How Can "Big Data" Be Harnessed to Enhance Congestion Management

Karlson Hargroves, Daena Ho, Daniel Conley, Peter Newman

Curtin University Sustainability Policy Institute, Curtin University, Perth, Australia

Email address:

charlie.hargroves@curtin.edu.au (K. Hargroves)

Abstract: Traffic congestion is a key issue facing transport planners and managers around the world with many now asking if there are any promising technologies offering new solutions. In the US, the cost of congestion was $121 billion in 2012 and in 2015 alone Australia's capital cities were estimated to have a combined congestion cost of $16 billion, expected increase to $37 billion by 2030. With the rapidly growing availability of data and the ability to analyse large data sets this paper investigates the question "What role can 'Big Data' play to assist with congestion management?" There is great interest and hype around 'Big Data' and this paper provides a summary of an investigation into its value to assist in relieving congestion. The paper explores the emerging types of large data sets, considers how data will be sourced and shared by vehicles and transport infrastructure in the future, ad explores some of the associated challenges. Despite the opportunities of Big Data not being fully realised it is already clear that it presents a significant tool for transport planners and managers around the world to assist in managing congestion. The research is based on research undertaken with the Sustainable Built Environment National Research Centre (SBEnrc).

Keywords: Big Data, Predictive Congestion Management, Technology Enabled Transport

1. Introduction

There are multiple definitions of Big Data. Most commonly, the term is used to broadly characterise data sets so large they cannot be stored and analysed by traditional data storage and processing methods. Large volumes of data are now available from a growing number of sources, however, this is only one dimension of its complexity. The velocity at which data is received and the variety of information available adds to the challenge of creating value. Further, data is now produced in multiple formats, languages and software configurations depending on where the data is sourced.

It is these three characteristics (referred to as the three V's – Volume, Velocity and Variety) that distinguish 'Big Data' from other forms of data. The emergence of such large and complex data sets has primarily been the result of a decrease in the cost of sensory and observational technologies in conjunction with mass digitisation of systems and processes around the globe. Combined with a vast amount of information available in a worldwide network of distributed archives, data from large-scale sensor networks and computer simulations is now creating an immense resource to be harnessed.

Not only can data be used to observe direct traffic related phenomenon (such as streamlining traffic signal timings and ramp metering) but the interrogation of traditional and non-traditional data streams provides a unique potential to identify linkages between previously seemingly unrelated data and transport related activities. Because of the extremely large volumes of information produced, 'Big Data' requires analysis in order to produce meaningful results. The term 'Big Analytics' is used to describe the processing of multiple massive data sets to extract useful algorithms and information that can be visualised, say by a traffic control centre.

Big Data provides the promise of going beyond what could be referred to as 'Small Data' - data such as traffic counts, average velocity, temperature conditions, traffic light signal durations etc. – to include consideration of literally hundreds of data sources that stand to inform congestion management

efforts, such as data streamed directly from vehicles, data about car parking, data about public transport, data on social events that may affect traffic, meteorological data, etc and even data previously thought unrelated to congestion management.

Given the economic impacts of congestion it is critical that new data streams are used to inform both traffic management and transport planning. Congestion reduction has economic, environmental and health benefits that can be greatly enhanced by harnessing new data sets. Firstly, reducing peak-time congestion defers required capital investment: an additional road or highway does not have to be built if peak-time traffic is no longer an influential factor. In addition, road congestion has a financial cost. In the US, the cost of congestion was $121 billion in 2012, which equates to $818 per commuter per year. [1] According to a study by BITRE, Australia's capital cities were estimated to have a combined congestion cost of $16 billion AUD in 2015, with an expected increase to $37 billion AUD by 2030. [2]

Further, reduced vehicle wait times in traffic jams reduces vehicle exhaust, thus reducing carbon emissions and air pollution. In the US alone, 25 million tonnes of CO_2 per year was emitted from vehicles stuck on congested roads. [3] In addition, inhaling vehicle exhaust for extended periods has also been linked to human health problems such as brain-cell damage. [4] These negative externalities all point to the increased need to manage road congestion, and the growing availability of data might provide part of the solution.

2. Big Data and Transport Systems

2.1. Value Created by 'Big Data' for Transport Systems

The use of data to inform transportation is not a new phenomenon; traffic systems have long produced streams of observational and sensory information, however in today's terms this is typically referred to as "Small Data". In 2014, a study by the Australian Government Bureau of Infrastructure, Transport, and Regional Economics (BITRE) identified a number of available data collection technologies and concluded: '*Recent and emerging technologies offer significant opportunities for collecting more information, more cost effectively, about personal travel activity and road use, that can better inform day-to-day network management, long-term infrastructure planning and road user travel choices*'. [5]

Beyond this there is growing evidence that transport managers are harnessing an increasing scale of data which can achieve a higher level of efficiency, which leads to cost savings, reduced energy demand, better delivery of services, improving life quality and reducing environmental impacts. In this paper, transportation refers to systems of mobility including vehicles, roads, railways, subways, buses, taxis, bicycles, ferries and share-rides. Each transportation mode plays an essential role in mobility of a city and if properly harnessed can move people and products to their destination safely and efficiently at a reasonable cost.

Efforts to reduce congestion through mode shifts to public transport and better management of the road network lead to a number of benefits, such as:

a. Enhanced liveability of cities due to less lost time in congestion,
b. Faster, cheaper journeys that reduce wear and tear on vehicles and the road network,
c. Attracting businesses to cities by providing better and more efficient mobility,
d. Reduced environmental impact such as air pollution and greenhouse gas emissions, and
e. Easing the stress on the city transport budget and maximising the benefit of expensive transportation assets.

According to the Australian Bureau of Transport and Resource Economics, in Australia '*the avoidable social costs of traffic congestion will rise to about $20 billion by 2020*'. [6]

2.2. Collecting Big Data

The rapid rise in the capacity of data storage options (both in-house and remotely) along with the increase in computational ability means that there is great potential to harness additional value from both existing and new data sources that can inform the working of a modern city, especially its transportation systems. Data from transport systems is highly varied and comes in three broad categories:

1. *Highly structured datasets* that originate from technology implemented to address well-defined problems (e.g. data from automatic toll road payment transponders for the use in processing toll road payments or data from intersection sensors on traffic flows and time of day usage of the road network) which can be considered 'Small Data' until it rises in volume to a level that cannot be analysed using traditional methods.

2. *Unstructured datasets* that are produced from any interaction between road users and digital infrastructure. Given the explosion in the use of mobile phones, personal computers, sensors, cameras, and devices, there is huge (yet largely untapped) potential to harness these data streams to inform congestion which is now moving into the realm of 'Big Data', and

3. *Data from seemingly unrelated sources* that stand to provide insights into the behaviour and functioning of the transport system, such as the price of parking at particular public carparks, the level of fines for illegal parking, the amount of people walking more than 1 kilometre to public transport, weather conditions etc., which is typically not included in small data techniques, or needs to be manually inputted from other sources by traffic control centre operators.

Currently access to data is not a concern as there are a multitude of data sources available which produce a wealth of information (however it may be the case that additional data sets that are currently not available may be more valuable to congestion management than those that are currently openly available). The challenge is to harness the data by processing and interpreting it both at the higher levels of trends and scenarios and at the lower levels related to the day to day management of transportation infrastructure. High

data volumes mean that it takes time to process and advanced computing technologies are required to improve response times. [3] Currently, data is used to inform trip times and route selection; however, Big Data can be used to inform predictive analyses and the development of advanced user information platforms.

This analysis requires programs and technologies that extract value from multiple data streams and historic data bases which contain data that may seemingly be disconnected from transportation but show correlations that would otherwise be hidden. It is the combined 'Big Analytics' processing of available data streams within which the true potential value of Big Data exists. [7]

2.3. Data Analysis and Analytics

Because of the volume, velocity and variety of modern data streams and historic databases there are inherent challenges in the analysis and harnessing of this information. In particular, the different data formats and languages in which data is stored may lead to difficulties in processing using data mining algorithms. [8] However, the potential rewards are impressive. The availability of 'Big Data' provides insight into actual passenger and road use behaviour, as opposed to reported behaviours and preferences which may not present the whole picture. [9]

The multilayered nature of Big Data also allows data mining programs to find correlations and convergent traveller preferences across multiple platforms such as surveillance cameras, smartphone and metro card (smart card) use, and sensors. [10] This can aid in the development of transport demand projections that take into account multiple modes of transport (private vehicles, public transport, cycling etc.), forming a big picture overview of expected travel patterns that can inform long term transport planning and infrastructure investment.

Communicating value from 'Big Data' to road users, planners and operators is crucial for the improvement of transport networks. To curtail road congestion before the congestion becomes severe, Big Analytics algorithms must be able to communicate with traffic lights and other traffic control systems when real-time congestion pre-cursors match historical information on severe congestion events. [11]

Demand projections can inform transport infrastructure investment for planners and ensure that implemented projects are catered specifically to customer demand. In the case of public transport, the provision of real-time traffic conditions and accurate wait time estimations greatly improves customer perceptions of service effectiveness. [12] As such, the benefits of using Big Data must be weighed against the effectiveness and efficiency of Big Analytics technologies, as well as the costs incurred in using these analytic procedures.

3. Options to Harness Data to Inform Congestion Management

There are three main avenues for Big Data to be harnessed to inform congestion management, namely:

1. *Mitigation of existing traffic jams through real-time data use*: Real-time information from a variety of sources (such as traffic signals, vehicle counters, CCTV streams etc) is being used in many cities across the world as a form of congestion management. This is now being expanded to include technology that communicates directly with vehicles to both send and receive data.

2. *Avoidance of traffic jams through predictive strategies*: Harnessing multiple streams of data and comparing to historic databases can allow for going beyond real time data analysis of small data streams to allow predictive algorithms to blend real time data with historical data sets on commuter habits and preferred routes to allow for predictive traffic management. The predictive ability of such platforms is enhanced by selecting indicator data streams and stores however these are not always obvious.

3. *Create sophisticated public transport routings*: By accessing multiple streams and sources of data digital platforms can produce high-resolution information which can be used to build public transport demand maps which result in the better allocation of public transport resources. Research in this area is still relatively new, however the potential to blend in numerous data streams and stores related to why travellers are using transportation services, such as shopping timings, public events, climate conditions, can create a more sophisticated understanding of factors that affect patronage.

3.1. Real-Time Congestion Management

Many well-established congestion management strategies currently use real-time data, albeit from a limited set of traditional sources. Multiple types of software exist which respond quickly to real-time changes in traffic volume, traffic movement demands, and direction of travel. In Australia, three main types of software are currently used to inform traffic control systems, namely SCATS, STREAMS and InSync. [13] SCATS is used in most capital cities and stands for 'Sydney Co-ordinated Adaptive Traffic System', which monitors real-time traffic signals and vehicle volumes to coordinate adjacent traffic signals to reduce traffic congestion and optimise traffic flow (with the option for user intervention by control system operators). The use of SCATS has been shown to correspond to a reduction in overall travel times, vehicle stops, fuel consumption and waiting times at red traffic signals. [14] STREAMS is based on the SCATS system and has been implemented in Queensland, with promising results.

InSync is another adaptive traffic control system that uses cameras installed at traffic intersections to detect and manage traffic conditions. InSync differs to SCATS in that it does not use the concept of cycle lengths, splits, and offsets, but rather uses the concept of a finite state machine which consists of all possible states within the intersection. This means that at any given moment, a specific state can be identified which will lead to an appropriate signal transition. Main Roads Western Australia has recently developed a tool called

NetPReS (Network Performance Report System), which integrates data from multiple sources and reports road network performance in terms of multiple indicators. The tool is currently limited to historical performance but is expected to be expanded in to real-time performance analysis.

In terms of data visualisation and reporting, real-time congestion reporting is so commonplace that it has become both ubiquitous and expected. The best example of this is the 'Google Maps' directions application, which incorporates real-time congestion information in a readily accessible form known to almost all first-world road users. While real-time congestion mitigation techniques have been implemented extensively across multiple cities and countries, any real-time strategy only has a limited scope to improve traffic conditions. This is primarily because it is already too late to avoid the congestion once it has been observed. Real-time mitigation strategies are often based around deterring additional traffic from moving towards this area through traffic signals or responsive road tolls, but neither strategy eliminates the existing congestion. As such, great interest is now focused on predictive strategies which seek to curtail a traffic jam before it even begins and on how data streams and stores can inform transportation infrastructure planning and investment. [15]

3.2. Predictive Congestion Management

The question is now being raised as to how effectively the use of 'Big Data' can predict traffic conditions, and the short answer is 'yes, but not perfectly, and not right now'. The long answer is that while high-resolution data collected from millions of sources may contain the necessary information to predict travel patterns and identify problem areas, the challenge is to process this information to extract the useful information and correlations to be compared to historical data in order to create a well-informed prediction of future conditions. [16]

Even with such a process no prediction will be perfect given traffic conditions are also influenced by non-linear factors such as vehicle collisions caused by abrupt changes in traffic dynamics or irrational human responses. Even a perfectly safe vehicle under perfect geometric and environmental conditions may still crash due to sudden changes in road dynamics, interruptions to the normal flow of traffic, or a disruption inside the vehicle. In order to better improve congestion prediction models, aggregated traffic flow variables (e.g. assuming vehicle speed to be equal to the speed limit) can be replaced with real-time data in order to create a more realistic model. [17]

There are currently a number of 'Big Analytics' traffic prediction systems in development. An early mover in this space is the global company HERE that processes information collected from over 2 billion traffic probes per day and compares it to historical data since 2011 using algorithms to generate predictions of road traffic congestion issues. [18] Microsoft has also developed various software to predict traffic conditions, with some software platforms taking into account unexpected traffic conditions, and have achieved promising results. [19]

3.3. Public Transport Planning and Deployment

Public transport systems are increasingly equipped with automated data collection systems, which can be harnessed along with other data streams to provide insight into passenger demand and identify optimal public transport networks, routes and connections. [12] Analysis of such data can provide information on passenger needs and behaviour, as well as provide an assessment of system performance and real-time conditions. Furthermore, such data analysis can allow road and transport organisations to quantify the costs of service deficiencies. Crucially, quantification also allows an even-handed and simulation based evaluation of possible solutions (eg. timetable synchronization), allowing each solution to be ranked based on its cost-effectiveness and user experience benefit. [12]

Two key Big Data sources are of interest for public transport networks: 1) Automated Vehicle Location (AVL) data, provided by mobile phones, and 2) Automated Passenger Counting (APC) data, provided by smart cards (metro cards), surveillance systems (ie. video cameras), Wi-Fi and Bluetooth trackers, and sensors connected to assets, signals and switches. Currently, both AVL and APC data are used for system performance evaluation. However, neither data source has been used extensively in system planning and development, making such data a largely underused resource. If harnessed, these forms of data can inform projections of passenger volumes, which are essential in the effective prediction of future demand and can act to enable the design and optimisation of transport networks. [12]

Furthermore, the analysis of data sources such as AVL and APC data can replace large, costly, and often overstated surveys on travel habit and stated preferences. Algorithms can directly construct travel demand based on observed travel patterns and provide a basis for public transport planning such as tactical planning (a mid-term plan that involves service frequency, timetabling, and vehicle and crew scheduling) and strategic planning (a long-term plan and concerned with overall network and service design such as stop positioning and line topology and capacity). [20, 21] Not only does using Big Data reduce the costs of transport surveys, it also provides more detailed, high-resolution information such as seasonal effects and within-day and day-to-day demand variations which are essential in timetabling. [22]

There is great interest in ways to mine the topology of a public transport network, and this has led to the development of tools to support this such as the 'Density Consensus Clustering' approach. [23] This approach seeks to deduce static knowledge of a public transport network by means of a GPS time series data. The proposed method is able to be developed to generate static data, manage data changes and to check on sudden detours in real time. The creators of the approach suggest that the infrastructure required to collect data using this approach is small and low-cost, comprising of one main server and on-board units for each vehicle.

Using Big Data to inform public transport systems requires the collecting and processing of multiple data streams to input

into prediction algorithms. Information is generated by analysing traffic data and public transport vehicle data through the application of machine learning techniques, which can utilise large amounts of data to reveal complex patterns. Service disruptions can also be mitigated by offline analysis of passenger behaviour during severe disruptions, which allow the adjustment of transport lines to high demand areas. [12] Ultimately, prediction algorithms built based on Big Data can be used to optimise future transport networks as well as monitor, schedule and manage disruptions in real-time.

4. The Future of Technology Enabled Transport

It is anticipated that in the near future vehicles will increasingly act as mobile computers which produce, process, and react to a constant stream of input data, including the locations of other cars and objects, real and expected traffic conditions, opening hours for car parks, the level of ridership and trip time on public transit alternatives, and optimal routes to a given destination. Vehicles will be able to harness data from various sources, such as other vehicles, road conditions, and road infrastructure, and beyond.

Vehicles that have these capabilities are being referred to as 'connected vehicles' and extensive access to multiple communications services is required for effective operation of such vehicles. [24] Vehicle manufacturers around the world are in a race to embed greater levels of technology into vehicles with the goal of eventually providing what is referred to as an 'autonomous vehicle', which is a vehicle that does not require a driver. Weather this goal is achieved or not, or in fact even preferable given concerns that it will lead to increased congestion, the majority of benefits can be reaped weather the driver takes their hands off the wheel or not.

Connected vehicles provides the potential to harness high-velocity vehicle-generated information streams and data from associated infrastructure in real time. To clarify, there are three types of vehicle related data transfers: Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I), and Vehicle to Everything (V2X).

4.1. Vehicle-to-Vehicle (V2V)

Vehicle-to-Vehicle communication allows transmission of information between vehicles, creating the potential for organisation and cooperation to prevent collisions and map vehicle information across networks to reduce congestion. It is anticipated that such a connected vehicle could intervene to prevent an accident based on information it is receiving from other vehicles and the transport infrastructure, for example braking before a collision occurs, either caused by the driver or due to the behaviour of other vehicles the driver cannot see till it's too late.

A number of car manufacturers are now testing V2V prototypes with Toyota announcing in 2016 it will increase its V2V enabled fleet test size to 5,000 vehicles in Ann Abor, Michigan. [25] However data compatibility is proving to be a challenge for the industry with the Mercedes E-Class designed to communicate with other E-Class models. In order to streamline efforts the United States Department of Transport propose to require that all new cars from 2021 have V2V technologies with a standard V2V frequency to allow cross make and mode communication. [26]

4.2. Vehicle-to-Infrastructure (V2I)

An example of Vehicle-to-Infrastructure communication is when data is sent from mobile telephones riding in vehicles to a private company (if particular software is active) to be aggregated to provide estimates of traffic travel times. However this is only the start what the concept of V2I is capable of. In the near future vehicles themselves will send sophisticated data streams that include multiple variables directly to transport infrastructure to be used for congestion management, emergency response, and predictive analysis. Visa versa infrastructure can communicate directly to vehicles to provide alerts, nominate optimal speeds to reduce congestion and travel time, and create open corridors around emergency vehicles and public transport vehicles. For instance, Audi's traffic light information system released on select 2017 Audi Q7 and A4 models will inform drivers of the timings until traffic lights change to green. [27] Third party applications such as the 'EnLighten' application provide similar information to motorists, harnessing traffic signal timings to provide travel speed recommendations, and is now being installed in BMW vehicles. [28]

Vehicle-to-Everything (V2E)

Vehicle-to-Everything (V2X) combines V2V and V2I while also communicating with pedestrians, devices and networks, essentially allowing a vehicle to communicate with all surrounding elements on a road network that may affect it. The application of V2X approaches 'Big Data' proportions and harnesses a wide array of data streams and inputs in order to provide road users with information to create safer and more efficient road networks.

One of the crucial points of connected systems is uniformity as connected vehicles should have compatible technology so that vehicles on the road can communicate with infrastructure and each other regardless of their make and model. In this vein, progress on harmonisation of technology standards internationally has delayed implementation and adoption of connected vehicles in Australia. [29] The Australian Communications and Media Authority is in the process of consultation with industry to develop a regime for the authorisation of 'Cooperative Intelligent Transport Systems (C-ITS). [30]

Transport Certification Australia and Austroads are also working on a system to ensure the security, robustness, and credibility of C-ITS systems in Australia. [31] Testing of connected vehicles has also commenced in NSW, which has implemented a C-ITS testbed in Illawarra for 60 participating heavy vehicles fitted with V2V and V2I technology which broadcasts on the 5.9 GHz radio spectrum. [32] In an internal report, Main Roads WA also recommends the staged installation of road-side units which have connection

capability, especially for planned road developments. [33]

5. Privacy Concerns

The exponential growth of mobility-related data will trigger significant changes in the transport industry accompanied by rising concerns relating to the adequacy of regulations ensuring privacy. [34] Even data that is said to be 'anonymous' may still be able to be linked to specific individual sources if cross-referenced with other sources of related data, especially as much of the data is currently shared with private companies with little accountability. Not only do traffic management centres have to tackle this issue, they also have to decide on whether the data is reliable enough. Much of this data right now has to be verified with other data sources such as sensors and camera footage or still-shots. In addition, companies may need to migrate to non-relational (NoSQL) databases to accommodate and process large unstructured data sets. These NoSQL databases usually use external security enforcing mechanisms; hence to reduce security breaches, companies have to use additional security software, reviewing security policies for the 'middleware' between the operating system and the NoSQL database, while also toughening the NoSQL database itself to match its counterpart relational databases. [35]

The multi-tiered nature of Big Data means that transaction logs are stored in multi-tiered media. In smaller datasets, IT managers can manually move data between tiers, giving them a measure of control; however, as the dataset grows exponentially, auto-tiering is likely to become increasingly necessary for big data storage management. As auto-tiering does not keep track of where the data is stored, unauthorised access to data stores is less easy to detect and data breaches may occur. Thus, new mechanisms must be developed to prevent data theft and maintain the 24/7 availability. [36]

In Australia, the Privacy Act regulates and protects personal information, including the Australian Privacy Principles (APPs) which define the standards, rights and obligations in relation to handling and assessing personal information. Big Data changes how key privacy principles—which include data collection, minimisation of data retention and use limitation—are applied. However, as the APPs are technologically neutral, corporations and other organisations can adapt their Big Data handling policies to protect personal information while also retaining the maximum use from the information derived from Big Data analysis. [37]

According to the Privacy Act, organisations must take reasonable steps to implement practices, procedures and systems that protect personal information. These organisations must also be able to deal with privacy related complaints from individuals. A systematic risk management approach must be used to identify reasonable steps according to the size of the entity, its resources and the complexity of its operations. Organisations dealing directly with Big Data must take more rigorous and detailed privacy protection procedures than an entity handling the results of Big Data analytics. [38]

One possible technique to ensure privacy is to use de-identification to remove the personal identifiers such as addresses and date of birth, as well as any other unique individual characteristics. This means that the Privacy Act no longer applies. [38] However, this technique is not foolproof and if de-identified datasets are matched to other datasets or other information, it is possible that individuals can be re-identified. [39]

6. Conclusions

As vast amounts of data continue to explode across a huge array of platforms, Big Data is increasingly relevant in providing high-resolution information to optimise transport systems. The exciting implications of Big Data have yet to be fully realised with most traffic management systems drawing on real-time, for-purpose 'small data'. Big Data has a high potential to prevent congestion and has achieved substantial returns, however its realisation is not without challenges.

Although Big Data can provide key information to evaluate, plan and improve transport systems, the key challenge in its utilisation is the fact that the extensive volume of information requires multiple modes of data analysis and processing. Because so much information is available, software and programs must be developed which can sift out irrelevant information and focus on key features of the data which will provide necessary inputs into transport prediction patterns.

However, due to the scale of data, data variety and rapid frequent changes, it is a challenging task to integrate, visualise, analyse and respond to queries. Current data analytics systems provide limited analysis capabilities with long response times of several minutes, which is an impediment for real-time data analytics. Recently, in-memory computing techniques have been found to achieve significantly higher efficiencies, with processing speeds of approximately one second. Multiple IT firms are actively working in this field, with researchers currently investigating new methods to improve processor speed and responsiveness.

Further, when using Big Data for future transport volume projections, highly specialised and accurately calibrated data mining programs must be used in order to develop accurate and robust projections, because the sheer volume of information available makes analysis difficult. The algorithms and projections developed using Big Data must also be properly calibrated against real-life transport volume scenarios in order to ensure that the projected system performance is sufficiently accurate.

These challenges must be overcome in the future in order for Big Data to be accurately, effectively and efficiently harnessed for congestion management and emergency response. Yet as a whole, what transport planners stand to win is far greater than what they could lose, in particular enhanced systems for managing transport systems and the timely prediction of when bottlenecks will occur to allow transport planners to devise methods to prevent congestion in these areas, effectively deferring capital investment in

transport system expansion.

References

[1] Texas A&M Transportation Institute 2013, 'As traffic jams worsen, commuters allow extra time for urgent trips', *Texas A&M University*, February 5.

[2] Bureau of Infrastructure, Transport and Regional Economics (BITRE) 2015, *Traffic and congestion cost trends for Australian capital cities*, Commonwealth of Australia, Canberra. (ISBN 978-1-925216-99-8)

[3] Mullich, J. 2013, 'Drivers avoid traffic jams with Big Data and Analytics', *Bloomberg L. P.,* New York.

[4] Hotz, R. L. 2011, 'The hidden toll of traffic jams', *The Wall Street Journal*, November 8.

[5] BITRE 2014, 'New traffic data sources – An overview', *Australian Department of Infrastructure and Regional Development,* Canberra.

[6] Transmax 2015, 'Streams ITS', Transmax Pty Ltd 2015.

[7] International Transport Forum 2015, 'Big Data and Transport', *International Transport Forum.*

[8] Hassanien, A. E., Azar, A. T., Snasael, V., Kacprzyk, J., Abawajy, J. H. (ed) 2015, op. cit. ISBN: 978-3-319-11056-1.

[9] van Oort, N. & Cats, O. 2015, 'Improving public transport decision making, planning and operations by using Big Data: Cases from Sweden and the Netherlands', *IEEE 18th International Conference on Intelligent Transportation Systems*. DOI 10.1109/ITSC.2015.1.

[10] Carter, K. B. 2014, Actionable Intelligence: A guide to delivering business results with Big Data fast!, John Wiley & Sons, Singapore. (ISBN: 1118920651)

[11] Sawyers, P. 2015, 'How Microsoft's using big data to predict traffic jams up to an hour in advance, Venturebest, April.

[12] van Oort, N. and Cats, O. 2015, op. cit. DOI 10.1109/ITSC.2015.1.

[13] Fernando, B., Gray, E., Kellner, J. 2013, 'A review of current traffic congestion management in the City of Sydney', *Infrastructure Australia*, Canberra.

[14] Main Roads WA 2014, 'SCATS', *Government of Western Australia,* Perth.

[15] Lu, H. P., Sun, Z. Y., Qu, W. C. 2014, 'Big Data-Driven Based Real-Time Traffic Flow State Identification and Prediction', *Discrete Dynamics in Nature and Society*, vol 1. (DOI: 10.1155/2015/284906)

[16] Jie Xu et al. 2015, 'Mining the Situation: Spatiotemporal Traffic Prediction with Big Data', *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4.

[17] Hossain, M. 2012, 'A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways', *Accident Analysis & Prevention,* vol. 45. (http://dx.doi.org/10.1016/j.aap.2011.08.004)

[18] Highway Engineering Australia 2015, 'Big data: The key to unlocking the future of traffic, transport and infrastructure', *HEA*, vol. 47, no. 2, pp. 40-41.

[19] Horvitz, E. 2011, 'Predictive Analytics for Traffic', *Microsoft Research*, Sept 26.

[20] Ma, X., Wu, Y. J., Wang, Y., Chen, F. and Liu, J. 2013, 'Mining smart card data for transit riders' travel patterns', *Transportation Research Part C: Emerging Technologies*, Vol. 36, pp. 1-12, http://dx.doi.org/10.1016/j.trc.2013.07.010.

[21] Zhao, J., Rahbee A. and Wilson, N. H. M. 2007, 'Estimating rail passenger trip origin-destination matrix using automatic data collection systems', *Computer Aided Civil and Infra. Eng.*, Vol. 22, pp. 376-387. (DOI: 10.1111/j.1467-8667.2007.00494.x)

[22] Berkow, M., El-Geneidy, A. M., Bertini, R. L. and Crout, D. 2009, 'Beyond generating transit performance measures', *Transportation Research Record*, Vol. 2111, pp. 158-168. (DOI: 10.3141/2111-18)

[23] Fiori, A., Mignone, A., Rospo, G. 2016, 'DeCoClu: Density consensus clustering approach for public transport data', *Information Sciences*, vol. 328, no. 1, pp. 378-388.

[24] Weeratunga, K. and Somers, A. 2015, Connected vehicles: Are we ready?, Main Roads Western Australia internal report, Perth.

[25] Asian Development Bank 2016, 'Safety and Intelligent Transport Systems Development in the People's Republic of China', *Asian Development Bank,* Section B. 8.4.

[26] National Highway Traffic Safety Administration 2016, 'US DOT advances deployment of connected vehicle technology to prevent hundreds of thousands of crashes', *United States Department of Transportation*, Washington DC.

[27] Herndon, Virginia. Audi USA Homepage, Press Release (August 15, 2016) "Audi announces the first vehicle to infrastructure (V2I) service – the new Traffic light information system".

[28] Zurschmeide, J. 2015, 'Stop wasting gas in the city with the app that knows when traffic lights will change', *Digital Trends*.

[29] ACMA 2016, 'Proposed regulatory measures for the introduction of C-ITS in Australia', Australian Communications and Media Authority, Australian Government, accessed 25 Jan 2017.

[30] ACMA 2016, 'Proposed regulatory measures for the introduction of C-ITS in Australia', op. cit.

[31] TCA 2017, 'Cooperative Intelligent Transport Systems (C-ITS)', Transport Certification Australia, Australian Government, accessed 25 Jan 2017.

[32] Centre for Road Safety 2016, 'Cooperative Intelligent Transport Initiative', *NSW Government*, accessed 25 Jan 2015.

[33] Weeratunga, K. and Somers, A. 2015, Connected vehicles: Are we ready?, op. cit.

[34] International Transport Forum 2015, 'Big Data and Transport', op. cit.

[35] CSA 2012, *Top ten Big Data Security and Privacy Challenges*, Cloud Security Alliance, Rolling Meadows, Illinois.

[36] CSA 2012, *Top ten Big Data Security and Privacy Challenges*, op. cit.

[37] OAIC 2015, *Consultation draft: Guide to Big Data and the Australian Privacy Principles*, Office of the Australian Information Commissioner, Canberra.

[38] OAIC 2015, *Consultation draft: Guide to Big Data and the Australian Privacy Principles*, op. cit.

[39] OAIC 2016, *Privacy business resource 4: De-identification of data and information*, Office of the Australian Information Commissioner, Canberra.