directive did not, and there are few other examples of regulatory efforts to manage the cumulative integration of Locational data with other sources of information about individuals.

Conclusion

The different cultures involved in using, providing, exploiting and regulating LBS have been explored with special reference to transport applications. These cultures and approaches differ depending on the position in the power hierarchy from consumer, through technical provider and marketer to legislative bodies.

As is common in such cases the privacy and overall implications of such a new technical capacity are somewhat ahead of legislative responses, and indeed community perceptions of the need to engage regulatory bodies at an early stage..

However the combination of LBS and mass data aggregation and mining systems change the nature of the daily range of previously unrecorded activities to change the basis of one’s identity to:

‘You are where you have been and .... with whom, and when’ (Wigan and Clarke, 2008)

How much will this affect travel and ITS related behaviour? Or will the slow expansion of the surveillance state have reduced user sensitivities by the time that the negative effects become more widely felt?
Location Based services offer a great deal: to individuals, marketers, enforcement bodies transport and traffic researchers and the community. What is needed to secure these gains against to downsides discussed?

Legal frameworks have been proposed, but community attitudes will have the final say. We can only hope that the excellent data collection and service provision potentials will remain in future to be reaped responsibly.

The bridge between expanded locational data acquisition and provision, often unintended, and privacy and surveillance has certainly not yet been bridged. While many transport data collection systems obey sound ethical rules for collection, use and onward transfer of location specific personal data, this is not the case in many other areas.

Is the gap too large to bridge before the momentum of large scale data integration processes moves on to close the commercial and surveillance gaps?

If asked, we would estimate that as of now the answer is: Probably yes.

References

Blumberg, A.J. and Eckersley (2009) On locational privacy and how to avoid losing it. Electronic Frontiers Foundation, August 7pp at


