Understanding the travel patterns on London Rail network: The use of Oyster card data

Mr Qian Fu

PhD student

Institute for Transport Studies, University of Leeds

ABSTRACT

The growing use of smartcard systems in public transport greatly enhances the ability to collect massive quantities of traveller information. Not only is it employed to be a convenient way of fare payment, but the diverse travel data that it gathers can also help to enlighten both operators and researchers in many different ways. The Oyster system, which is managed by Transport for London (TfL) and implemented across the Greater London area, is one of the most successful applications. By virtue of the Oyster card data, both the system-wide performance and details of the anonymous card users' travel history can be observed, concerning every journey stage performed by all TfL modes of transport and most suburban National Rail wherever the Oyster is applicable. With the initial processing of a sampled set of Oyster card data, a typical O-D pair (Waterloo - King's Cross St. Pancras) is selected and interrogated in this paper. Then various investigations for travel time variability and reliability of service are presented, categorised by different time periods within a day and also between different days (weekdays and weekends). This paper mainly reports on initial analyses on the following specific objectives: (a) to make clear the context under which the data is generated and applied as well as the role that Oyster card data plays; (b) to explicitly expound what sorts of information the Oyster card data contains; and (c) to further examine the variations of travel times between certain periods with reference to various aspects of travel patterns discovered within the Oyster data.

INTRODUCTION

To look at public transport network, different travellers' behavioural intentions may imply different travel patterns concerning e.g. different planned destinations, departure times and desirable travel routes which lead the travellers to complete their journeys. What's more, travel patterns vary potentially due to the travellers' responses towards the reliability of the transit service, especially as for those commuters who could gain experience of the network performance in terms of day-to-day variations in their travel times. Therefore, a good understanding of such different travel patterns under various backgrounds is essential to the efficient public transport planning, operations and travel demand management.

As has been widely put into use and being increasingly improved during the past decade (Pelletier *et al.*, 2011), smartcard systems are playing active roles in data collection of fares and travel information for public transport. The due expense of a travel pass would be validated or deducted automatically via the intercommunications between the smartcard and card reader. In the meantime, a record of every transaction occurred, accompanying with its various related data (e.g. start and end locations and the associated timestamps) is created and stored. Such information can provide rich opportunities for both future planning and research. Thus the utilisation of smartcards for public transit substantially enhances the ability of capturing numerous travel data of many kinds, and largely releases the restraints of conducting manual surveys (Bagchi and White, 2005). Some representative examples include Oyster in London (Lathia *et al.*, 2010, Seaborn *et al.*, 2009, Chan, 2007, Uniman, 2009, Lathia and Capra, 2011), Passe-Partout PLUS in Gatineau (Trépanier *et al.*, 2009, Morency *et al.*, 2007), Chicago Card (Plus) in Chicago (Utsunomiya *et al*., 2006), and T-money in Seoul (Park *et al.*, 2008, Jang, 2010), to name but a few.

This paper mainly presents initial analyses with the purpose of: (a) making clear the context under which the data is generated and applied as well as the role that Oyster card data plays; (b) explicitly revealing what sorts of information the Oyster card data contains; and (c) further examining the variations of travel times given certain periods by looking at the travel patterns from different aspects.

BACKGROUND

Transport for London & Oyster scheme

Being as an integrated part of the local government, Transport for London (TfL) basically undertakes responsibility for the transport system and its services across the Greater London area. With regard to the aspects of public transport, TfL is taking charge of the management and operations for multifarious transit services running on a massive and intricate network (TfL network). It involves a variety of transport modes including London Buses (Bus), London Underground (LU, commonly known as Tube), Dockland Light Railway (DLR), London Overground (LO), London Tramlink (Tram), London River Services (LRS), and some other customised services such as Dial-a-Ride particularly for disabled people, etc.

In accordance with the zonal fare scheme, the London Rail network (here we define London Rail, or simply Rail, as a generic term which refers to LU, DLR, LO and National Rail (NR) services) is carved out into 11 fare zones. A central zone (Zone 1) is based in the centre of London, which is surrounded by 5 roughly concentric ring-zones (Zone 2 - Zone 6) radiating outward one by one. In addition, there exist 3 ancillary zones (Zone 7 - Zone 9) that are positioned to the northwest central London, plus 2 further zones (Watford Junction and Grays). Unlike buses and trams where unitary fare regime is adopted throughout, fare charged on using the London Rail services mostly depends on how many zones a passenger travels through.

Since in 2003 the Oyster scheme (simply Oyster) was introduced, the Oyster card has then become the most popular method of fare payment as utilised by London's public transport users. According to the Factsheet of TfL by November 2010, over 80% of journeys on TfL network have been paid through the Oyster. Not only can it be employed to make the fare payment on most TfL modes (e.g. LU, DLR, LO, LRS, Bus and Tram), but also on the bulk of suburban NR services of which terminals are within the schemed fare zones mentioned above. However, three TfL rail modes that are specified on the 'Standard Tube Map' (i.e. LU, DLR and LO) share the same rule of charging fares, which is distinct from the others (i.e. LRS, NR, and Bus & Tram) that adopt their respective fare schemes.

In practice, passengers choose to buy different types of ticket products in line with their own trip purposes and route choices. Different alternatives that passengers purchase may also suggest different travel patterns. At a broad level, the fare charging plan applies for two categories, 'Pay-As-You-Go' (PAYG) and Travelcards. In general, Oyster card acts as PAYG by default, in which case the cardholders need to deposit some credits before they start to travel and the fares would be calculated and deducted once the trip or a journey stage is finished. Meanwhile, diverse alternative travel passes can be chosen in the form of Travelcards. They are designed to adapt different purposes pursued by different public transport users (e.g. travelling without frequency limits between certain fare zones within a certain period of time, and getting a student discount or free passes). Such Travelcards can be 'attached' to Oyster cards (except 1-Day Travelcard only valid for a day pass) and they would be prioritised to be validated by card readers. Furthermore, an Oyster cardholder can combine both PAYG and Travelcards together into one Oyster card, so that any excess fare that is caused when travelling beyond the prescribed zones shown on Travelcards can be covered by PAYG credits.

The implementation of smartcard like Oyster is not only an efficient tool of public transport fare collection but, more importantly, it also gathers a wealth of travel information (i.e. Oyster transaction data) that reflects how the transport system is made use of by the cardholders. Hence mining the smartcard data is playing a significant role in public transport planning and operations of which the foremost task is to find out what information the data can pass on to us. Nevertheless, also note that, the traditional paper tickets are also available at every station to the travellers who do not hold Oyster cards, and the Oyster data records do not correspond to the entire ridership in London.

The Oyster card data

The Oyster card data is collected automatically as the Oyster card being touched on a card reader. Miscellaneous information is then generated and appropriately stored in separate data subsets. They gather both aggregated statistical data (e.g. the count of entries and exits at each station, and the number of journeys grouped by different transport modes, time periods, stations and ticket types) and, on the individual level, detailed (but anonymised) travel history of individual Oyster users and the fare payment information. In addition, each of Oyster card users' trips is presented seriatim as per transaction in the data set of travel history which is mainly targeted in our following analysis. Table 1 describes what sorts of information is included in Oyster data.

Table 1. Information contained in the Oyster card data

General category	***Information***	***Element***
Bus & Tram	Boarding information: amounts of boarding by different times, dates, ticket products (with validity) and locations	Time segment (a certain period of time in a day), day and date
		Location (or name of fare stage)
		Route direction and route number
		Ticket product along with validity information
		Counts of boarding
Rail services (including Tube, DLR, LO, some NR)	Aggregate information of rail journeys: amounts of rail transactions of each O-D pair, by different times, dates, types of card and ticket products added (PAYG, Travelcards and Mixed-payment), and the passenger flows at each station)	Time segment (or hour), day and date
		Station of origin and destination
		Ticket product along with validity information
		Oyster card type (unfinished journeys only)
		Counts of all rail journeys
		Counts of entries and exits through ticket barriers (as well as the category of the ticket barrier)
The purchase of different ticket products	Purchase information of different types of ticket products associated with each Oyster card user, including both PAYG and other products of tickets (e.g. Travelcards, Freedom Pass, etc.)	Oyster card number (encrypted)
		Date and location (or agent) of purchase
		Credit added and the new balance (PAYG only)
		Type of ticket product (other than PAYG), along with its specific scope of use, e.g. start and expiry date, valid time and zone)
History of each Oyster user's transactions	Details of the every Oyster transaction (except purchase or top-up) on all the transport modes where the Oyster is applicable, involving time, day and date, mode of transport, start and end location, and fare payment, etc.	Day and date
		Oyster card number (encrypted)
		Number of sequence over all transactions conducted by each Oyster card
		Mode of travel
		Route direction and route number (Bus only)
		Station of origin and destination
		Information of travelled distance (Tube and DLR only) and zones
		Entry and exit time of each transaction
		Ticket product used and its validity information
		Information about fare charged (cap, full or discounted fare price)

To go into the sequenced travel history records, each row of the data corresponds to a transaction made by a traveller using an Oyster card identified by a unique encrypted number (Oyster card ID). We presume that a card is only used by one traveller who is in turn represented by an Oyster card ID. When travelling on Rail services, passengers are required to present Oyster cards at both their stations of entry and exit. However, this is not the case upon most occasions when they interchange within the Rail network in the course of their trips. Particularly at some transfer stations, passengers will have to, for some reason, touch out first and then touch in at the same (transfer) station where they intend to interchange from one service line to another. These special cases are called 'Out of Station Interchanges' (OSIs). Hence each transaction record could be deemed as either a complete trip, or just a journey stage that is defined as a part of a journey/trip performed on a single mode of transport (TfL, 2009). And accordingly, a complete trip can be made up of just one journey stage or several. That is, the trip can be achieved by a single transport mode (e.g. London Underground), or it could involve a couple of modes (e.g. boarding DLR for the first journey stage and then transferring to London Overground for the second). Here, we further define that a journey stage refers to a part of a journey/trip completed on a single transport mode without any interchange, which allows us to get insight into the latent route choices in the course of a complete journey.

By far, we used a sample set of Oyster card data to start up the initial analysis. The sampled data set involves travel history associated with 5% of the overall Oyster cardholders travelling within 28 days (without any public holidays included), namely, from 6 February (Sunday) to 5 March (Saturday) 2011. It amounts up to 11,418,860 records of transactions made by 347,646 travellers, and whereby it can be inferred that there should be about 6,952,920 Oyster cardholders travelling on Oyster services network during the 28-day period. In the following sections, we present our initial analyses based on this extracted dataset.

INITIAL ANALYSES

Aggregate analyses of the given Oyster data

As many researchers pointed out that there would always be some error records existing in the automatic fare collection (AFC) systems data (Jang, 2010, Utsunomiya *et al.*, 2006), and the Oyster is no exception. Nonetheless, before the raw data was cleaned, its integrity was maintained for some aggregate analyses in order for us to attain statistical information as accurate as possible. By drawing support from exploiting the raw Oyster data, we are aiming at bringing out a big picture of the system-wide situation.

It can be seen from Figure 1 that the Bus trips comprise a fairly large proportion up to almost 65% of the more than 11 million sampled transactions. By contrast, only a third of the total records are associated with the other Oyster services, except LRS that is not included in the given data set. (Heathrow Express Terminal Five is counted in the Tube network.)

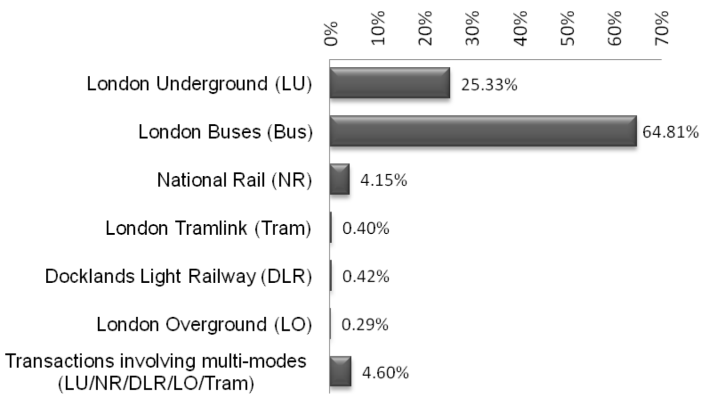
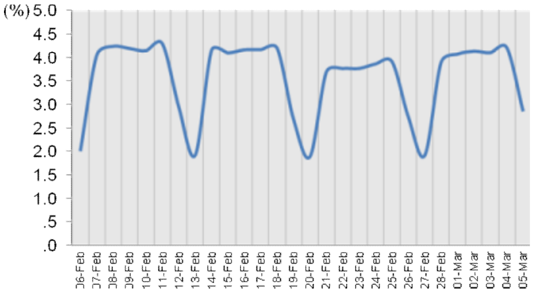
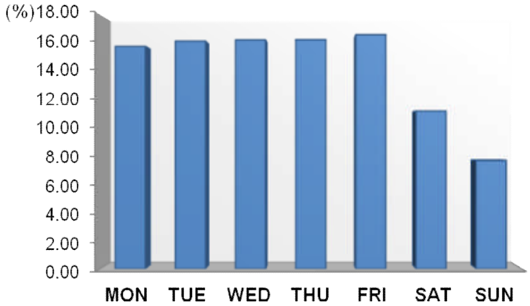


Figure 1. The proportion of transactions made on different modes

We define that the weekend refers to either Sunday or Saturday, and the weekdays refer to the working days from Monday to Friday. In our given sample data, there are 20 weekdays plus 8 weekends. Figure 2 shows the proportions of Oyster transactions made grouped by different weekdays and weekends within the sampled period. Evidently, people travel much less frequently during the weekend than on weekdays. Besides, there are differences of service frequency between working days, Saturday and Sunday. Thereby the journey time distributions on weekdays for any given O-D pair are then supposed to hold obvious distinction from weekend.



(a) Daily proportion of transactions



(b) Grouped proportion of transactions by day

Figure 2. The proportion of the Oyster transactions made on different days over the sampled 28-day period: **(a)** daily proportion of transactions; and **(b)** grouped proportion of transactions by day

As is shown in Figure 3, based on the sampled Oyster records, about 60% cardholders made only one trip a day from 6 February to 5 March. In the case of such low frequency of transactions made per day, we believe that there must be a fair portion of visitors or tourists, who would potentially make quite a few trips within a couple of days. Apart from that, suppose that a commuter goes to and back from work on every weekday and has one or two trips during weekends. Thus we could make an assumption that he/she might make use of the Oyster service without or with at most one observable interchange during a complete trip. Then, normally, a commuter that uses Oyster could be daily involved with 1.5 to 4.5 transaction records. According to the statistical figures, over 35% Oyster users are counted in this range. Nonetheless, there are still about 5% transactions related to people who travelled much more frequently. Some of them (Staffs) even made more than 20 transactions per day. Therefore, it would be interesting to try to distinguish between different groups of Oyster users such as commuters with Travelcards, staff travel and concessionary (e.g. elderly and disabled). Given any O-D pair, the travel times may vary much among these users.

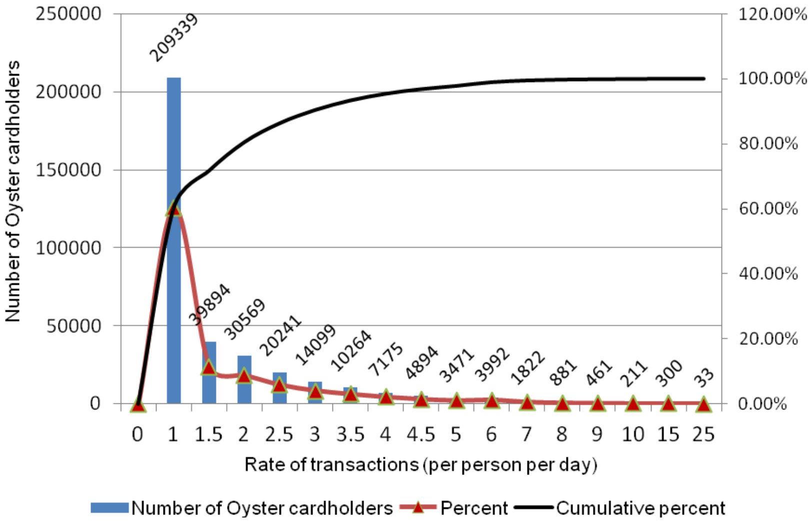


Figure 3. The distribution of Oyster transactions over 28 days

The Oyster users can be roughly identified in the light of the purchased ticket products. All the transactions could be generally categorised by 3 groups of Oyster users in terms of their ticket payment methods: Pre-payment (Pay-As-You-Go), Tickets (e.g. Travelcards, Freedom Pass, etc.), and Mixed-payment which involves both of the aforementioned types of methods with respect to one Oyster transaction. Figure 4 reveals the proportions occupied by purchasers of different types of tickets. Note that PAYG (35.51% of the total) is being applied to all the modes included in the data set wherein about 54.8% corresponds to Bus and Tram transactions as sorted by different modes in particular to PAYGs.

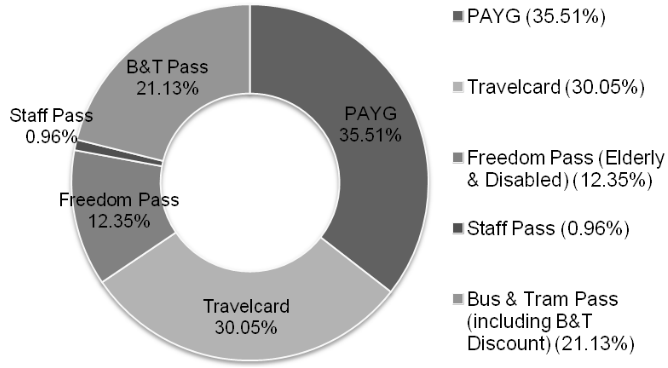


Figure 4. The proportion of different types of tickets that are purchased

All the results presented above are based upon the raw data that contains errors. For the sake of further looking at the travel pattern and service reliability of any specific pair of origin and destination on the network, valid data is indispensable. The criteria and of cleaning the data as well as the processed results are described in the following sections.

Cleaning the raw data

The initial steps taken to disposal of the errors in the sample Oyster card data is mainly intended for finding out the issues with respect to unreliable data records so as to avoid their interferences to further analysis. The general criteria to sort out our concerned data are as follows: a) both origin(s) and destination(s) are identified for some travellers might fail or forget to tap out with Oyster cards; b) as for any valid journey stage that we would consider, the start location is supposed to be different from the end location; and c) the travel time of every journey stage should be valid (i.e. a positive number) as it can be calculated as the difference between the times of touching out at the exit and touching in at the entry. These errors may be caused by occasional non-synchronization of the system time between origin and destination. Note that any of those calculated travel time (Oyster travel time) corresponds to the time in minutes that a traveller spent between the ticket barriers of his/her start and end station.

Consequently, we are not considering the trips taken by bus and tram, because passengers do not need to show or touch out with their Oyster cards when they are trying to get off or step off-board. In that case, those passengers' alighting locations of taking bus or tram, and hence the travel times, are not able to be observed by the Oyster card data (but only the fare stage in light of the boarding location can be recorded accompanying with the bus route ID/number and heading direction).

Finally, after excluding about 75% of the raw data, only 2,919,256 transactions remain, which are all related to London Rail services except NR. The records regarding Bus and Tram are filtered out, and the mode shares based on the processed data set is displayed by Figure 5. It reflects the proportions of the transactions made on different transport modes based on the processed sample data. It should be noted that many journey stages may have been ignored as a result of dropping the error records. We would be then unable to trace or reproduce a complete trip due to any missing journey legs. This, however, does not matter the cases that interchanges are completed without touching out of the stations, for instance, people make interchanges when travelling on London Underground.

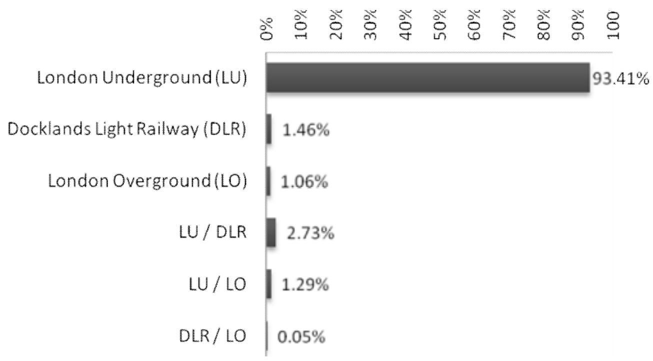


Figure 5. The proportion of transactions made on different modes based on the processed sample data

We then pick out a typical O-D pair to examine the travel time variations within a day and between days. Suppose that passengers would rather travel with 'Tube' whenever available. Because it can provide relatively more frequent services, and the 'Standard Tube Map' (that only consists of LU, DLR and LO) is the primary reference source for travellers in London. We were also interested in looking at whether different travel times might explicitly suggest that different routes were chosen. Thus, the selection criteria should be in accordance with the following principles: firstly, there is no direct service but passengers will have to interchange once; secondly, alternative routes as well as optional service lines are available between the origin and destination; thirdly, 'National Rail' services are avoided or unavailable, as are the cases of OSI, since only the users of the 'Standard Tube Map' are taken into account. Consequently, the O-D of 'Waterloo Underground station - King's Cross St. Pancras Underground station' is selected, and 764 transaction records were obtained based on the given Oyster sample that contains the travel history of 5% of all Oyster users in 28 days.

Table 2. Information of the selected O-D (Waterloo - King's Cross St. Pancras)

Origin (LU)	Service Tube line	Interchange Station	Service Tube line	Destination (LU)	JP-TT (min)	Average headway (min)	Estimated total TT (min)
Waterloo	Northern	Warren Street	Victoria	***King's Cross St. Pancras***	19	3.5	21
	Northern	Leicester Square	Piccadilly		16	3.5	18
	Northern	Euston	Northern		19	3.5	21
	Northern	Euston	Victoria		19	3.5	21
	Bakerloo	Oxford Circus	Victoria		15	4	17
	Bakerloo	Piccadilly Circus	Piccadilly		16	4	18
	Jubilee	Green Park	Victoria		18	3.5	20
	Jubilee	London Bridge	Northern		23	3.5	25
	W & City	Bank	Northern		21	6.5	24
King's Cross St. Pancras	Victoria	Warren Street	Northern	***Waterloo***	N/A	3.5	N/A
	Victoria	Euston	Northern		16	3.5	18
	Victoria	Green Park	Jubilee		16	3.5	18
	Victoria	Oxford Circus	Bakerloo		19	3.5	21
	Piccadilly	Piccadilly Circus	Bakerloo		15	3.5	17
	Piccadilly	Leicester Square	Northern		17	3.5	19
	Northern	Euston	Northern		18	4	20
	Northern	London Bridge	Jubilee		23	4	25
	Northern	Bank	W & City		22	4	24

Note: 1). JP – Journey Planner; TT – Travel time; W & City – Waterloo & City service line
2). N/A indicates that there is no related information available on Journey Planner
3). JP-TT is the average estimated journey time in accordance with the result searched via JP on 16/11/2011.

Essential information regarding our example O-D is revealed in Table 2. It exhibits all the reasonable alternative routes that travellers can choose (involving the related service lines and alternative interchange stations). As for this example, each Oyster transaction record refers to a (part of) trip that comprises two journey stages. Furthermore, the journey times estimated by TfL Journey Planner (an online search engine that helps travellers to plan journeys by all means of public transport and bicycles in and around London) are collected. Each JP-TT is composed of the train running times on both of the journey stages plus the time spent for interchange. And the average headways are roughly estimated according to the varying frequencies of the related service lines. Since we presume that passengers arrive randomly at the station and the service frequency is high, each of the estimated total travel time listed in the last column is calculated by the equation of 'Estimated total TT = JP-TT + ½ average headway', where a half of the average headway corresponds to any passenger's mean waiting time for the related service line. Note that this is merely an approximate calculation that provides a rough note for reference. In the following analysis, only northbound service (i.e. services from Waterloo to King's Cross St. Pancras) is investigated. Still, the same procedures could be also adapted to the southbound service as well as any other O-D examples.

INVESTIGATION ON A SELECTED O-D EXAMPLE

Initial investigation on the typical O-D example by mining the Oyster data

As is illustrated in Figure 6, the Oyster travel times range from 15 minutes up to 73 minutes, yet very few people spent more than 30 minutes in travelling between the O-D of Waterloo to King's Cross St. Pancras. Despite the obvious existence of 3 'break points', the ridership shared by different routes is not clear via comparisons with the 'estimated total travel time' specified in Table 2. This is because, on the one hand, the sample size is too small to gain a relatively accurate travel time distribution; on the other hand, this curve lies over the entire sampled period wherein the travel patterns may potentially vary between different time segments (e.g. Peak and Off-peak). It is thereby necessary to further review the data in terms of different periods of time.

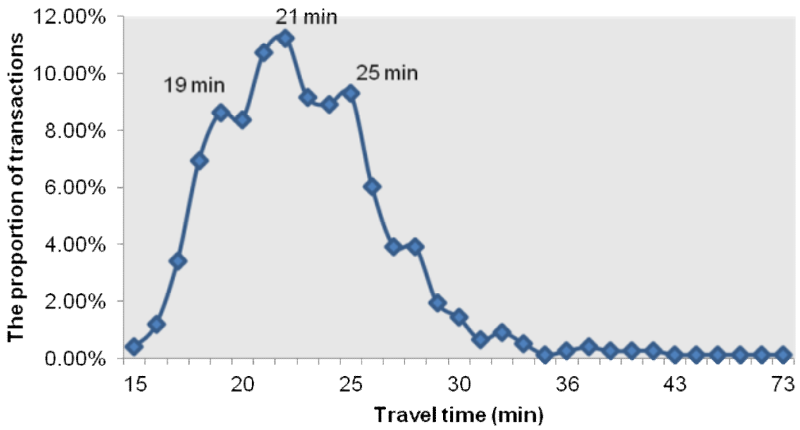


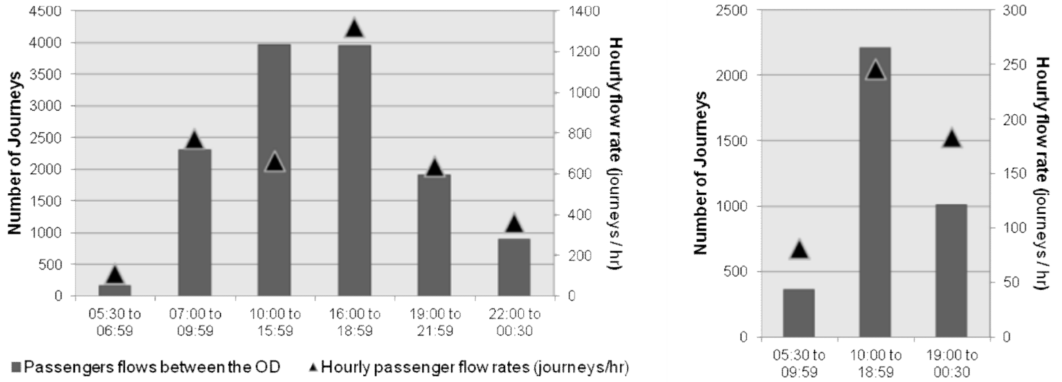
Figure 6. The distribution of the Oyster travel times between the Underground stations of Waterloo and King's Cross St. Pancras

TfL specifies two distinguishing periods of service when different single fares are applied. They are: Peak hours including morning peak from 6:30 - 9:30 and evening peak from 16:00 - 19:00 from Monday to Friday (weekdays but not counting in any public holidays), and Off-peak hours for the rest of service times.

Passenger flows

Before we go further with the variability of the travel times, the passenger flows as well as hourly flow rates between this O-D pair is examined by different time segments on weekdays and weekends (See Figure 7). As each journey refers to an Oyster transaction, the count of journeys can be also deemed as the passenger flows passing through the selected O-D pair.

From Figure 7 we can see clearly that, the hourly rate of passenger flows is much higher during the Peak time, especially evening Peak hours, than that at Off-peak.

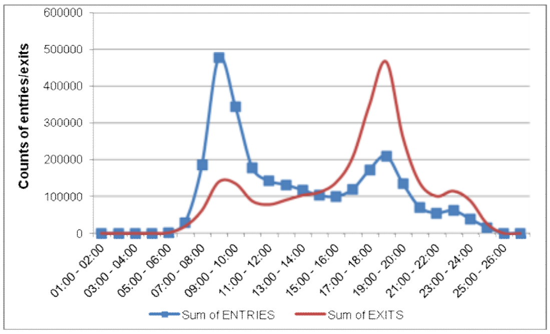


(a) Weekdays **(b)** Weekends

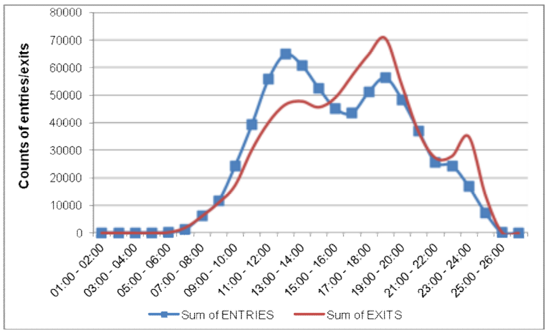
Figure 7. Count of journeys made through the example O-D, along with the hourly flow rates by different time segments on: **(a)** weekdays; and **(b)** weekends

Crowding situations

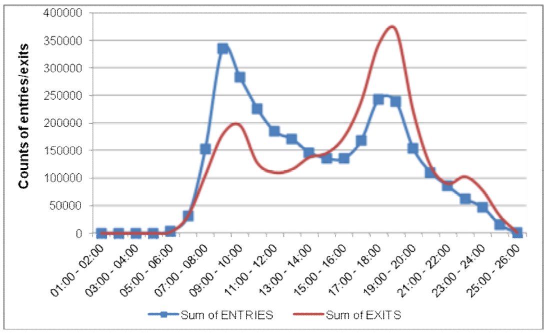
Moreover, as is shown in Figure 8, the peaks of passenger flows are reached at about 8:00 - 9:00 and 18:00 - 19:00 during the weekdays, while on the weekends an influx of travellers does not turns up at both stations until around 12:00 - 13:00. It is also noticed that, in the morning, there are larger amounts of Oyster travellers swarming into either station (touching in) than that of those who touch out. It is just the reverse in the evening hours that the accumulated volumes of passenger outflows from the stations are higher than that of the entries. Under the circumstance of Peak hours, it is believed to be crowded at ticket barriers, in pedestrian channels, on platforms as well as onboard. Undoubtedly, passengers' travel times would be affected then for various issues incurred by such crowding situations.



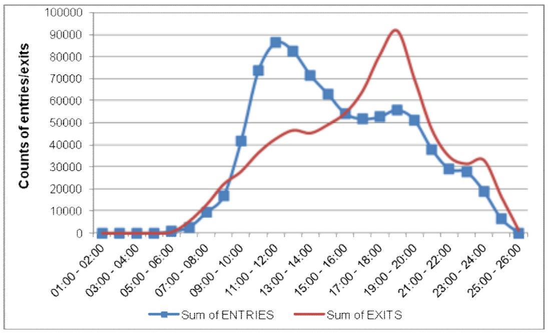
(a) Waterloo, on weekdays



(b) Waterloo, on weekends



(c) King's Cross St. Pancras, on weekdays

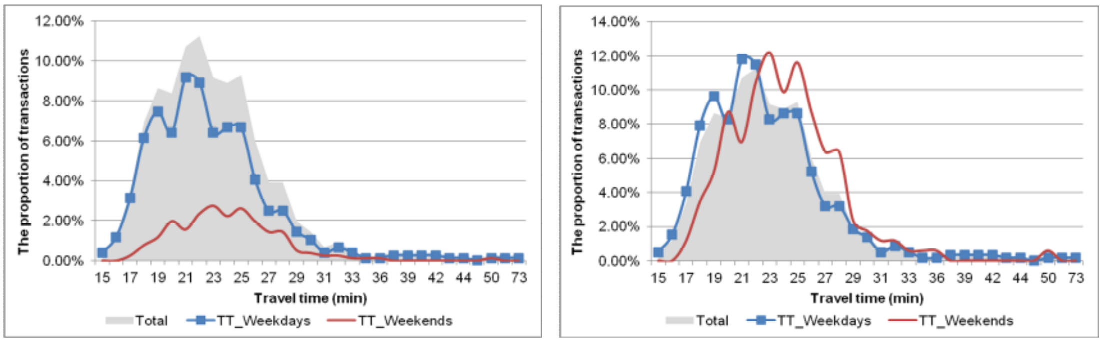


(d) King's Cross St. Pancras, on weekends

Figure 8. The distributions of the counts of entries and exits through the ticket barriers, by stations and different hour periods (weekdays and weekends): **(a)** Waterloo, on weekdays; **(b)** Waterloo, on weekends; **(c)** King's Cross St. Pancras, on weekdays; and **(d)** King's Cross St. Pancras, on weekends

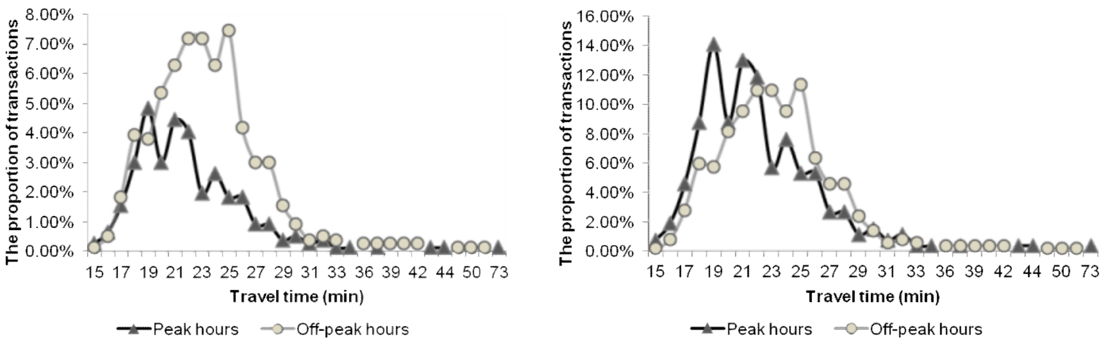
Travel time variations

Being further interrogating our example O-D pair with the help of Oyster data, the following figures (Figure 9, Figure 10 and Figure 11) basically show 3 sets of comparisons of the variations of travel times by different periods of time. They are presented by different travel time distributions between weekdays and weekends (Figure 9), Peak and Off-peak hours (Figure 10), and AM and PM Peak hours (Figure 11).



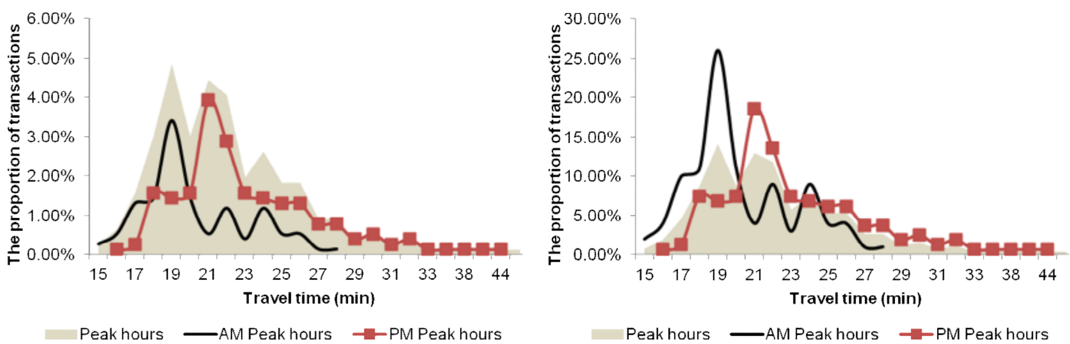
(a) Based on total samples **(b)** Based on their own respective samples

Figure 9. Comparisons of the travel time distributions between *weekdays* and *weekends*: the proportion of transactions is **(a)** based on total selected sample size; and **(b)** based on their own respective samples



(a) Based on total samples **(b)** Based on their own respective samples

Figure 10. Comparisons of the travel time distributions between *Peak hours* and *Off-peak hours*: the proportion of transactions is **(a)** based on total sample size; and **(b)** based on their own respective sample size



(a) Based on total samples **(b)** Based on their own respective samples

Figure 11. Comparisons of the travel time distributions between *morning (AM) Peak hours* and *evening (PM) Peak hours*: the proportion of transactions is **(a)** based on total sample size; and **(b)** based on their own respective sample size

We take Figure 9 for example with the view of an explanation to the figures. The shade area in Figure 9 delineates the overall distribution of Oyster travel times (as that shown in Figure 6). Figure 9(a) depicts the travel time distributions of weekdays' and weekends', respectively, within the gross scope (i.e. the proportion of transactions indicated by either curves is derived based on the total 764 transactions). In Figure 9(b), those curves are dependent upon their own sample sizes. A distinct shift is noticed, which illustrates that the travel time on the weekends is relatively longer than that during the weekdays.

CONCLUSION

For public transport a smartcard system can not only be used as a integrated ticketing tool for fare collection and/or payment verification, but also the card holder's every trip (or journey stages) could be recorded and traced anonymously for further level of data analysis. Since there are millions of passengers (over 80% of the total) who use Oyster cards travelling across London, a great quantity of travel data with rich information is gathered. The potential for understanding and use of smart card data have been put forward in a breadth of contexts.

This paper demonstrates how Oyster data can help us to gain a better understanding of travel patterns on London Rail network. The scope of applying the Oyster scheme is firstly elaborated, along with the description of what types of information the Oyster data contains. We then explain the methods of processing and cleaning the raw data for the sake of obtaining meaningful information from further analyses. Indeed, plenty of the travel information can be captured by the Oyster system, such as the amounts of rail journeys of all O-D pairs sorted by different time periods within a day and also on different days. In addition, more details of the card users' travel history are accumulated by journey stages, relating not only to those on London Underground but also Bus, London Overground, Tram, Dockland Light Railway and the bulk of suburban National Rail. Since the Oyster card data also keeps records of the origin and destination stations and trip stages and the corresponding times of entry and exit; the information of travel times, and thus the travel time distributions and the O-D matrices can be obtained accordingly. Then a variety of measures for travel time variability are derived at the same time.

In line with our stated criteria, we select one typical O-D pair on the basis of the London 'Standard Tube Map', namely, the Underground stations of Waterloo and King's Cross St. Pancras. Various aspects of information are revealed via mining the corresponding Oyster card data. Based on the selected O-D as well as its related Oyster data, we demonstrate our methods by examining different periods (i.e. weekdays and weekends; Peak and Off-peak; and morning Peak and evening Peak) to reflect the likely passenger flows, the crowding situations and the variations of travel times. The above very approximate estimates could be greatly improved with more data.

ACKNOWLEDGEMENT

The author appreciates his funding support from China Scholarship Council and University of Leeds. He would also like to express his gratitude to the staff of Transport for London. Special thanks go to Andrew Gaitskell for his continued support on issues concerning Oyster card data and he is grateful to his supervisors Dr. Ronghui Liu and Dr. Stephane Hess for their guidance and comments.

REFERENCE

BAGCHI, M. & WHITE, P. 2005. The potential of public transport smart card data. *Transport Policy,* 12**,** 464-474.

CHAN, J. 2007. *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data.* S.M. Thesis, Massachusetts Institute of Technology.

JANG, W. 2010. Travel Time and Transfer Analysis Using Transit Smart Card Data. *Transportation Research Record: Journal of the Transportation Research Board,* 2144**,** 142-149.

LATHIA, N. & CAPRA, L. 2011. How Smart is Your Smartcard? Measuring Travel Behaviours, Perceptions, and Incentives. *13th ACM International Conference on Ubiquitous Computing (Ubicomp).* Beijing, China.

LATHIA, N., FROEHLICH, J. & CAPRA, L. 2010. Mining Public Transport Usage for Personalised Intelligent Transport Systems. *Proceedings of the 2010 IEEE International Conference on Data Mining.* Sydney, Australia: IEEE Computer Society.

MORENCY, C., TRÉPANIER, M. & AGARD, B. 2007. Measuring transit use variability with smart-card data. *Transport Policy,* 14**,** 193-203.

PARK, J., KIM, D.-J. & LIM, Y. 2008. Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea. *Transportation Research Record: Journal of the Transportation Research Board,* 2063**,** 3-9.

PELLETIER, M.-P., TRÉPANIER, M. & MORENCY, C. 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies,* 19**,** 557-568.

SEABORN, C., ATTANUCCI, J. & WILSON, N. 2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record: Journal of the Transportation Research Board,* 2121**,** 55-62.

TFL 2009. Travel in London, Report number 1. Transport for London.

TRÉPANIER, M., MORENCY, C. & AGARD, B. 2009. Calculation of Transit Performance Measures Using Smartcard Data. *Journal of Public Transportation,* 12**,** 76–96.

UNIMAN, D. L. 2009. *Service Reliability Measurement Framework using Smart Card Data: Application to the London Underground.* S.M. in Transportation, Massachusetts Institute of Technology.

UTSUNOMIYA, M., ATTANUCCI, J. & WILSON, N. 2006. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning. *Transportation Research Record: Journal of the Transportation Research Board,* 1971**,** 119-126.