On Locational Privacy, and How to Avoid Losing it Forever
Related Issues: Locational Privacy, Privacy

August 2009

By Andrew J. Blumberg and Peter Eckersley, August 2009
PDF file

Also available as a PDF
in English and Bulgarian.

Over the next decade, systems which create and store digital records of people's movements through public space will be woven inextricably into the fabric of everyday life. We are already starting to see such systems now, and there will be many more in the near future.

Here are some examples you might already have used or read about:

* Monthly transit swipe-cards
* Electronic tolling devices (FastTrak, EZpass, congestion pricing)
* Cellphones
* Services telling you when your friends are nearby
* Searches on your PDA for services and businesses near your current location
* Free Wi-Fi with ads for businesses near the network access point you're using
* Electronic swipe cards for doors
* Parking meters you can call to add money to, and which send you a text message when your time is running out

These systems are marvellously innovative, and they promise benefits ranging from increased convenience to transformative new kinds of social interaction.

Unfortunately, these systems pose a dramatic threat to locational privacy.
What is "locational privacy"?

Locational privacy (also known as "location privacy") is the ability of an individual to move in public space with the expectation that under normal circumstances their location will not be systematically and secretly recorded for later use. The systems discusssed above have the potential to strip away locational privacy from individuals, making it possible for others to ask (and answer) the following sorts of questions by consulting the location databases:

* Did you go to an anti-war rally on Tuesday?
* A small meeting to plan the rally the week before?
* At the house of one "Bob Jackson"?
* Did you walk into an abortion clinic?
* Did you see an AIDS counselor?
* Have you been checking into a motel at lunchtimes?
* Why was your secretary with you?
* Did you skip lunch to pitch a new invention to a VC? Which one?
* Were you the person who anonymously tipped off safety regulators about the rusty machines?

* Did you and your VP for sales meet with ACME Ltd on Monday?
* Which church do you attend? Which mosque? Which gay bars?
* Who is my ex-girlfriend going to dinner with?

Of course, when you leave your home you sacrifice some privacy. Someone might see you enter the clinic on Market Street, or notice that you and your secretary left the Hilton Gardens Inn together. Furthermore, in the world of ten years ago, all of this information could be obtained by people who didn't like you or didn't trust you.

But obtaining this information used to be expensive. Your enemies could hire a guy in a trench coat to follow you around,but they had to pay him. Moreover, it was hard to keep the surveillance secret — you had a good chance of noticing your tail ducking into an alley.

In the world of today and tomorrow, this information is quietly collected by ubiquitous devices and applications, and available for analysis to many parties who can query, buy or subpeona it. Or pay a hacker to steal a copy of everyone's location history.

It is this transformation to a regime in which information about your location is collected pervasively, silently, and cheaply that we're worried about.

Threats and opportunity

Some threats to locational privacy are overt: it's evident how cameras backed by face-recognition software could be misused to track people and record their movements. In this document, we're primarily concerned with threats to locational privacy that arise as a hidden side-effect of clearly useful location-based services.

We can't stop the cascade of new location-based digital services. Nor would we want to — the benefits they offer are impressive. What urgently needs to change is that these systems need to be built with privacy as part of their original design. We can't afford to have pervasive surveillance technology built into our electronic civic infrastructure by accident. We have the opportunity now to ensure that these dangers are averted.

Our contention is that the easiest and best solution to the locational privacy problem is to build systems which don't collect the data in the first place. This sounds like an impossible requirement (how do we tell you when your friends are nearby without knowing where you and your friends are?) but in fact as we discuss below it is a reasonable objective that can be achieved with modern cryptographic techniques.

Modern cryptography actually allows civic data processing systems to be designed with a whole spectrum of privacy policies: ranging from complete anonymity to limited anonymity to support law enforcement. But we need to ensure that systems aren't being built right at the zero-privacy, everything-is-recorded end of that spectrum, simply because that's the path of easiest implementation.

Location Based Services That Don't Know Where You Are

Surprisingly, modern cryptography offers some really clever ways to deploy road tolls and transit tickets and location searches and all the other mobile services we want, without creating a record of where you are. This isn't at all intuitive, but it's really important that policymakers and engineers working with location systems know about it. This section lists just a few examples of the kinds of systems that are possible.

Automated tolling and stoplight enforcement

In many metropolitan areas, drivers are encouraged to use small electronic transponders (FastTrak, EZpass) to pay tolls at bridges and tunnels. As momentum builds behind nuanced usage tolling and congestion pricing schemes, we expect to see an explosion of such devices and tolling methods.

For simple point tolls (e.g. bridge tolls), protocols that cryptographers call electronic cash are an excellent solution. In its cryptographic sense, electronic cash refers to means by which an individual can pay for something using a special digital signature which is anonymous but which guarantees the recipient that the can redeem it for money; it acts just like cash! See this paper for the details of a modern implementation. Thus, a driver "Vera" would buy a wad of electronic cash every few months and "charge up" her transponder. As Vera drives over bridges and through tunnels, the tolling transponder would anonymously pay her tolls.

For more complicated tolling systems (in which the price depends on the specific path taken), a somewhat more involved implementation can be used (discussed in detail in this technical paper).

Straightforward but privacy-insensitive implementations of congestion-pricing systems simply track drivers and use the tracking information to generate tolls. For instance, you might have all of the cars using a little radio gadget to report their location all the time. As Vera drives throughout the congestion pricing area (e.g. down a street in central London), the gadget says "Hi, this is Vera's car." That creates a record of everywhere Vera went. Equivalently, one might put cameras everywhere which record Vera's license plate as she drives and keeps track of everywhere she goes to subsequently compute his tolls. Both of these solutions violate Vera's locational privacy.

The less obvious but much better way to run such tolls is to have Vera's gadget commit to a secret list of "dynamic license plates" — a long list of random-looking cryptographic numbers. This commitment takes the form of a digital signature given to the tolling authority. As Vera drives through the tolling region, her gadget cycles through these numbers rapidly, sending the current number to the monitoring devices she passes. None of those numbers actually identifies Vera, and since they keep changing there's no way to string them together to track her.

But, at the end of the month, Vera has to pay her road toll by plugging the gadget in her car into her computer. The computers execute a fancy cryptographic process called a "secure multi-party communication". At the end, her computer proves that she owes $17.00 in road tolls this month, without revealing how she acumulated that total. The committment exchanged at the beginning ensures that Vera can't cheat: she can't prove a lower total if she actually drove across a bridge with the gadget active.

This kind of approach can be used to solve various automated traffic enforcement needs, as well. For instance, every time Vera passes a traffic light a monitoring device can collect the current "dynamic license plate". Although again, the collected data can't be used to track Vera around, if Vera runs a red light the system can detect this and issue Vera a ticket.

Location-based search

A location-based search on a mobile device is another important example. Phones are starting to be able to locate themselves based on the signal strength or visibility of nearby wireless networks or on GPS data. Naturally, companies are also racing to provide search tools which use this data to offer people different search results depending on where they are at any given moment. The naive way to do

mobile location search is for the device to say "This is Frank's Nokia here. I see the following five WiFi networks with the following five signal strengths". The service replies "okay, that means you're at the corner of 5th and Main in Springfield". Then your device replies, "What burger joints are nearby? Are any of Frank's friends hanging out nearby?". That kind of search creates a record of everywhere you go and what you're searching for while you're there.

A better way to do location-based services and search is something like this: "Hi, this is a mobile device here. Here is a cryptographic proof that I have an account on your service and I'm not a spammer. I see the following five wireless networks." The service replies "okay, that means you're at the corner of 5th and Main in Springfield. Here is a big list of encrypted information about things that are nearby". If any of that encrypted information is a note from one of Frank's friends, saying "hey, I'm here", then his Nokia will be able to read it. If he likes, he can also say "hey, here's an encrypted note to post for other people who are nearby". If any of them are his friends, they'll be able to read it. (An excellent and detailed discussion of a related approach via secure multi-party computation is presented in this paper.)

Transit passes and access cards

Another broad area of application is for passcards and devices allowing access to protected areas; for instance, passcards which allow access to bike lockers near train stations, or cards which function as a monthly bus pass. A simple implementation might involve an RFID card reporting that Bob has checked his bike into or out of the storage facility (and deducts his account accordingly), or equivalently that Bob has stepped onto the bus (and checks to make sure Bob has paid for his pass). This sort of scheme might put Bob at risk.

A better approach would involve the use of recent work on anonymous credentials. These give Bob a special set of digital signatures with which he can prove that he is entitled to enter the bike locker (i.e. prove you're a paying customer) or get on the bus. But the protocols are such that these interactions can't be linked to him specifically and moreover repeated accesses can't be correlated with one another. That is, the bike locker knows that someone authorized to enter has come by, but it can't tell who it was, and it can't tell when this individual last came by. Combined with electronic cash, there are a wide-range of card-access solutions which preserves locational privacy.

Privacy concerns and anonymized databases

We should note that even the existence of location databases stripped of identifying tags can leak information. For instance, if I know that Vera is the only person who lives on Dead End Lane, the datum that someone used a location-based service on Dead End Lane can be reasonably linked to Vera. This problem is widely acknowledged (and studied) in the context of epidemiological data as well: it turns out to be relatively easy to deduce the identity of individual disease victims from "anonymized" geographic information about the location of cases. Generally speaking, one solution to this problem is to restrict the use of location-based services to high density areas. There are more complicated cryptographic solutions that are also possible. See this paper for a discussion (and proposed solution) to this problem in the context of collection of aggregate traffic statistics, and this paper for discussion of "differential privacy", a formalization of ideal privacy guarantees in the face of the existence of databases.

For more information

Safely and correctly implementing such modern cryptographic protocols can be a substantial engineering challenge. And implementing them efficiently takes work. But it can be done — this is exactly the kind of cryptographic software that protects the security of our financial network (e.g.

ATMs), makes it safe for us to buy things online, and encodes our phone calls. Big software contractors (e.g. IBM and Siemens) maintain large staffs of cryptographers.

We've linked to some of the sources that would be useful for engineers who want to understand how these protocols work. But, if you're a policymaker or an engineer and you have questions about how these methods work, don't hesitate to contact us: we can point you at literature and connect you with experts to answer your questions.

Why Should Private Sector Firms Prioritize Locational Privacy?

We believe that governments have a civic responsibility to their citizens to ensure that the infrastructure they deploy protects locational privacy. But there are also financial reasons for the private sector to go to some length to design privacy into the locational systems they build.

Avoid legal compliance costs

If a corporation retains logs that track individuals' locations, they may be subject to legal requests for that information. Such requests may come in different forms (including informal questions, subpoenas or warrants) and from different parties (law enforcement or civil litigants). There are complex legal questions as to whether compliance with a particular request is legally required, optional, or even legally prohibited and a liability risk.

This legal complexity may even involve international law. For instance, US corporations which also have operations in the European Union might be subject to European data protection laws when EU citizens visit the United States and use the US company's services.

Corporations with large locational datasets face a risk that lawyers and law enforcement will realize the data exists and begin using legal processes to obtain it. The best way to avoid this costly compliance risk is to avoid having identifiable location data in the first place.

Obtain a competitive edge

The public is slowly becoming aware of the potential downsides of having their location tracked on a continuous basis. The ability to demonstrate reliable privacy protections will increasingly offer firms a competitive edge if they can persuade individual customers — or government clients — that their product offers more robust and trustworthy privacy protections.

Isn't there an easier/different alternative?

Using cryptography and careful design to protect location privacy from the outset requires engineering effort. So it's important to ask whether there are other adequate ways to preserve privacy in these systems. Unfortunately, we believe the alternatives are unreliable or harder to implement and enforce.

Data retention and erasure

One kind of protection you might hope for is that your location records will be deleted before your adversary gets to them. If the company that's offering you a fancy location search on your cell phone doesn't need to remember your history a week later, perhaps they can be persuaded to forget it quickly. Perhaps they promise that they will.

Unfortunately, there isn't much basis for optimism on the data retention front. Search companies have incentives to keep extensive records of their users' queries, so that they can learn how to improve their results (and sell more effective advertisements). Storage space is cheap and getting cheaper. Tolling agencies have incentives to keep extensive records of toll usage, to settle complaints and provide

aggregate statistics and accounting data.

Even if the collecting outfit does promise to delete the data after a set interval, there's no guarantee that they're actually going to do that properly. Firstly, secure deletion tools are necessary to make sure that deleted data is really gone; many sys admins will fail to use them correctly. Secondly, all it takes is the flip of a switch to suddenly change policies from deletion to retention. To make matters worse, there's no guarantee that a government won't suddenly pass a law requiring such companies and government agencies to keep all of their records for years, just in case the records are needed for "national security" purposes. This last concern isn't just idle paranoia: this has already happened in Europe, and the Bush administration has toyed with the same idea.

And as for government agencies, experience so far with data retention has not been reassuring. An interesting example is provided by automated tolling data (records from FastTrak and EZpass). Different states have made different promises about how long they keep the data, and there have been varying degrees of effectiveness in carrying out these promises. Data has often remained available for subpeona after a number of years. Legal penalties for the violation of these promises are currently minimal.

Limiting data retention is an important protection for privacy, but it's no substitute for the best protection: not recording that information in the first place.

Opting out

Sometimes people respond to these sorts of worries with the claim that the free market will solve this problem. "People who are worried about privacy shouldn't use these services," they say. "If people really care, a company offering privacy as an explicit feature will arise."

We don't believe this is an acceptable viewpoint — there is too much coercion in play. Often, there's no adequate replacement for the service in question, and it is or will soon be a dramatic hardship to avoid its use. Suppose that parts of the United States began to adopt mandatory "pay as you drive" insurance, or congestion pricing, that was based on location tracking. In most parts of the United States, it's not really reasonable to suggest that people who are worried about privacy shouldn't drive (or shouldn't drive to their religious institution of choice). And in the case of location-based services, it's clear that the deck is stacked against people choosing to take inconvenient measures to protect themselves: it's too hard to know what is being recorded by whom, too hard to know what options there are to avoid being recorded, and too hard to keep researching these questions as you interact with new pieces of technology. In this environment, people simply haven't adjusted to the potential for the loss of the reasonable expectation of privacy in public places, and our standard intuitions haven't kept up with advances in technology.

Cell phones and credit cards already create a trail

It's true that most cell phones provide some amount of tracking information to the carriers as long as they're on, and that credit card records provide a pervasive trail of activity. This is no reason to surrender further locational privacy, but rather a reason to fight for better practices or laws for cell phone technology and credit card data. The problems we're having now with identity theft make it clear how problematic the handling of sensitive personal data is.

Law-abiding citizens don't need privacy

Another common response to worries about locational privacy is to say that law-abiding citizens don't need privacy. "I don't commit adultery, I don't break the law," people say (and tacitly, "I'm not in the

closet, and I don't belong to any non-majority religious or political groups").

One answer to this concern is a reminder that there are more subtle reasons for needing privacy. It's not just the government, or law enforcement, or political enemies you might want to be protected from.

* Your employer doesn't need to know things about whether, when, and where you went to church.
* Your co-workers don't need to know how late you work or where you shop.
* Your sister's ex-boyfriend doesn't need know how often she spends the night at her new boyfriend's apartment.
* Your corporate competitors don't need to know who your salespeople are talking to.

Preserving locational privacy is about maintaining dignity and confidence as you move through the world. Locational privacy is also about knowing when other people know things about you, and being able to tell when they are making decisions based on those facts.

Suppose that an insurance company manages to obtain a record of Alice's movements over the past year, and decides that there is some aspect of that record which is grounds for raising her premiums or denying her coverage. The problem with that decision is not just that it is unfair, but that Alice may have no ability to dispute it. If the insurance company's reasoning is misinformed, will Alice have a practical way of knowing that and disputing it?

The "I've got nothing to hide" argument against privacy is criticized at greater length in this article.

Conclusion

In the long run, the decision about when we retain our location privacy (and the limited circumstances under which we will surrender it) should be set by democratic action and lawmaking. Now is a key moment for organizations that are building and deploying location data infrastructure to show leadership and select designs that are responsible and do not surrender the locational privacy of users simply for expediency.