Insights from Wi-Fi data

Proposed Pilot

Contents

Background

In 2014, TfL approached the Information Commissioners Office (ICO) to discuss using data generated from the TfL in-station Wi-Fi access points. The aim of using this data was to improve both the operation of the London Underground network and provide better customer information. TfL conducted a Privacy Impact Assessment as part of this to identify and reduce any risks or privacy concerns this could have generated. Whilst Wi-Fi data was considered initially, the proof of concept, which assessed passenger crowding on the Victoria Line was delivered using train telemetry data, Rolling Origin Destination Survey (RODS) data and Oyster smartcard ticketing data instead of Wi-Fi data.

The purpose of this document is for TfL to engage the ICO as we are now planning to run a pilot using Wi-Fi data on a subset of the London Underground network. We propose collecting, hashing and analysing the media access control (MAC) address [hereafter referred to as 'Wi-Fi data'] of customers connecting to selected Wi-Fi access points to determine the value of this data against specified use cases. A new privacy impact assessment will be completed as part of this.

Use Cases and benefits of analysing Wi-Fi data

Wi-Fi data has the potential to provide TfL with a far greater understanding of customer behaviour on the London Underground network enabling TfL to improve the operation, planning and information provision on London Underground. Four key use cases have been identified where the collection, processing and analysis of Wi-Fi data could benefit both TfL and our customers:

1. **Customer Information**: TfL could provide historical and near real time information to customers informing them of on-train, platform and station crowding and advise customers of the best route and travel options. Where disruptions and delays occur, patterns of travel from Wi-Fi data could be used to identify customers with similar patterns in our ticketing data enabling us to deliver more efficient and effective refunds to customers impacted.

2. **Medium and long term transport network planning**: Currently TfL carries out the Rolling Origin Destination Survey (RODS) to understand passenger journeys, route choices and movements. The survey underpins organisation wide planning decisions ranging from timetables, to train assets, and the priority and design of station upgrades. RODS incurs the same constraints and bias inherent in any self-completion survey; it is only carried out with a small sample of customers, at a small sample of stations and in a short time period limiting its usefulness in capturing the variability and dynamic nature of the network and customers movements within in it. By using Wi-Fi data, merged with aggregated Oyster and Contactless ticketing data we would have a far richer data source to ensure optimal and evidence based decision making for a wide range of planning decisions. This would also enable financial savings through reduced survey, processing and modelling costs.

3. **Operations and Safety Information**: Understanding our customer's behaviour is vital for day-to-day, event and disruption management of the Underground network. Through collecting and processing Wi-Fi data we would be better able to operate our network and to deploy staff more effectively to best meet customer needs; ensuring a safe environment for all who use the network.

4. **Financial**: TfL is under increasing financial pressure. The Department of Transport grant we receive (£591m in 2015/16) will be removed from 2018. In addition, fares are to be frozen over the current mayoral term (2016 to 2020). To ensure we continue to provide a world

class transport network we will need to spend less and generate more. One area for generating more is the commercial advertising partnership for in-station advertising assets. By understanding customer behaviour in relation to these assets we will be able to significantly increase the revenue we generate.

Pilot Proposal

Informing Customers

We will communicate to our customers and the public that we are collecting and processing Wi-Fi data on a subset of our network. The proposed channels detailed below are methods we will consider using to inform as many customers as practicably possible.

1. In station signage and notices,
2. In station announcements,
3. Press release(s),
4. Newspaper articles or advertisements,
5. Dedicated TfL website page,
6. Emailing registered TfL customers.

Customers will be advised that they can opt out by turning off the Wi-Fi capability of their device.

Informing Employees

Currently station staff frequently use Wi-Fi on mobile work devices (phones and tablets). This enables TfL to provide staff with the latest information and easy to use tools to help them deliver excellent customer service. We will work with our Employee Communications team to provide staff with the information on how this may affect them. Communication channels are likely to include:

1. Team Meetings and briefings,
2. Staff magazines and bulletins (electronic and hard copy),
3. Emails,
4. TfL Yammer Network.

Pilot Period

We plan to run this pilot for four weeks commencing on Monday 5 September 2016. We will communicate to customers the intentions of the pilot both before the pilot begins and during the pilot through our available communication channels.

Data Collection Area

TfL proposes collecting and processing data from only those stations in Figure 1. This is a total of 54 stations (Appendix 1).

Figure 1: Proposed Area where data will be collected, hashed and anonymised and analysed.
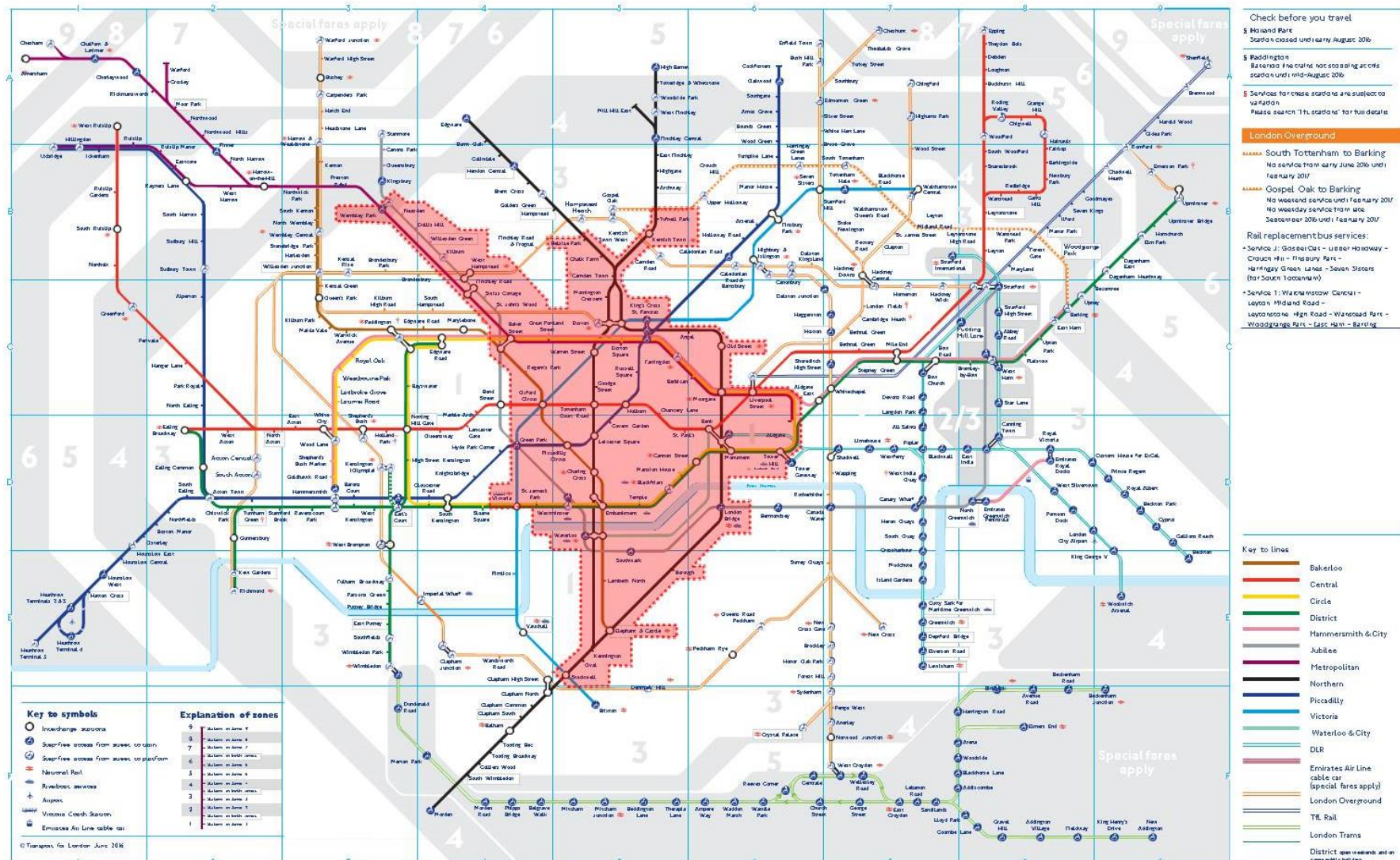
*Tottenham Court Road has no Wi-Fi provision. Based on information http://content.tfl.gov.uk/wifi-underground-overground-map-october-2015.pdf

Rationale for Data Collection Area

This area has been carefully selected to enable TfL to collect the variety of data required to test the use cases whilst minimising the amount of data collected during the pilot. The justification for selecting this as the data collection area are:

1. *Understanding route choice*. We want to understand the various routes (lines and stations) taken by customers between their entry station and exit station (which we may already know from ticketing data). This data will enable us to better plan the network. Rolling Origin Destination Survey (RODS) currently provides this information based on a sample of customers but this is deliberately carried out on days without disruption or events. We believe Wi-Fi data has the potential to replace RODS and deliver a superior, larger, more frequent and cheaper data set. It would provide us with a continuous data set and we could understand seasonal changes and trends and monitor the impact of a range of factors. We have selected part of Zone 1 as it has a dense number of stations, offers a high number of interchange options, has multiple routes between two stations, and customers are likely to show a far greater propensity to interchange compared to outer zones which are served by fewer routes and options. Further to this we also need an area where we can monitor entries and exits. The perimeter effect of the circle line will enable this. We have selected the Metropolitan Line and Jubilee Stations between Finchley Road and Wembley Park to understand what line and train a customer takes when there are two options where one is faster than the other (5 minutes on the Metropolitan Line versus 12 minutes on the Jubilee Line).

2. *Understanding crowding and train assignment.* We want to understand if we can use a combination of Wi-Fi and train scheduling and telemetry data to measure on train crowding and to help us understand how customers interact with crowded trains. This will assist communications with customers about the specific levels of crowding they may experience on their journeys and allow them to make informed choices about the type of journeys they make. To fully understand this we have chosen some Zone 1 stations where demand and crowding are most prominent. Furthermore, we have selected branch lines to enable us to assess where customers may choose to remain on the platform as they wait for a train to their destination and not because the train is crowded and prevented them from boarding. To understand the impact of branch lines we have selected part of the Northern Line that includes the City and Bank branch and the Edgware and High Barnet branches. We have not

selected the whole of the Northern Line but two stations after the branch which we think is the minimum required in order to test and validate our outputs.

3. *Understanding in station movements.* TfL has a variety of stations on its network from small stations served by one line to major termini with multiple lines and many possible walking routes within stations to reach platforms. To understand in station movements we have selected some of the largest and most complex stations where a variety of movements are possible including large termini (Kings Cross St. Pancras, Euston, Waterloo, Liverpool Street, Victoria), large interchange stations (Green Park, Bank, Monument, Westminster, Baker Street, Embankment) with up to 10 platforms, and small stations where station movements should be straightforward (St. James's Park, Barbican, Temple, Euston Sq.).

4. *Scaling seen devices to network totals.* Wi-Fi data will only be a sample of the population. Some customers will not carry a Wi-Fi capable device, others will have Wi-Fi turned off, and some may not be within range of the Wi-Fi access points. We are also likely to double or treble survey some customers who carry and connect to the network with multiple Wi-Fi enabled devices. By selecting the proposed area we will be able to understand how we scale the devices seen to network totals and remove any bias within the data.

5. *Understanding the impact of mobile network connectivity on Wi-Fi.* Only 45 per cent of the London Underground network is in tunnels and therefore has limited data connectivity. For the remaining 55% it may be possible to connect to a radio network (GPRS, EDGE, 3G or 4G) through a mobile network provider for data connectivity. We wish to understand how the availability of radio networks impact the number of customers connecting to station Wi-Fi. To facilitate this we have selected the stations between Finchley Road to Wembley Park on the Metropolitan and Jubilee line. Here the trains run outside of tunnels and customers should be able to choose between Wi-Fi and mobile radio network.

Pilot data access and security

Data will be stored in an on-estate secure data warehouse and will be hashed and anonymised to prevent the identification of an original MAC address. A limited number of analysts will have access to the data during the pilot. These analysts are required to complete annual tests on Privacy and Data Protection, and Information Security. Analysts accessing this data are familiar dealing with personal and sensitive data through their access and management of the Oyster and Contactless Payment Card data.

Pilot Review and Next Steps

During the four week pilot we will collect, process and start analysing the data. We will only process and analyse data collected during the four week pilot. It is likely that we will continue to process and analyse data collected during the pilot after the pilot has ended to continue testing the use cases. Once we have completed the analysis of data against the use cases we will review the pilot. We will ask for questions to determine the success of the pilot:

1. Did the collection, processing and analytics of data deliver envisaged benefits and value against the use cases stated earlier in (Use Cases and benefits of analysing Wi-Fi data)'

2. What, if any was the response from our customers and the public to TfL undertaking this pilot?

3. What, if any was the response from organisations that govern and regulate TfL?

4. What, if any was the response from specialist interest groups?

Following this review, TfL will share its findings with regulatory and governing bodies and decide whether:

1. To cease collecting and processing Wi-Fi data,

2. To undertake an additional or extended pilot to allow more analysis or feedback,

3. To commence full network data collection to enable continuous analysis of Wi-Fi, creating transactional and aggregated data.

Based on the option selected, appropriate communications will take place with stakeholders, customers and staff.

Appendices

Appendix 1 – List of Stations for proposed pilot

Aldgate	London Bridge
Angel	Mansion House
Baker Street	Monument
Bank	Moorgate
Belsize Park	Mornington Crescent
Blackfriars	Neasden
Borough	Old Street
Camden Town	Oval
Cannon Street	Oxford Circus
Chalk Farm	Piccadilly Circus
Chancery Lane	Regent's Park
Charing Cross	Russell Square
Covent Garden	St. James's Park
Dollis Hill	St. Paul's
Elephant & Castle	St. John's Wood
Embankment	Stockwell
Euston	Swiss Cottage
Finchley Road	Temple
Green Park	Tower Hill
Holborn	Tufnell Park
Kennington	Victoria
Kentish Town	Warren Street
Kilburn	Waterloo
Kings Cross St. Pancras	Wembley Park
Lambeth North	West Hampstead
Leicester Square	Westminster
Liverpool Street	Willesden Green

Contact Information

EVERY JOURNEY MATTERS

For further information please contact **Lee McGirr** or **Ryan Sweeney**.

Lee McGirr, Privacy Advisor
Email: ███████████████████
Phone: ████████████

Ryan Sweeney, Data Services Manager
Email: ████████████████████████
Phone: █████████████

EVERY JOURNEY MATTERS