27th April, 2010

Copy of Google's submission today to several national data protection authorities on vehicle-based collection of wifi data for use in Google location based services

Summary

Google's location based services rely on a variety of signals to attempt to provide the user with an approximate location.

These signals include GPS, the location of cell towers and the location of wifi access points.

Google collects information about cell towers and wifi access points in a variety of ways, including using information provided by cellphone handsets and computer applications, as well as radio receivers attached to vehicles.

For the sake of ease Google uses the same vehicles which collect imagery for the Street View service to collect information about wifi access points. This is, however, a separate product from Google Street View, and the only connection between Street View and Google location services is the shared use of the physical vehicle and equipment for collecting data.

Data about wifi access points is collected passively, and is only that which is broadcast publicly on the wifi radio network. It is visible to anyone else with a wifi receiver - including other users with wifi enabled devices like laptop computers and smartphones.

The data is collected in aggregate by Google; it is not tied to any particular user. The data is used in aggregate to improve Google's location based services and is not shared directly with users. Applications that return information to users of Google's location based services only receive geocoded locations that are individual to the user's request - they do not receive specific information about an access point.

The operators of access points may personalise the data that they broadcast (e.g. by changing the default non-personal SSID to one which contains a full name), and in doing so they choose to broadcast this data to the public domain to be received by any wifi-enabled equipment in range. The SSID is not, however, disclosed in Google's location based services.

Google's collection of location relevant information

Google provides location based services to users of a number of its products, and to the operators and users of other products and services through the free publicly available Geolocation API.

Location based services rely on a variety of location indicators associated with the user's device to help identify the device's location. For example, GPS enabled devices can provide a highly accurate geocoded location using information from GPS satellites. However, many devices are not GPS enabled, and/or are used in situations where obtaining a GPS signal will take too long, or might not even be possible (e.g. indoors, where there is no line of sight between the device and the satellites). Therefore other location indicators are often used to help locate the user's device, albeit perhaps not to the same level of accuracy as GPS.

These include proximity to cell towers and proximity to wifi access points.

To be able to use these location indicators Google has collected information to help create a database of known cell tower and wifi access points. This information is collected in a variety of ways, including using information provided by cellphone handsets and computer applications, as well as information collected by wifi radio receivers. Google is not the only company to collect such data - global companies like Skyhook have collected similar data in many countries around the world, and even just within Germany many organisations are involved in similar projects, including Gammax Systems GmbH, Magic Map, IT2media and the Fraunhofer Institute.

As part of Google's efforts it operates an internal project which uses wifi radio receivers attached to vehicles to collect wifi access point information in areas which the vehicle passes.

As a matter of practicality, the equipment is fitted to the same vehicles which are used to collect Street View imagery and 3d laser data for use in products like Google Maps. However, the collection of wifi access point information is not part of the Street View product, and the only association between location based services and Street View is the shared use of vehicles for practical and cost purposes.

Our collection of wifi access point information is widely known, indeed it is on Wikipedia and has featured in articles in the New York Times amongst others. We have also provided a wealth of information about our location based services and collection of location indicator data on our various blogs. e.g. My Location feature on Maps for mobile, My Location on the desktop, Gears geolocation API

How does the technology work?

Visibly attached to the roof of each vehicle is a commercially available Maxrad BMMG24005 omnidirectional radio antenna. This antennae receives publicly broadcast wifi radio signals within range of the vehicle.

The vehicle travels at normal road speeds, and so spends only a very short amount of time within the range of any given wifi access point.

The signals are initially processed onboard in the car, using software including the standard Kismet open source application. The data is then further processed when transferred to servers within a Google Data Centre, and used to compile the Google location based services database.

The equipment within the vehicle operates passively, receiving signals broadcast to it but not actively seeking or initiating a communication with the access point.

The information visible to the equipment is that which is publicly broadcast over the radio network, using the 802.11 standard. This includes the 802.11b/g/n protocols.

The equipment is able to receive data from all broadcast frames. This includes, from the header data, SSID and MAC addresses. However, all data payload from data frames are discarded, so Google never collects the content of any communications. In addition, the operator of the access point can choose to restrict the SSID from broadcast, and in many cases this will mean that the SSID is not received (although this may vary depending on the way the access point broadcasts data).

The equipment also separately records the signal strength and channel of the broadcast at the point at which it was received by our equipment, and is able to establish the protocol used (i.e. 802.11b/g/n).

It is possible to identify from the data received if an access point is encrypted - this may be included in the data sent in the frame header but in any event will be self-evident from the presence of encryption within the frames generally. However, while the information within the data frames will always reliably indicate to us if an access point is encrypted, we cannot reliably determine whether an access point *is not* encrypted. For example the data packet received by our equipment could be truncated or corrupted, meaning that we do not see the use of encryption within the broadcast. This does not, however, mean that the network is not encrypted - merely that we did not receive enough data to establish whether encryption was used or not. Aside from the use of encryption, the operator of an access point may apply other higher level access controls to restrict access to the network, which we are not able to detect. We can therefore never reliably determine if an access point is ‘open’.

How do we use wifi access point data?

The data which we collect is used to provide location based services within Google products and to users of the Geolocation API. For example, users of Google Maps for Mobile can turn on “My Location” to identify their approximate location based on cell towers and wifi access points which are visible to their device. Similarly, users of sites like Twitter can use location based services to add a geolocation to give greater context to their messages.

Google currently uses 2 pieces of the data collected during the driving operation to build its database and provide location based services - the MAC address of the access point and the GPS co-ordinates of the vehicle at the point at which the access point was visible. This data is stored in aggregate form, and is used to provide the location based service.

Google location based services using wifi access point data work as follows:

- The user’s device sends a request to the Google location server with a list of MAC addresses which are currently visible to the device;
- The location server compares the MAC addresses seen by the user’s device with its list of known MAC addresses, and identifies associated geocoded locations (i.e. latitude /

longitude);
- The location server then uses the geocoded locations associated with visible MAC address to triangulate the approximate location of the user;
- This approximate location is geocoded and sent back to the user’s device.

The only data which Google discloses is a triangulated geocode which is an approximate location of the user’s device. At no point does Google publicly disclose MAC addresses from its database (in contrast with some other providers in Germany and elsewhere).

There has been speculation that Google will make available a map or list of wifi access points, including identifying the SSID of each access point and/or identifying those which are open. This is not a service which Google provides.

Raphael Leiteritz, Product Manager, Google