# BusViz: Big Data for Bus Fleets

**Afian Anwar** (Corresponding Author)
Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology
32 Vassar Street
Cambridge, MA 02139 United States of America
afian@csail.mit.edu

**Amedeo Odoni**
Operations Research Center
Massachusetts Institute of Technology
77 Massachusetts Avenue
Bldg. E40-149, Cambridge, MA 02139
United States of America
arodoni@mit.edu

**Nelson Toh**
Public Transport Quality Division
Land Transport Authority of Singapore
10 Sin Ming Drive, S(575701)
Singapore
nelson_toh@lta.gov.sg

5340 words + 7 figures + 0 tables
November 15, 2015

## ABSTRACT

Transit agencies in major cities make hundreds of decisions every day on how to best allocate scarce resources. How adequate and reliable is service on a given bus route? When to add capacity to a popular route? How should a bus route service a neighborhood? Transit planners make decisions by using whatever information is available. This includes ground surveys, public feedback and a large element of experience and intuition. In this paper, we present BusViz, a web-based application that helps bus service operators and transit regulators to make better decisions by using large streams of field data to monitor and visualize the performance of bus fleets. We describe the system architecture and user interface of the application, and illustrate how the Singapore Land Transport Authority (LTA) is using it to iterate and evaluate ideas with data driven rigor and share quality of service statistics with various stakeholders.

## INTRODUCTION

In a fast growing number of cities around the world, data about a passenger's trip is captured by a wide array of embedded tags and sensors every time he or she boards a bus, takes a train or hails a taxi. On board Global Positioning System (GPS) devices, electronic fare cards and surveillance cameras produce a rich stream of data recording vehicle arrivals, locations, travel times and occupancy. This explosion of Big Data, produced routinely in the course of everyday operations, opens up new possibilities for transport operators and regulators to understand network performance, monitor service levels, such as crowding and delays, and make adjustments and improvements as necessary.

In this paper, we describe BusViz, a web-based tool that helps transit agencies monitor and improve the performance of bus fleets. BusViz uses passenger fare card data to create a dynamic digital replica of the operation of a city's bus transportation system. The data come from a simple marker - when people tap their fare cards to get on or off a bus, they leave a digital trace of their trip's origin and destination. We use this marker to infer crowdedness (Figure 1), waiting times, bus arrival frequency and much more, thereby allowing transit agencies to extract detailed performance metrics from their bus fleets. In addition to providing a valuable tool for operators of bus services, BusViz also addresses an urgent, unmet need of many regulatory agencies worldwide which have restructured their public bus services to a government contracting model where in exchange for bus operators agreeing to meet service and reliability standards, the transit agency subsidizes operations by funding - partially or fully - the purchase and maintenance of new buses and, in some cases, even driver salaries.

For such arrangements to work, transit agencies need to be able to monitor and enforce key performance indicators on their operators. We present BusViz as an example of a new generation of web-based tools that provide critical assistance to bus system operators and transit regulatory agencies in performing such tasks efficiently and inexpensively, without the use of surveys, and in identifying ways for substantially improving service.

The main contributions of the paper are:

- a review of current approaches and limitations of existing methods used by transit agencies to monitor and analyze reliability and quality of service
- a description of BusViz, a web-based visualization and analysis tool that uses passenger fare card data to allow transit agencies to better understand bus network performance and behavior and
- a case study of how this tool is used by a large transit agency to: reduce reliance on expensive and labor intensive surveys; identify possibilities for improving service; and share and disseminate quality of service information and statistics to stakeholders.

## RELATED WORK

Measuring bus fleet reliability and quality of service is particularly important to transit agencies because this information is used to inform resource allocation policies and investment decisions into capital resources and operational improvements. Traditionally, this information has been collected using surveys conducted by human observers who typically position themselves at bus stops to measure bus arrival times and passenger overcrowding. For example in Singapore, bus passenger surveys are routinely used to respond to public inquiries and feedback and verify that instances of poor service reported by commuters are not one-off events caused by unusual traffic or weather conditions (1). In general, two kinds of surveys can be identified. One is based on passenger satisfaction, where passengers are individually polled and asked questions about their transit experience (2, 3). The other focuses on compiling vehicle journey logs, whereby hired observers board vehicles and take note of travel time, ridership, and punctuality during their trips (4, 5). Analysts then use multi-criteria evaluation techniques on such survey data to make recommendations or identify operational deficiencies and areas of improvement (6). Alternatively, some transit operators have installed video cameras on buses (7) and bus stops (8) in an effort to automate data collection.

These methods are popular because they present an administratively simple way for agencies to obtain "ground truth" data but, as noted in (9), they are costly, time-consuming and limited to small sample sizes. In contrast, by using individual passengers as a virtual distributed network of sensors, bus operators and transit agencies can collect bus service information and statistics at any point on a bus route without the use of expensive fixed equipment or human surveyors.

Our approach is inspired by recent academic efforts that use operational transit data to develop and test a wide variety of performance measures. References (10) and (11) demonstrate how speed, travel time, and intersection delay data collected by buses could be used to accurately characterize general traffic flow. References (12) and (13) describe how such data can be used to estimate operating efficiency and service reliability with a high degree of accuracy, while (14) relates service reliability to passenger demand and (15) evaluates arrival predictions using ground truth bus location data. Similar to other web-based, open-source planning tools such as Open Trip Planner (16) and Transit Mix (17), BusViz builds on this body of work by making it easy for agencies and operators to assess the quality of the connectivity of bus networks and their reliability.

## MOTIVATION

We next motivate the need for BusViz by illustrating how automated tools that turn data into actionable information can help transit agencies better monitor and communicate reliability and quality of service. BusViz was developed over a period of several months as part of a research collaboration with the Land Transport Authority (LTA) of Singapore. We observed that even though LTA collected a large quantity of fare card data each day, its planners still relied on inefficient tools to plan and manage Singapore's bus transportation network. As a result, they did not necessarily have a holistic and complete assessment of network health. Some planners used markers to draw bus routes on street directory maps. Others relied on spreadsheets and Excel macros to keep track of service performance and customer complaints. Even when data were available, LTA staff often had to write custom SQL to retrieve the relevant information. It was not uncommon for such queries to "time out" because the amount of data returned was too large for LTA's servers to handle. Two specific instances requiring heavy data analysis are those when the Authority has to (i) respond and act on public feedback or (ii) monitor whether the performance of private transport operators is consistent with their contractual obligations.

### Use Case: Acting on Feedback

In many cities such as London and Singapore, the local transit authority plans routes and establishes minimum service standards for bus lines managed by commercial operators (18). They must ensure that these minimum service standards are met, and investigate complaints from the general public. In a typical scenario, complaints by members of the public are handled by a local political representative, who acts as liaison between his/her constituency and the transit authority. The local representative meets with the transit authority regularly to provide feedback and lobby for better service (e.g. increased frequency for a local feeder route). To evaluate a request, the transit authority must identify the bus stops, if any, with poor service, request recent data from the bus operating company, process the data in-house to extract performance metrics manually at the offending bus stops and finally, produce a report summarizing such metrics as on-time performance, average headway, service reliability and average passenger load in table form. The results are then compiled and shared with the representative, and if, action is necessary, brought to the attention of the bus company. This process can take several months during which service could deteriorate further as operational problems remain unresolved. If operational data is not available, a survey company is typically engaged to record statistics at the affected bus stops.

**Use Case: Contractual Performance Monitoring**

It is common for transit authorities to specify quality of service targets that must be achieved by transport operators in their jurisdictions. Under a model that has been applied successfully in London, Stockholm, Copenhagen, Seoul and Perth (19) prospective operators of transport services bid for sets of bus routes in an urban market and receive certain subsidies in exchange for meeting the agreed targets. This gives operators revenue certainty while allowing the transit authority to dictate performance metrics as they see fit (20). Typically, the contract between the regional transit authority and bus operators specifies that the operator will pay penalties or earn bonuses (21) depending on the level of service offered (22). But for this system to work, the transit authority must have the capability to monitor and enforce key indicators of the performance of the operator.

Singapore offers an example. In exchange for bus operators agreeing to meet service and reliability standards, Singapore's government recently overturned a longstanding policy of no subsidies for public transport operations by funding the purchase, maintenance and driver salaries for over 1000 new buses (23). The specified standards are tracked using a combination of bus GPS data, smart card records and on-site audits (24). However, analyzing the data can be difficult and time consuming without using efficient software tools.

## TECHNICAL APPROACH

As with a lot of publicly held data, the issue in the above Use Cases is not that information on bus arrivals, occupancy and service reliability does not exist - it is just that it is not very accessible or easy to process, visualize and understand. BusViz addresses this problem by building an API (Application Programming Interface) layer that connects passenger fare card data to a web-based visualization layer, thereby allowing LTA to query their data at scale.

BusViz consists of three main components. The first is the algorithmic layer, which is used to perform multi-pass computations that mine the EZ-Link data stream for transactions. In Singapore, whenever a passenger taps an EZ-Link (25) fare card to board a bus, her position, time of boarding and the license number of the bus are stored in an onboard computer and uploaded to a central EZ-Link server when the bus returns to its home depot. By keeping track of where, when and how each passenger got to his final destination and grouping this data by origin-destination pairs, we can tell exactly how many people took buses from one location to another. By storing a running total of the number of people on each bus, we can infer how long commuters waited on average, how many buses went by that were so full commuters could not get on, and what the passenger volume of each bus was throughout the day. The computing power now available means that our algorithms are able to process this data at a scale that was previously unimaginable even a few years ago, allowing us to quickly turn raw fare transaction data into useful, actionable information for LTA. In practice, this preprocessing step is handled by EZ-Link on their proprietary backend and forwarded to PLANET (Planning for Land TrAnsport NETwork), LTA's enterprise data warehouse project.

The second component is the API layer, which is responsible for connecting data stored in PLANET to our algorithms for batch processing. Our API layer does so by performing an integration with PLANET. We expose an API endpoint in PLANET that our API layer can use to pull data at scheduled intervals. We process this data and store the results in a separate, high-fidelity database. This enables analytics built on this data to be served to a large number of concurrent users without affecting the live PLANET server. Alternatively, users can also download PLANET data as a CSV file and upload it to BusViz using its data import functionality.

Finally, the third component is the front-end visualization layer. This layer accesses data stored in the analytics database to display finely tuned Javascript- and HTML-based visualizations on a web browser. In doing so, bus passenger data is mapped to visually intuitive objects that interact and animate to tell a story, thereby allowing users to discover patterns and switch effortlessly between different spatiotemporal views, boundaries and scales. Because it operates entirely in the cloud, the visualization layer can be accessed securely on any web browser, whether from a desktop computer, tablet or smartphone. This simplifies the development process considerably because one no longer has to account for different operating systems, and maximizes interoperability, by allowing BusViz to run on secure computing environments (typical of government agencies) that have restrictions on installed software.

## DATA

The EZ-Link card is widely used in Singapore as a stored-value, contactless smart card to pay for public transportation, parking and even retail transactions. Passengers may still pay cash for public transport rides but are charged more. The overwhelming majority of passengers thus opt for the convenience and cost savings provided by EZ-Link. Some 96% of all the public transportation trips in Singapore are made with the card (26). Since each card ID is linked to a single person, the associated data can be used to obtain detailed assessments of travel behavior and city-scale mobility patterns.

By assuming that each person correctly taps their card when they board and alight, we can equate the arrival time of a bus at a stop to the earliest card entry time of the first passenger who boards or alights at that stop. Similarly, the departure time is set equal to the greatest of the card entry times of passengers who board or alight at the stop. If no passengers alight or come on board at a certain bus stop, we estimate the arrival time by linearly interpolating arrival times at the bus stops immediately before and after. Fortunately, since Singapore's bus system is heavily utilized, such occasions are rare.

The data for each trip is stored locally on the bus's on-board computer unit (OBU). When buses return to their depots at the end of each trip, the data is downloaded from the OBU to a central server via WIFI. The log files are then consolidated centrally and stored on a shared database.

Using the algorithms described in (27, 28, 29), the BusViz API layer aggregates the data into a new format, recording the time-stamp, bus route ID, bus ID, bus stop ID of the current bus stop, direction (inbound or outbound), maximum capacity and number of passengers on each bus. Records are logged each time the bus arrives at a bus stop, thereby allowing us to track the approximate position of each active duty bus over the course of a day. Since each record contains the unique ID of the bus and the bus stop ID, we can cross-reference this information with an index of the GPS locations of each bus stop to reproduce the bus's trajectory.

## BUSVIZ WEB APPLICATION

In this section, we describe the features and design considerations of the front end BusViz web application that LTA's transit planners use to highlight trends, identify recurring problems in service quality and calculate bus stop level reliability and performance metrics. At this stage of its development, BusViz serves as a convenient, fast and highly intuitive management information system and provides a tool for experienced users to identify problems visually ("eyeballing") and seek solutions to them. The visualizations are written in Javascript and done entirely in the web browser, using a combination of Three.js (30) (a wrapper for WebGL) for 3D rendering, D3.js (31) for charting and mapbox.js (32) for mapping.

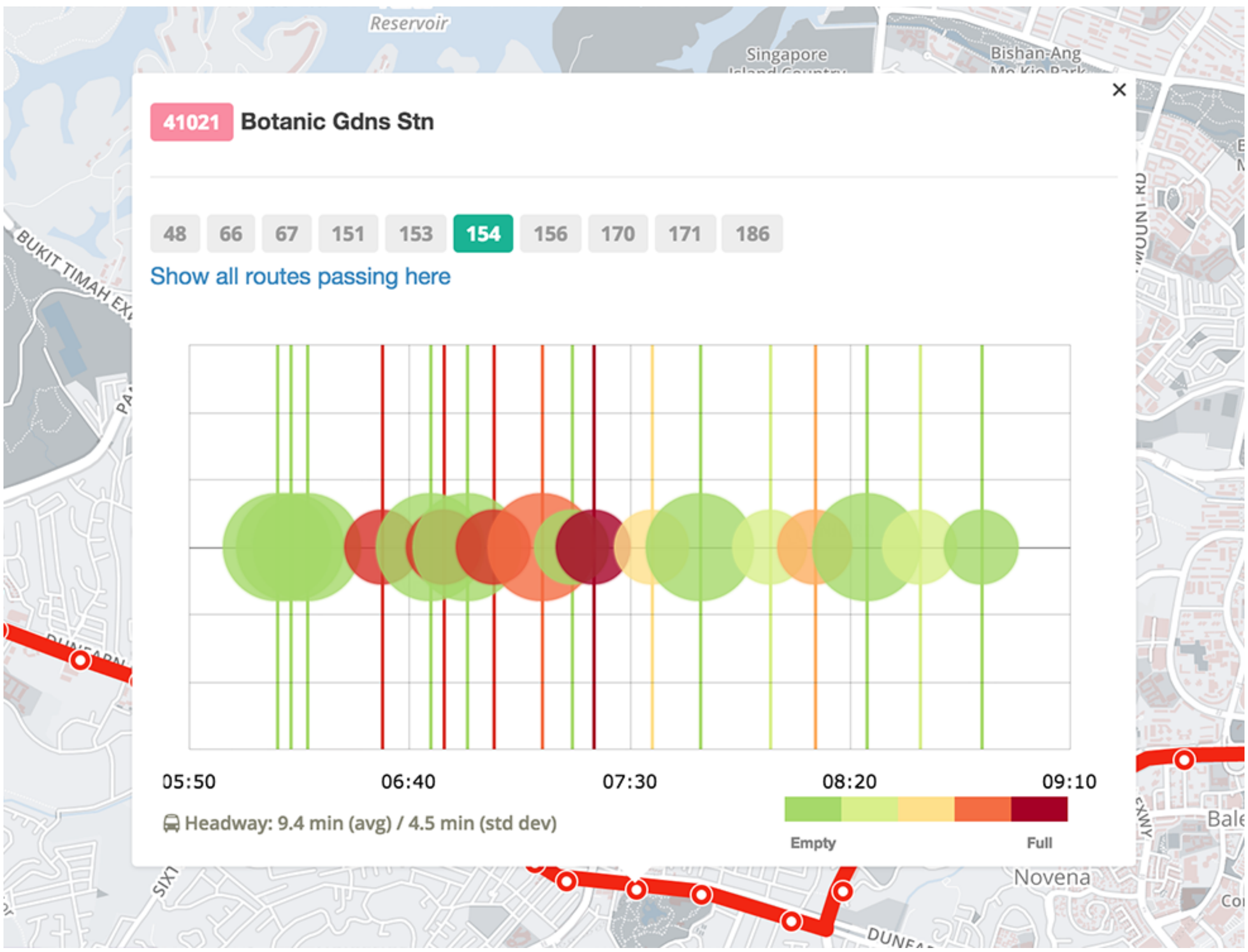


**FIGURE 1** : BusViz's bus stop view showing the number of people on board buses arriving at a bus stop. Bus loads are visualized as colored circles: dark green (empty) to dark red (completely full). Large circles represent double deck buses and small ones, single deck buses. Headway average and standard deviation is shown below the chart.

### Bus Stop View

The *Bus Stop View* (Figure 1) reduces the need for transit authorities to place human surveyors at bus stops to monitor bus arrivals and passenger loads. Users can add bus services to an online map to display the direction and physical path that each bus service takes (Figure 3).

Users have the choice of selecting a single bus service or selecting origin and destination bus stops to retrieve a set of common corridor ("overlapping") bus services that run in parallel with each other. Many popular routes or route segments in Singapore are served by more than one bus service. Considering bus services in isolation underestimates actual service availability. To identify overlapping bus routes between any two specified bus stops, we store the sequence of bus stops from each bus service as a string and find the longest common substring (33) starting and ending with our origin and destination bus stops. We then return all bus services that have at least two bus stops in common (Figure 2).

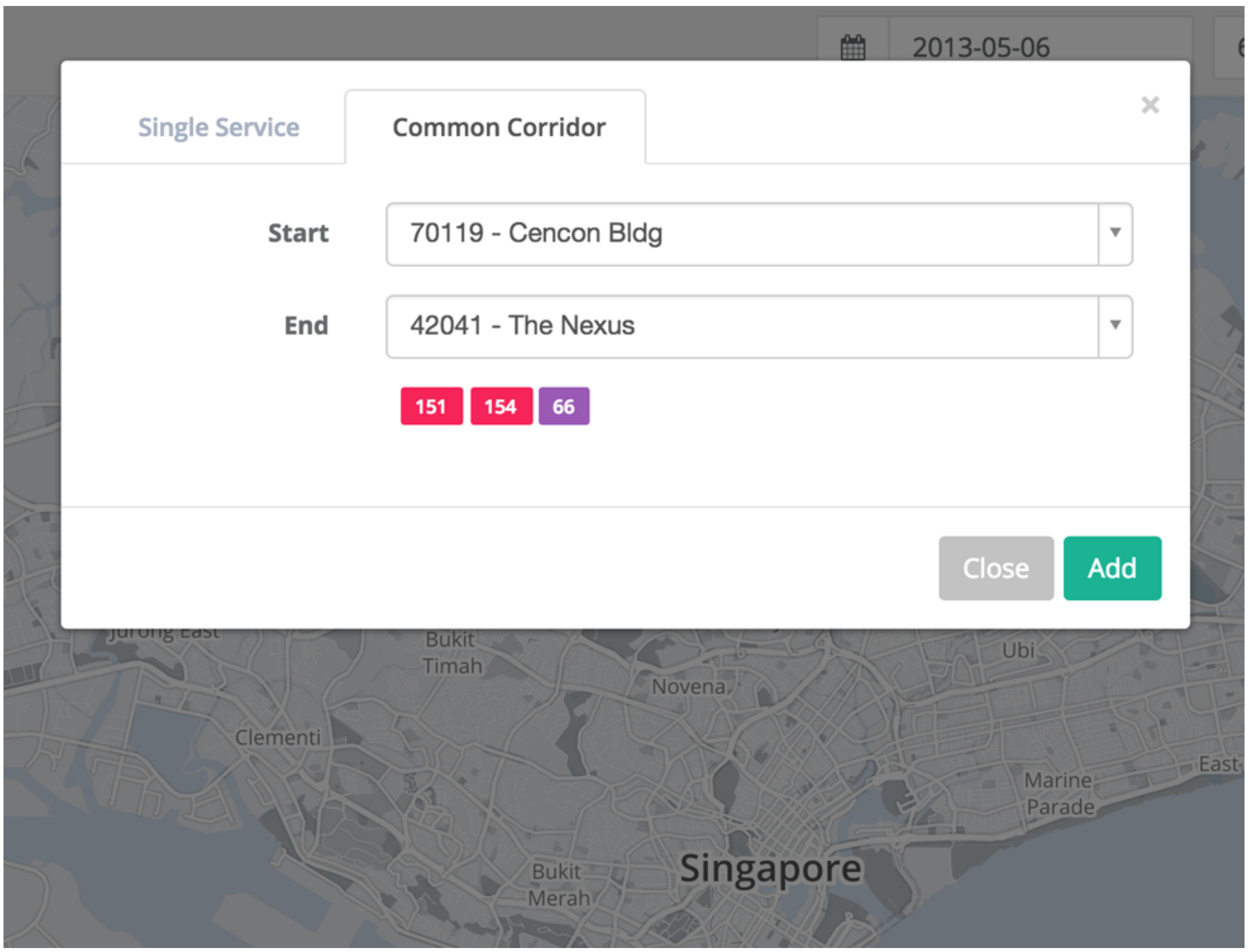


**FIGURE 2** : Common corridor bus services are automatically retrieved from a user specified origin and destination bus stop.

Selecting a bus service highlights the bus stops on its route and clicking on a bus stop produces a pop up (Figure 1) showing a visualization of the number of people on board buses arriving at a bus stop. Bus loads can also be visualized as colored circles, as in Figure 1: dark green (empty) to dark red (completely full). Useful statistics such as the average and standard deviation of headways are shown below the chart. By selecting different days and time periods, users can see how service performance at the bus stop level varies.

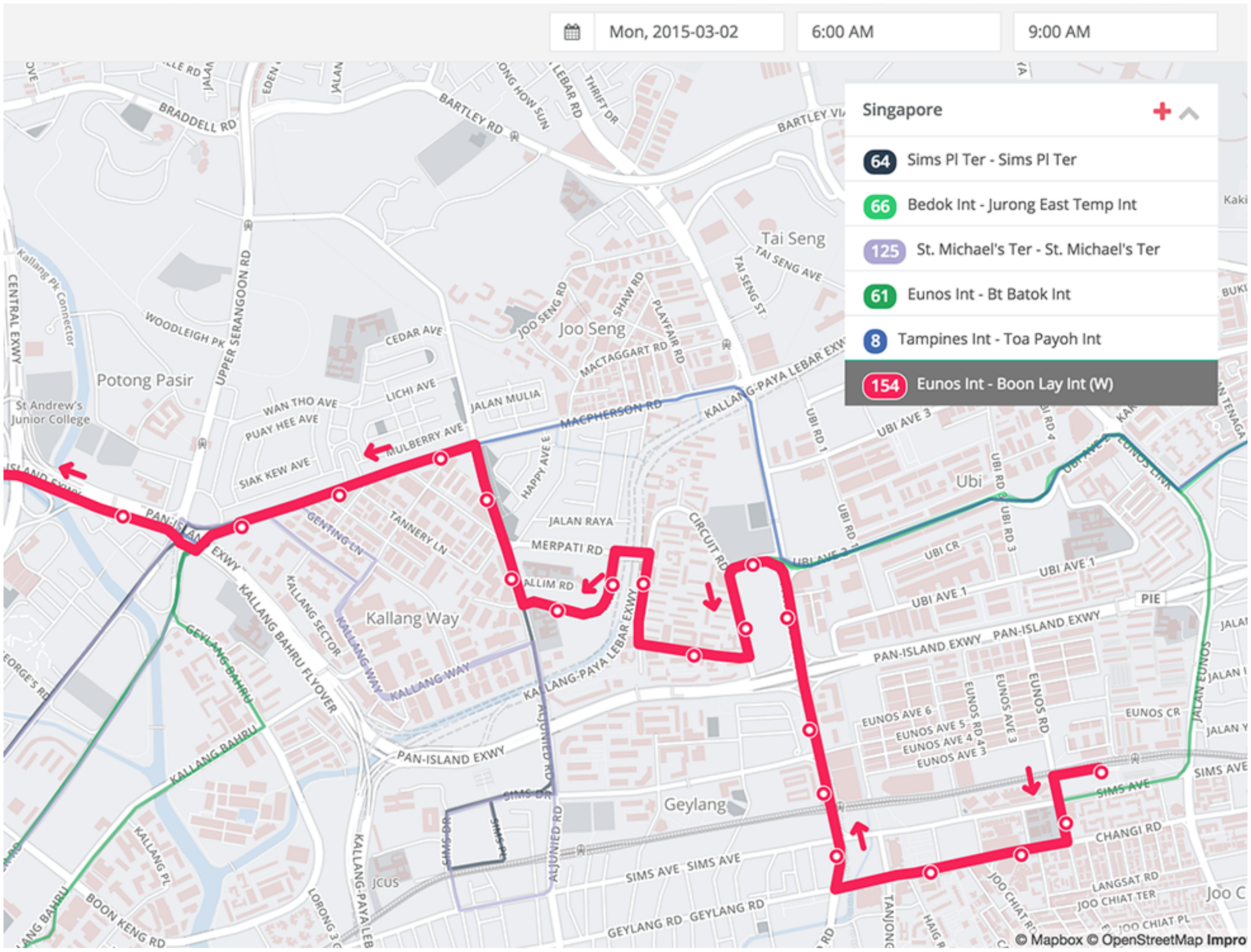


**FIGURE 3** : Bus services can be added to a map and when selected, BusViz shows the service's route direction and bus stops. Multiple services can be added and overlaid on top of one another.

### Time Space Diagram View

The time-space diagram is a tool used widely by planners for the design and analysis of transportation systems. When referring to bus services, time-space diagrams typically track the "trajectories" of individual buses on their routes, with each point (x, t) on a trajectory indicating the time $t$ at which a point $x$ along the route is reached. Thus, by drawing a vertical axis over the location $x_S$ of stop $S$, the time-space diagram identifies the times at which successive buses pass a bus stop. The time separations between consecutive bus arrivals at a given location are the "headways" and the distance separations between consecutive buses at a given instant the "spacings". By studying how the absolute and relative positions of buses change over time, planners can understand the performance characteristics of a single bus trip or of a bus route or of any specified set of bus routes.

Until very recently, the time-space diagrams used by technically advanced transit agencies such as the LTA were created through spreadsheets or commercial business intelligence software

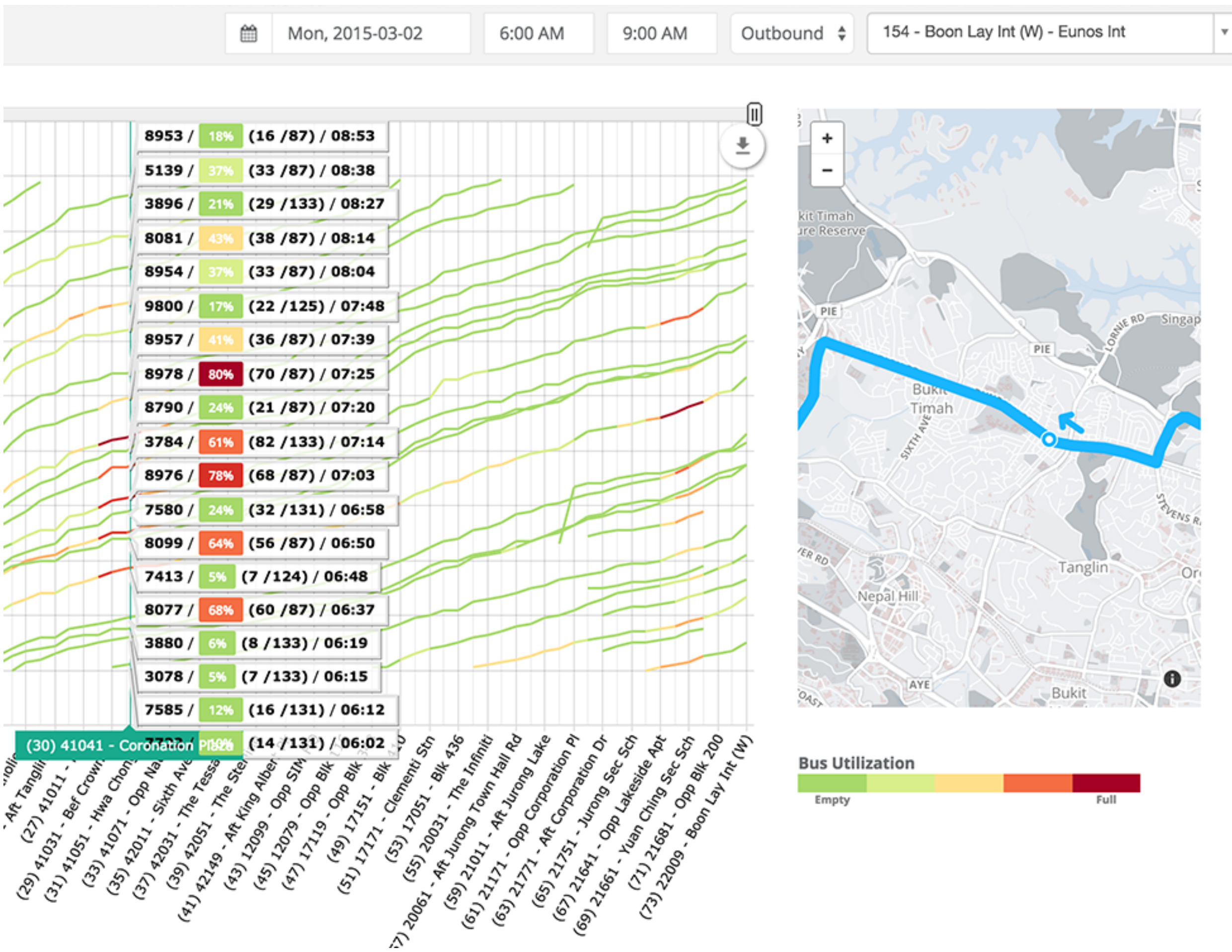


**FIGURE 4** : The Time-space Diagram view summarizes the performance of a bus service along it's route. Overlapping trajectories indicate bus bunching while orange and red lines indicate congestion. When a bus stop is clicked on the chart (left) its location is shown on the map (right).

such as Tableau. They make it possible to identify some specific performance characteristics e.g. the presence of bus bunching when observing overlapping bus trajectories, but more sophisticated analysis e.g. relating bus bunching to passenger loads or detailed examination of morning peak periods is cumbersome or impossible.

The BusViz *Time Space Diagram View* (Figure 4) improves on this by offering users the opportunity to generate instantaneously any number of displays with the option to specify the time period, geographical area and set of bus routes to be studied. Moreover, it produces dynamic time-space diagrams that color bus trajectories according to the number of passengers on board (dark green for empty, dark red for completely full). This allows users to immediately identify critical route segments and the times when a route is congested. Hovering over a bus stop on the time-space diagram brings up detailed statistics (arrival time, passenger load, bus capacity) for each bus arrival (Figure 4). Clicking on a bus stop shows its location and surroundings on a map, thereby allowing the user to identify physical characteristics in the area that may be contributing to the problem. To facilitate detailed analysis, the time-space diagram can be zoomed-in, scrolled and panned to show specific segments of the route. Moreover, as is also the case with the *Bus Stop View*, users can pull data from different combinations of days, times and bus services to obtain statistically meaningful samples of operational performance.

We provide below two simple examples of specific BusViz Time-Space Diagram application contexts that are already used extensively. The first (Figure 5) shows a Time-Space Diagram for a typical day's operations on a long bus route in Singapore. Clearly, significant bus bunching and bus overloading occurs in segments of the route after roughly 06:30. Inspection of the diagram immediately suggests a potential solution that would consist of increasing the frequency of service by about one bus run per hour between 06:00 and 09:00 and re-adjusting bus departure times from the origin. Conversely, analogous Time-Space Diagrams can help identify bus routes where the overall quality of service, as measured by waiting time for buses and by bus loads, would not be significantly reduced by modest reductions in bus frequencies and appropriate adjustments in bus departure times.

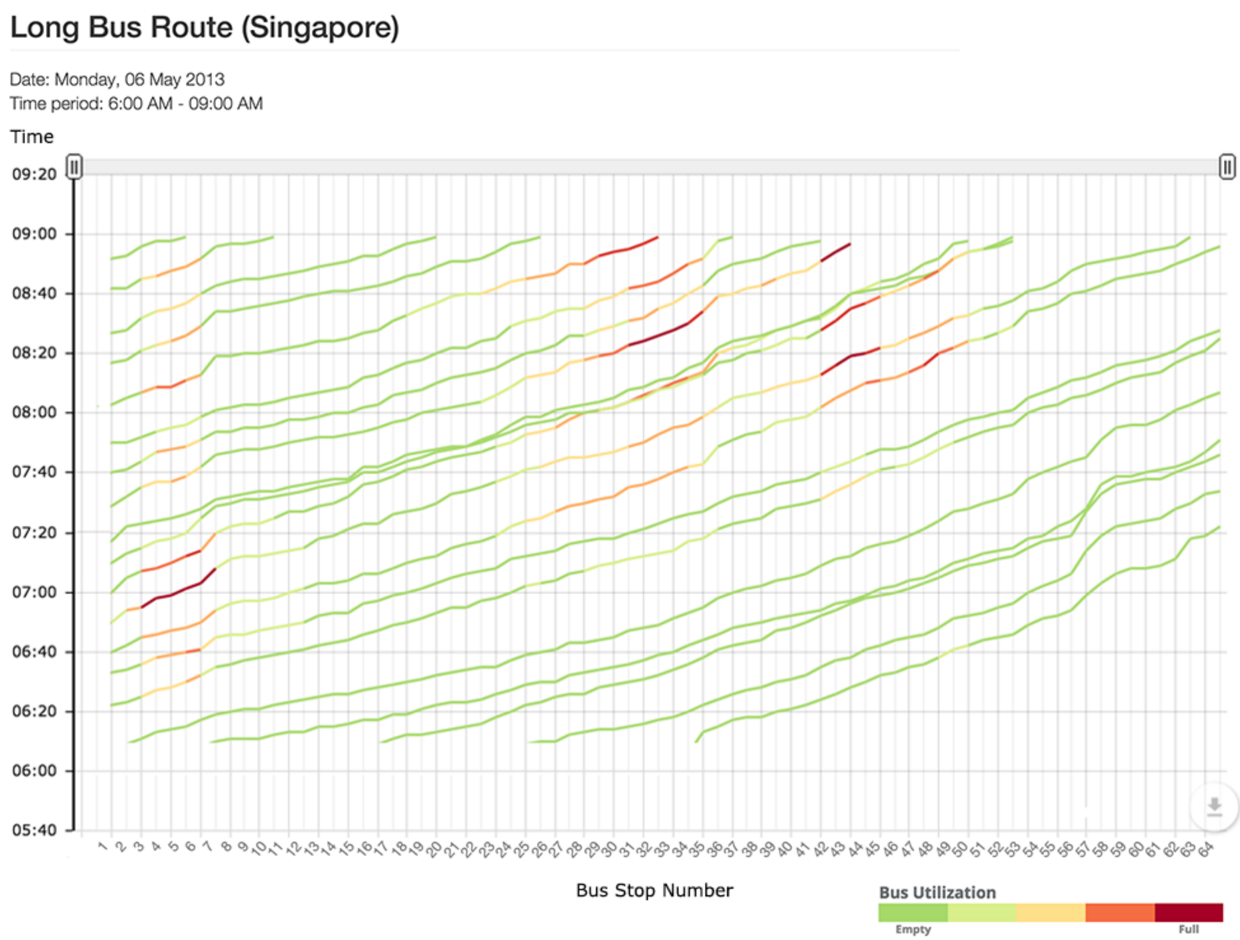


**FIGURE 5** : The Time-space Diagram view summarizing the performance of a long (64 stop) bus service along its route. Significant bus bunching and overcrowding can easily be seen at bus stop #30 at 8.20 am and at bus stop #44 at 8.00 am.

The second example highlights the usefulness of BusViz as a diagnostic tool. Figure 6 shows a BusViz display of the standard deviations of headways between bus arrivals at all the

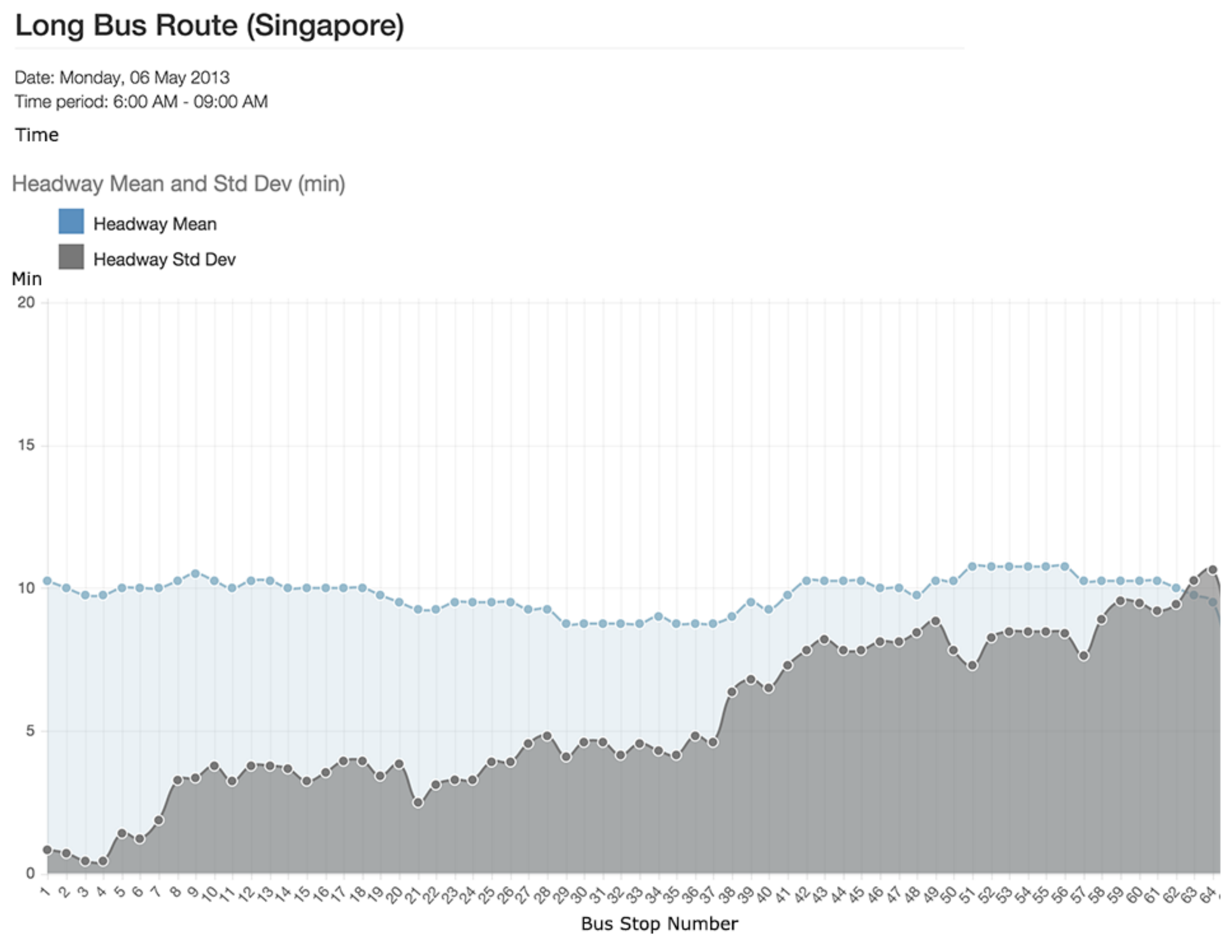


**FIGURE 6** : Headway mean and standard deviation of a long (64 bus stop) bus service in Singapore. As expected, headway mean stays constant (about 10 min) along the route but the headway reliability (as measured in standard deviation) steadily deteriorates from 0.8 min to 10.3 min at the end of the route.

stops of a popular bus route (the same as in Figure 5). The average headway between buses on this route is about 10 minutes. It can be seen that the standard deviation of the headways increases from about 0.8 minutes at stops near the origin of the route to about 10.3 minutes at stops nearer its end. This means that headways become increasingly unpredictable and unstable along the route, with negative implications for level of service, as measured by average waiting time at bus stops. It is well known that the average waiting time for buses depends not only on the average value of the headway, but also on its "regularity", as measured by the standard deviation. This is particularly true in the case of bus routes with high frequency of service, where prospective passengers tend to arrive at bus stops independently of the detailed bus schedule - especially, if headways are highly unpredictable in the first place. In such cases (see, e.g. (34)) we have:

$$E[W] = \frac{E[H]}{2} + \frac{\sigma^2(H)}{2E[H]} \tag{1}$$

where $E[W]$ and $E[H]$ are the expected ("average") values of, respectively, the waiting time at a stop $W$ and the headway, $H$, between bus arrivals and $\sigma^2(H)$ is the headway variance. Note that, when the standard deviation of the headways, $\sigma(H)$, is large, the second term on the right contributes significantly to average waiting time. For instance, when $\sigma(H) \approx E[H]$, the average waiting time $E[W] \approx E[H]$, or twice the "naive" estimate of $E[W] \approx \frac{E[H]}{2}$. Applying (1) to the bus service in Figure 6, setting $E[H] = 10$ min, $\sigma(H) = 0.8$ min near the beginning of the route and $\sigma(H) = 10.3$ min near its end, we see that passengers near the end of the route experience an average waiting time of roughly 10.3 min vs. only 5.03 min near the beginning of the route. This identifies the bus route as "problematic" and a candidate for mitigation measures.

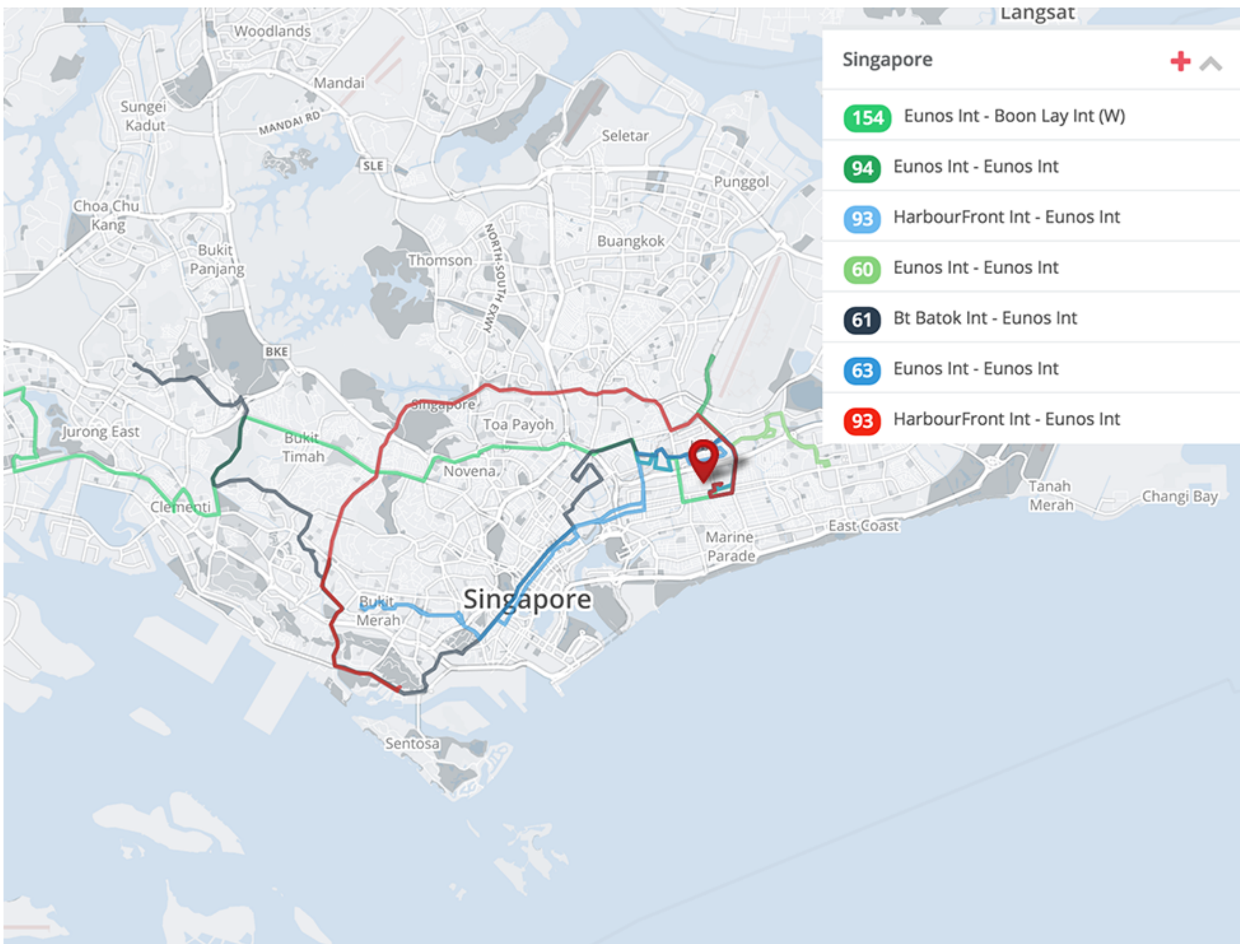


**FIGURE 7** : Map view showing all bus services originating from Eunos Interchange. It is clear that although this area is well served by bus services going to the central and western parts of Singapore, there is limited bus service to the East and North.

Beyond identifying such problematic routes, the LTA further uses the Time Space Diagram View to distinguish between passenger congestion caused by insufficient capacity, congestion caused by bus bunching and poor headway adherence, and congestion caused by a combination of both. The first case is characterized by evenly spaced buses that are consistently crowded while the second by large headway variability and buses that alternate between being completely empty and completely full. Mitigation measures may include: increasing bus frequency, increasing bus capacity (e.g. replacing some single-deck buses with double-deckers), "plugging" some headways with extra buses (e.g. Figure 5), utilizing appropriate dynamic control strategies, such as bus holding (35, 36, 37), bus speed control (38, 39), stop-skipping (40, 41), or, even modifying the bus route itself, such as avoiding heavily congested intersections that increase unpredictability or splitting long routes into two shorter, interconnected ones (29).

### Map View

A third, more macroscopic perspective of bus services is provided by BusViz's *Map View* (Figure 7), which displays all bus routes passing through selected bus stops. Quite unexpectedly, *Map View* has been one of the most heavily utilized features of BusViz at the LTA, as it has proved to be an excellent tool for quickly answering many of the numerous questions the agency receives concerning network accessibility. For example, members of the public contact the LTA to complain about how their neighborhood is poorly served by the bus network. This type of inquiry has, in the past, been very time-consuming for LTA's transit planners to handle, because although the relevant information is publicly available, it is not immediately amenable to analysis and subsequent communication to the inquiring party. Instead of combing through hard copy street directories and bus guides to piece together bus routes, LTA planners can use *Map View* to obtain an overall picture of bus service accessibility in a particular area, as well as numerous related statistics and visual documentation. This is then used to either demonstrate that a neighborhood is adequately served by public transit, or as a starting point for building a case that additional bus services are needed in that area.

## CONCLUSION AND FUTURE WORK

In this paper, we showed how large amounts of data generated from urban transportation systems combined with clever algorithms and significant advances in computing power could be used to automatically generate performance metrics relating to service quality, reliability, accessibility and passenger comfort. We introduced BusViz, a web based application that allows transit regulators to make better decisions by using ground truth data to monitor and visualize the performance of bus fleets, and showed how it could reduce reliance on, for most practical purposes, the need for transit authorities to place ground surveyors at bus stops to monitor bus arrivals and passenger loads. Lastly, we showed how BusViz could be used as a diagnostic tool to identify bus bunching, optimally insert additional buses to increase service capacity and monitor service quality (measured by average passenger waiting time) of a bus route or of any specified set of bus routes.

Singapore was an ideal research partner for us because of its advanced data and performance monitoring capabilities. Smart card systems such as those in Singapore are becoming standard in many cities worldwide and we expect that as demand for urban transportation grows, so will the need for software such as BusViz that allows cities to use ground truth data to make better decisions. To facilitate the development and adoption of this type of software, it is essential that global standards for GPS/AVP/APC data be agreed on. In future work, we plan to extend BusViz to automate the process of searching for services with capacity bottlenecks and unacceptable congestion so that these services can be immediately escalated for remedial action.

## ACKNOWLEDGEMENTS

Support for this research has been provided by the Singapore-MIT Alliance for Research and Technology (Future Urban Mobility Project), SMART Innovation Center grants ING13057-ICT, EG11011 and contract LTA000ETQ14000583. We would like to thank the Singapore Land Transport Authority for making their data available for this exciting and fruitful collaboration.

## REFERENCES


[1] Singapore-Land-Transport-Authority. Straits Times: Mandarin Makeover. `http://www.psd.gov.sg/docs/default-source/module/news/-span-the-straits-times-mandarin-makeover-span-.pdf`, 2014. [Online; accessed 15-May-2015].

[2] Laura Eboli and Gabriella Mazzulla. A new customer satisfaction index for evaluating transit service quality. *Journal of Public Transportation*, 12(3):21–37, 2009.

[3] Eftihia Nathanail. Measuring the quality of service for passengers on the Hellenic Railways. *Transportation Research Part A: Policy and Practice*, 42(1):48–66, 2008.

[4] Yong Lao and Lin Liu. Performance evaluation of bus lines with data envelopment analysis and geographic information systems. *Computers, Environment and Urban Systems*, 33(4):247–255, 2009.

[5] Madhav G Badami and Murtaza Haider. An analysis of public bus transit performance in Indian cities. *Transportation Research Part A: Policy and Practice*, 41(10):961–981, 2007.

[6] Mohammad Nurul Hassan, Yaser E Hawas, and Kamran Ahmed. A multi-dimensional framework for evaluating the transit service performance. *Transportation Research Part A: Policy and Practice*, 50:47–61, 2013.

[7] Thomas J Kimpel, James G Strathman, David Griffin, Steve Callas, and Richard L Gerhart. Automatic passenger counter evaluation: implications for national transit database reporting. *Transportation Research Record: Journal of the Transportation Research Board*, 1835(1):93–100, 2003.

[8] Sumeet Jaiswal, Jonathan M Bunker, and Luis Ferreira. Relating bus dwell time and platform crowding at a busway station. *31st Australasian Transport Research Forum*, 2008.

[9] Rabi G Mishalani, Yuxiong Ji, and Mark R McCord. Effect of onboard survey sample size on estimation of transit bus route passenger origin-destination flow matrix using automatic passenger counter data. *Transportation Research Record: Journal of the Transportation Research Board*, 2246(1):64–73, 2011.

[10] Robert L Bertini and Sutti Tantiyanugulchai. Transit buses as traffic probes: empirical evaluation using geo-location data. *Transportation Research Record: Journal of the Transportation Research Board*, 1870:35–45, 2004.

[11] Mathew Berkow, John Chee, Robert L Bertini, and Christopher Monsere. Transit performance measurement and arterial travel time estimation using archived AVL data. *ITE District*, 6, 2007.

[12] Peter Gregory Furth, et al. Using archived AVL-APC data to improve transit performan -ce and management. No. Project H-28. 2006.

[13] Robert L Bertini and Ahmed El-Geneidy. Generating transit performance measures with archived data. *Transportation Research Record: Journal of the Transportation Research Board*, 1841(1):109–119, 2003.

[14] Kimpel, T. J. Time Point-Level Analysis of Transit Service Reliability and Passenger Demand. Ph. D. dissertation, Portland State University, 2001. 482.

[15] David T Crout. Accuracy and precision of TriMet's transit tracker system. In *Transportation Research Board 86th annual meeting*, 2007.

[16] Open Trip Planner. Multimodal trip planning and analysis. `http://www.opentripplanner.org/`, 2015. [Online; accessed 15-May-2015].

[17] Transit Mix. Plan better transit: visual, easy to use, and powered by data. `http://www.transitmix.net/`, 2015. [Online; accessed 15-May-2015].

[18] Donald Low, Tan Shing Bin, Leong Chin. The evolution of public transport policies in singapore. *Lee Kuan Yew School of Public Policy Case Studies*, 2013.

[19] Yahoo News. Bus, train routes to be open to competition soon: Minister https://sg.news.yahoo.com/-1-1-billion-bus-programme-to-be-completed-by-2014-120717386.html, March 2012.

[20] Straits Times. Government steering bus industry towards overhaul. *Straits Times, April 20, 2014*, 2014.

[21] Today Online. SBS, SMRT earn one million in total for improving bus services. `http://www.todayonline.com/singapore/sbs-and-smrt-earn-over-s1m-incentives-improving-bus-services`, 2015. [Online; accessed 15-May-2015].

[22] NB Hounsell, BP Shrestha, and Alan Wong. Data management and applications in a world-leading bus fleet. *Transportation Research Part C: Emerging Technologies*, 22:76–87, 2012.

[23] Land Transport Authority. *Singapore Land Transport Masterplan 2013.* http://bit.ly/1mdTcIB.

[24] Straits Times. Singapore budget 2014: Statefunded bus fleet to double to 1,000. *Straits Times, April 20, 2014*, 2014.

[25] Soi Hoi Lam and Trinh Dinh Toan. Land transport policy and public transport in Singapore. *Transportation*, 33(2):171–188, 2006.

[26] Silvester Prakasam. Evolution of e-payments in public transport. *JOURNEYS*, page 53, 2009.

[27] Wei Zeng et al. Visualizing mobility of public transportation system. *Visualization and Computer Graphics, IEEE Transactions on 20.12* (2014): 1833-1842.

[28] Artem Chakirov and Alexander Erath. Use of public transport smart card fare payment data for travel behaviour analysis in Singapore. 2011.

[29] Der-Horng Lee, Lijun Sun, and Alex Erath. Singapore study of bus service reliability in using fare card data. In *12th Asia-Pacific Intelligent Transpotation Forum*, 2012.

[30] Ricardo Cabello. Three. js. *URL: https://github. com/mrdoob/three. js*, 2010.

[31] Mike Bostock. Data-driven documents (d3. js), a visualization framework for internet browsers running javascript, 2012.

[32] Mapbox - an open source mapping platform for developers and designers at enterprise scale. https://www.mapbox.com/.

[33] Daniel S Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, 24(4):664–675, 1977.

[34] Richard C Larson and Amedeo R Odoni. *Urban Operations Research*. Dynamic Ideas, Belmont, MA, 2007.

[35] Mark D Abkowitz and Mark Lepofsky. Implementing headway-based reliability control on transit routes. *Journal of Transportation Engineering*, 116(1):49–63, 1990.

[36] Mark D Hickman. An analytic stochastic model for the transit vehicle holding problem. *Transportation Science*, 35(3):215–237, 2001.

[37] Yiguang Xuan, Juan Argote, and Carlos F Daganzo. Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis. *Transportation Research Part B: Methodological*, 45(10):1831–1845, 2011.

[38] Antoneta X Horbury. Using non-real-time automatic vehicle location data to improve bus services. *Transportation Research Part B: Methodological*, 33(8):559–579, 1999.

[39] Carlos F Daganzo. A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. *Transportation Research Part B: Methodological*, 43(10):913–921, 2009.

[40] Aichong Sun and Mark Hickman. The real–time stop–skipping problem. *Journal of Intelligent Transportation Systems*, 9(2):91–109, 2005.

[41] Felipe Delgado, Juan Muñoz, Ricardo Giesen, and Aldo Cipriano. Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. *Transportation Research Record: Journal of the Transportation Research Board*, (2090):59–67, 2009.