A FRAMEWORK FOR ANALYZING OPERATIONS CONTROL INTERVENTIONS ON METRO TRANSIT LINES USING AUTOMATICALLY COLLECTED OPERATIONAL AND PASSENGER TRAVEL TIME DATA

André Carrel (corresponding author)
University of California at Berkeley
109 McLaughlin Hall
Berkeley, CA 94720-1720
acarrel@berkeley.edu

Rabi G. Mishalani
The Ohio State University
2070 Neil Ave., Rm 470
Columbus, OH 43210
mishalani@osu.edu

Nigel H.M. Wilson
Massachusetts Institute of Technology
77 Massachusetts Ave., Rm 1-238
Cambridge, MA 02139
nhmw@mit.edu

John P. Attanucci
Massachusetts Institute of Technology
77 Massachusetts Ave., Rm 1-274
Cambridge, MA 02139
jattan@mit.edu

ABSTRACT

Operations control, the task of implementing the timetable in daily operations on a metro line, plays a key role in service delivery, as it determines the quality of the service as provided to passengers. Despite its critical importance, it is one of the most poorly understood aspects of rapid transit operations. This paper proposes a new framework for the study of rail operations control, which builds on the integration and analysis of data from multiple sources and on background knowledge about operations control. The framework takes into account the decision environment in which operations control takes place and acknowledges that the reliability of the resulting service depends on many factors endogenous to it, aspects previously not recognized in a comprehensive manner by researchers and practitioners alike. As part of the framework, it is shown how analysts can make use of automatically generated operational and passenger data, which are increasingly available and accessible to transit agencies and allow for addressing questions in operations control from multiple perspectives. As a result, this paper takes a distinctly different approach than previous research, which has relied heavily on modeling and on simplifying assumptions about the objectives and constraints of operations control on a metro line. An important ramification of this research is to highlight the possible opportunities that lie in data-rich environments and the importance of considering possible data uses before the procurement of new data collection systems.

The developed framework consists of four main elements. First, a thorough understanding of the operations controller's decision environment is necessary. The second element is the detailed reconstruction of train operations and the identification of control interventions from automatically collected data, such as signaling data. Third, a measure for assessing the impact of the interventions on operations is introduced. The fourth and final element is a set of passenger travel time and reliability measures adapted from previous research for the assessment of operations control decisions. To demonstrate its implementation, the framework is applied in a real-world context on the London Underground Central Line. The strategy employed by controllers to manage a disruption during the AM peak on one of the most congested line segments is analyzed, and the observed control decisions are assessed in terms of their impact on operations and passengers.

Keywords: Rapid transit, rail operations control, operational data analysis, research framework, service stability and reliability, performance measurement, disruption management and recovery, automated data, passenger travel times.

1. INTRODUCTION AND MOTIVATION

On any rail transit line, the basis for the service provided to customers is the service plan, which defines routes, service frequencies, and the span of service. The service plan provides the input to the development of the operations plan, which is the set of plans that fully describe the utilization of the agency's resources – rolling stock, personnel and infrastructure – in order to fulfill the service plan. It generally consists of a working timetable (including train movements which are not in passenger service), a crew schedule, a vehicle assignment plan and a crew roster. The implementation of the operations plan into everyday operations is the task of the transit agency's operations controllers. It is often not possible to operate the service exactly as called for by the operations plan due to unforeseen events (disruptions) and infeasibilities arising from congestion on the network or the unavailability of resources. Therefore, operations controllers need to modify the train schedule in real time to account for those problems.

Operations planners and managers of a metro line need to have a good understanding of the operations on the line in order to be able to make informed decisions on service policies, to validate assumptions made during the operations planning process and to understand where changes to the operations plan can have a positive effect on the functioning of the line. In particular, given the importance of operations control in the service delivery process, any effort to analyze operations on a line should include the study of operations control. The results do not only provide benefits for short-term planning, but can also be of use to the agency for long-term investment decisions. For example, a detailed analysis of the usage of infrastructure providing operational flexibility (e.g., crossovers and reversing tracks) can show where infrastructure upgrades or the deployment of spare trains, drivers or mechanics would be most sensible and have the largest impact on service stability. However, owing to its complexity, the field of operations control is one of the most poorly understood areas of transit operations. Despite recent advances in vehicle and signaling technology, operations control remains strongly driven by agency- and even line-specific considerations and is heavily reliant on human judgment and often informal practices, which are not commonly documented.

To help analysts account for this complexity, this paper introduces a framework to support the study of operations control for the purpose of better understanding and consequently improving line operations. The framework emphasizes the importance of analyzing questions related to operations control from multiple perspectives, including a careful examination of passenger movement data, reflecting the fact that operations control interacts with virtually all elements of and actors on a line. The framework makes use of train and passenger data from automated data collection systems, reflecting the fact that in recent years the availability of such data has markedly increased, thus potentially making a wealth of information on the functioning of a transit line available to analysts.

As described by Carrel (2009), transit service can be thought of as a process involving three levels. The overall service policies such as span of service, frequency and routing are determined at the management level of the transit agency. Service policy decisions are usually based on expected or actual demand, network connectivity considerations, financial constraints and political considerations. These policies are then used by the planning department, the second level responsible for developing an operations plan. The most important component of the operations plan is the working timetable, which features all train movements reflecting where and when transportation service should be provided to customers. The service delivery activities at a transit agency come together at the third level, the operational level, where the operations plan is implemented. This implementation involves all front-line staff (e.g., train operators, station supervisors) as well as vehicle and infrastructure maintenance divisions, engineers and operational support personnel. Operations control[1], an essential component at the operational level, consists of overseeing and coordinating the implementation of the operations plan and modifying it in order to cope with unforeseen events and short-term infeasibilities. It is the centralized, real-time control of schedule-based operations, which directly results in the daily operations provided to passengers. In order to make realistic and informed decisions, the higher service policy and planning levels rely on information about daily operations and the performance of the system, which is provided by the operational level. In addition, some feedback to these higher levels is provided by passengers in the form of complaints or responses to surveys.

It is important to note that in daily operations, passengers do not experience how the service was planned to be operated; instead, they experience the actual operations as implemented by operations controllers. Operations control directly governs the interaction between the supply of transportation capacity and the demand for it on a daily basis. The activities of the operations control center (or centers) of a metro system capture the integration of the network's central communication, decision-making and coordination functions. As discussed by Froloff et al. (1989), the transit line (or system) starts at a state of optimal operations. Due to unforeseen events (also known as disruptions) or delays, it then moves to a disrupted state. If nothing were done, operations would further deteriorate. To prevent this deterioration and to restore good operations, operations controllers take corrective actions towards a target state. These actions extend over time and the system only slowly moves towards its target state via other states. During this process, further unforeseen events may occur, prompting controllers to intervene again. This process is continuous over time, taking place from the daily beginning of service until the end.

In essence, operations controllers work with three important streams of input. The first one is real-time information on the current state of the line or system, which is delivered by the operations control system (OCS). This information is usually coupled with an automated or semi-automated signaling system which allows controllers to manage the line remotely, and which provides information on the position of trains on the network, along with supplementary data, depending on the design of the system. The supplementary data may consist of, for instance, the scheduled train run which a vehicle is covering and its destination, its lateness with respect to the schedule and the driver ID. Secondly, a controller has information on the target state of the system, which is generally the operations plan. Thirdly, operations

---

[1] The term "service control" is synonymous with "operations control" and "operations management".

controllers have communications channels at their disposal, with which they can exchange information with drivers, rolling stock and crew managers and any other personnel in the field.

In case the current state of the line deviates from the target state, i.e., if there are delays and disruptions or if rolling stock or crew become unavailable, then the controllers have a set of tools, termed “interventions”, which they can apply to the line in order to move it back to the target state. There are two types of interventions: Train- and crew-related interventions. The most important train-related interventions are train cancellations, holding trains at stations, short-turning trains before or extending trains past their scheduled destinations, and, on lines with branches, diverting trains to a different branch than they were originally scheduled to serve. Crew-related interventions generally consist of changes to driver-train pairings or the substitution of spare drivers for rostered drivers. A complete overview of train- and crew-related interventions as well as detailed descriptions of the different intervention types can be found in Carrel (2009).

In the next section the previous literature is discussed followed by the introduction of the framework and its elements in section 3. Section 4 presents an illustrative example of applying the framework in a real-world context. This application and much of the investigation supporting the development of the framework relates to the study of operations control of the Central Line, one of the longest and most congested lines of the London Underground network. As a result, the Central Line is referred to on a regular basis. (A line diagram is shown at the beginning of section 4, in Figure 2.)

2. LITERATURE REVIEW

2.1. Operational analysis

To the authors’ knowledge, there are two studies on the use of automatically collected data for the analysis of operations. Wile (2003), who explores how transit agencies can use automatically collected operational data for a variety of applications and establishes a range of agency functional needs that can benefit from such data, and Dixon (2006), who presents a tool to automatically reconstruct daily operations on heavy rail lines of the Massachusetts Bay Transportation Authority (MBTA) from data produced by the OCS.

The use of operational data is embedded in a framework developed by Rahbee (2001), who examined the analysis of operational problems on rail transit lines. His research was not strictly limited to operations control, but rather took a comprehensive view towards improving service quality on rail rapid transit lines, including the study of operations control. Rahbee states three main investigation goals that an analyst can pursue:

- Document objectives and constraints as they exist in the management of a particular line.
- Investigate to what degree operations control decisions are being made according to the agency’s objectives and guidelines.

- Investigate whether the agency's objectives and guidelines regarding operations control are properly thought out.

Although Rahbee explored the use of automatically collected data to study operational questions, he did not specifically focus on operations control. Moreover, due to the limited nature of automatically collected data at the time of his study, he did not consider the full range of data currently available to transit agencies. Therefore, the framework developed by Rahbee does not touch on the challenges and opportunities related to the integration of highly detailed data from multiple sources for operational analysis.

An extensive work based on the practitioner's point of view is a manual by Froloff et al. (1989). It focuses on bus service and was originally written in an effort to identify and systematically categorize objectives, constraints and techniques for bus operations control. Despite the focus on bus services, many of the topics addressed have never been published elsewhere and they form a good foundation for studying operations control on rail systems.

2.2. Research on operations control strategies

Although research on operations control strategies and dispatching dates back to at least 1972, much of the early work assumed that the dispatcher had little or no real-time information on the position of vehicles along the line. Control strategies developed by those researchers generally examine dispatching or holding strategies at predetermined control points on the line (e.g., terminals or time points), taking as input the timetable and possibly the distance between a vehicle and its immediate neighboring vehicles. The most notable early research into this topic was by Osuna and Newell (1972), Koffman (1978), and Abkowitz (1986).

The emergence of real-time information systems providing data from all vehicles serving a transit line has enabled recent research to take a broader approach to the holding and dispatching problem and to consider other types of operations control interventions such as short-turning, expressing, and deadheading. This development creates a strong tendency to use optimization models, calibrated with real-time data, that seek to find the optimal control strategies under disrupted conditions. Some notable work in this regard was done by O'Dell (1997) and Shen (2000). O'Dell developed a real-time decision support tool for service recovery on a rail line after a disruption. She formulated a linear programming model, which included holding and short-turning as feasible strategies for responding to a disruption. Shen followed a similar path in model development, but generalized it to allow any train to be held anywhere, and added expressing and short-turning strategies. Puong (2001) built on the models of O'Dell and Shen, developing a deterministic holding model that takes into account terminal capacity and the delays incurred by passengers left behind by fully loaded trains. The objectives embedded in the models developed by O'Dell, Shen and Puong are to minimize in-train or on-platform delays to passengers, and in all cases the duration of the disruption was assumed to be known. Other research on real-time operations control was conducted by Eberlein (1995), Adamski and Turnau (1998) and Walker et al. (2005).

2.3. Appraisal of previous research

Overall, it must be said that little research has been done on operations control considering all the "tools" an operations controller has and taking into account real-time data. Many of the research efforts so far dealt with simple line layouts and a restricted set of control interventions. This limitation may not be surprising given the complexity of the field. Moreover, all research involving mathematical models of a line to find optimal control strategies has made strong assumptions about operations on a rapid transit line and has focused narrowly on passenger travel time and headway regularity, which may limit its applicability in real-world situations. The major exception is Rahbee.

So far, only Froloff et al. have attempted to describe the decision environment of operations controllers, i.e., the complete description of variables, objectives and constraints that affect operations controller decisions. However, for research to be applicable in a real-world context, it should not only recognize the multi-faceted decision environment of operations controllers but also specifically address the interactions that exist and the possible conflicts that can arise between them. This issue is treated extensively by Carrel (2009) and Carrel et al. (2010), where a detailed description of the rail service controllers' decision environment is provided and the interactions between different objectives and constraints are illustrated. To the extent that they are helpful in the context of this paper, the findings are summarized in section 3.2.

O'Dell, Shen and Puong compared the performance of their models with a do-nothing scenario to demonstrate the effectiveness of their model formulations. However, these models are only applicable on lines that have centralized operations control, and it is unrealistic to assume that operations controllers would "do nothing" (or if that were the case, they would probably have good reasons to do so). While the passenger travel time savings with regard to a do-nothing scenario may be useful for comparison across models, the real question is whether any of the models are able to provide an improvement over operations control on the line as it exists in practice. Doing so obviously requires an analysis of operational and demand data and a reconstruction of the service on a daily basis, although such a comparison would make little sense unless the various other decision factors that drive operations control are also included in the models. Last but not least, summarizing the entire passenger experience in terms of travel (or waiting) time might be too crude. More research is needed into how passengers experience and value inconveniences like additional on-platform waiting time, waiting time on a held train, or being forced to alight early due to a short-turn.

This paper aims to remedy the fact that there is currently no comprehensive framework for the analysis of operations control questions, especially one that takes into account the increasing availability, accessibility and quality of train and passenger movement data. Such a framework would improve researchers' ability to observe and analyze interventions on metro lines, which in turn could advance the state-of-the-art towards more realistic service control models and implementable and effective methods while enhancing the understanding of passenger impacts.

3. DEVELOPED FRAMEWORK

3.1. Overall structure

Before the emergence of automated vehicle and passenger data collection systems, the rail transit sector was by no means a data-rich environment, prompting operations planners and analysts to rely on models for understanding operations on a line and the effect of potential changes. However, in recent years, the availability, accessibility and quality of operational and passenger data have greatly improved and the roles of data analysis and modeling in the operations planning process are shifting. Where possible, the analysis of operational data can often provide a considerable amount of information and answers to planning questions without a direct need for models. Moreover, the limitations of previous research studies of operations control show that there is a need for an approach that integrates multiple perspectives and data sources in order to gain a better understanding of operations control (and therefore of operations) on a line.

To achieve this, an analyst needs to consider the following elements: operations control decisions (i.e., interventions), the reasons for which the decisions are made, the level of service on the affected line segments and passenger travel times. These components and their relationships are shown in Figure 1. Furthermore, the two types of factors that lead to operations control intervention decisions are also shown in Figure 1, namely disruptions and general operating and demand conditions. If data on these factors are available, they can be included in the analysis.

The interventions represent deviations from the operations plan imposed by controllers. Measuring the resulting level of service allows the analyst to quantify the impact of those operations control strategies on the service. However, this is only an intermediate step since ultimately the concern should be with the impact on passengers. Once the level of service has been quantified, the link can be made to passenger travel times, either through calculation or measurement. Once the linkage is established, the analyst can directly infer the impact of operations control strategies on passengers, as shown in Figure 1 with a dashed arrow.

Nonetheless, the picture is not complete without an understanding of the reasons for the decisions, which requires a more qualitative approach. As previously mentioned, the decision to perform an operations control intervention is informed by the controllers' knowledge of operating and demand conditions as well as disruptions to the service. It is important to note that a disruption is not a direct cause for an intervention. Instead, it is either the root cause of a deviation of the service from internal service quality metrics, or the cause of a conflict between trains. In either case, the disruption in turn leads a controller to decide whether and how to perform an operations control intervention to restore service quality or to address the conflict. In this sense, since a controller may feel that a small disruption does not warrant an intervention, the decision not to make any changes to the service in the face of a disruption is also a possible output of operations control.

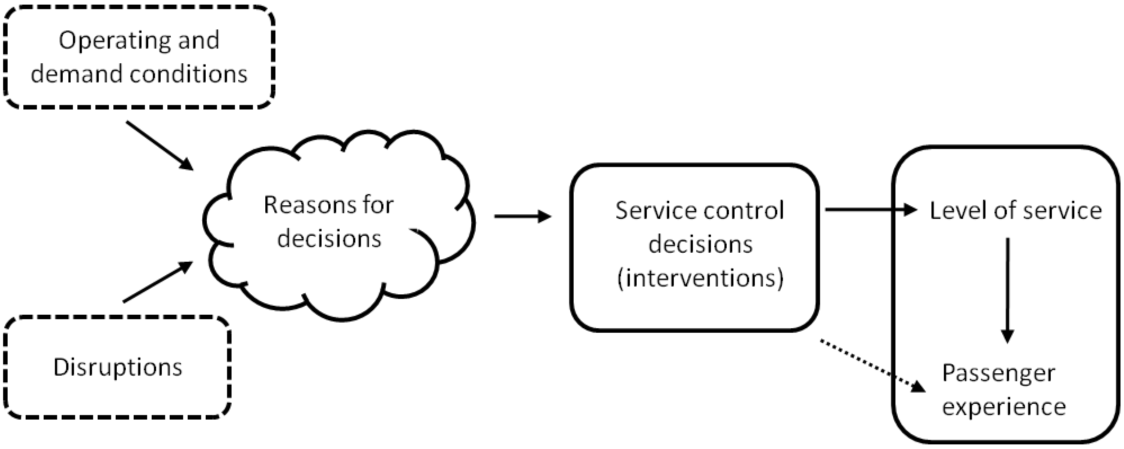

Figure 1: A framework for studying operations control on a metro line

This research framework focuses on the reasons for intervention decisions, the identification of operations control interventions and the assessment of their impact. Disruptions and the characterization of operating and demand conditions are only considered as necessary to understand the root causes for interventions. In the remainder of this section the individual elements of the framework are discussed in more detail, and in section 4 an illustrative example is presented showing how the elements are integrated to support the analysis of observed operations control strategies and their impact on service and passenger experience.

3.2. The operations controllers' decision environment

An important element of the framework is a description of the operations controllers' decision environment with all of the objectives and constraints encountered in daily operations. During extended visits to the control center of the London Underground Central Line, decision factors in operations control were observed, discussed with control center staff and compiled into a list which described the operations controllers' decision environment as completely as possible. As already mentioned, the findings are described in Carrel (2009) and Carrel et al. (2010), and therefore, only the aspects pertinent to the subsequent discussion and application presented in the next sections are pointed out.

As discussed earlier, a significant deviation from the train timetable or the crew schedule can trigger an operations control intervention, either to correct that deviation or to avoid a conflict. Adherence to the train timetable is directly related to the level of service and is a well-known driver of operations control. The major importance of crew management, on the other hand, has often not been sufficiently recognized in previous studies. It is one of the most complex aspects of operations control, as it is generally governed by numerous rules and regulations on maximum driving times, minimum layover and break times. Under disrupted conditions, these regulations pose a number of constraints on the possible recovery strategies that controllers can choose, and it was observed that the choice of recovery strategy is strongly driven by the time and location of crew reliefs. It is also noteworthy that operations controllers may lack real-time information on crews, since many OCS do not report crew lateness.

The level of service provided to passengers is obviously a key element in operations control. On low-frequency line segments with published timetables, passengers care about schedule adherence, but on high-frequency segments where passenger arrivals can be assumed to be random, headway regularity and destination sequence are of greater concern. However, to controllers the latter two criteria are not substitutes for schedule adherence, which is very important for the efficient management of assets (e.g., trains, infrastructure capacity, and staff).

It is important to note that uncertainty is omnipresent in operations control, as controllers are working with a stochastic system where it is difficult if not impossible to predict precisely how the system will evolve in the future, and decisions often have to be made under time pressure. The dominance of uncertainty is especially true in the case of disruptions, and it was observed that when choosing between recovery strategies, controllers would often favor those strategies where the least possible number of problems could arise, i.e., those where the outcome was subject to the least uncertainty.

Another decision factor encountered in the example presented later in this paper is passenger safety. For instance, trains stuck behind a blockage should be held in stations whenever possible since that greatly simplifies the evacuation of the trains in case the duration of the blockage turns out to be too long.

Generally speaking, there are complex interactions among the set of objectives, constraints and provisions by the operations plan – factors a subset of which were outlined above – that a controller has to consider when making a decision. In order to understand the rationale behind operations control interventions, a recognition and understanding of these factors and their interactions are necessary. Though the primary mission of operations control is to “repair” unreliable service, it was observed that in dealing with the other objectives and constraints encountered in daily operations, operations controllers could actually cause gaps in the service and delays, which would likely require remedies at a later stage. An example of such a situation is discussed in the example application presented in the next section.

3.3. Reconstruction of operations and diagnosis of operations control interventions

The second important element of the framework is the identification of operations control interventions, a prerequisite for which is the reconstruction of operations on the line. The challenge at this point is that the availability and quality of OCS data can vary between lines, so it is not possible to establish a unique procedure for all types of OCS data. As a matter of fact, a high level of detail is not essential for this task, although the more detailed the data and the higher its quality, the easier it is for an analyst to deliver an in-depth and precise description of what occurred on a line on any given day. The basic pieces of information required are the following:

- Actual times at which a train passed a set of predefined points along the line. In a modern, automated OCS, it may even be possible to obtain the arrival and departure times for all stations along the route.
- Numbers that clearly identify physical train units. Ideally, these would be carriage numbers if the OCS records them. If trains are assigned to vehicle blocks and retain their block number throughout the day, then these numbers can be used as identifiers.
- Identification of the scheduled trajectories train units are operating on, typically in the form of numbers that link the train units to the vehicle blocks they are assigned to.

This information allows the analyst to reconstruct the train movements throughout a day at a basic level, for example with a space-time plot. It also allows one to make the link between planned operations and the service that was actually operated by comparing the operations plan to the recorded train movements. This linkage leads to the next step, which is the identification of operations control interventions. Any change to the trajectory of a train in a metro system is necessarily the decision of an operations controller, and by searching for discrepancies between scheduled and operated service, the analyst can observe the actions of operations control. By comparing the origins and destinations of the operated and the scheduled train runs, the analyst can infer what actions were taken by operations control including short-turns, diversions, extensions, cancelations and unscheduled trips. Schedule adherence can also be inferred through this comparison and under certain conditions one can identify when a controller decided to dispatch a train early.

With more detailed information from the OCS, a nearly complete reconstruction of the trajectories of each train unit throughout the day is possible. Additional information allows for the identification of other operations control interventions, namely expressing trains, withdrawals en route, changes in platform and track assignments, the reassignment of trains to different vehicle blocks and unscheduled crew changes (e.g., the use of spare drivers). If the quality of the OCS data is sufficient, both the reconstruction of operations and the detection of operations control interventions can be automated. Carrel (2009) developed such an automated process and described the inference presented above in detail. While the automation is customized to the information available on the Central Line, it does serve as a blueprint for similar efforts for other metro lines. Such automation has several advantages over relying on manual logs completed by operations controllers. The latter are often problematic since such logs will generally be filled out only when controllers have time. During the most stressful moments in control centers, for instance during major disruptions, controllers are likely to be consumed by their operations management tasks, to the detriment of the reliability of the manual logs. Yet, the response to such disruptions would be of great interest for an analyst to study. Aside from that, a further potential problem is that controllers may record poor operations management decisions incompletely or not at all.

As can be seen in Figure 1, the identification of operations control interventions needs to be put in the context of the operating and demand conditions and disruptions which occurred at some point shortly before a decision was made. Often, there are manual disruption logs available, which can be of great help in understanding and interpreting observed operations control interventions. On the other hand, it may be more difficult to obtain data on the operating and demand conditions, but given that in most transit systems, daily and weekly demand patterns do not vary markedly except in the case of major events, the analyst can draw certain inferences about the boundary conditions from background knowledge on location and time of interest.

One of the most powerful ways to display intervention information is by indicating it on a space-time diagram, for example in the form of color-coded and numbered points, where the colors refer to the type of operations control intervention and the number refers to a list giving more information on the intervention. This depiction is not only helpful in determining how certain interventions are connected and thus form part of a larger strategy, but it also helps to visualize the immediate operational environment of a operations control intervention, for example a bunch of trains, a large gap in the opposite direction or a downstream blockage. An example of a simplified version of a space-time plot in matrix form without train trajectories is discussed and presented in section 3.4.

Aggregate counts of interventions have also proven to be a good measure for comparing the operating conditions across days, as they are a proxy for the number and severity of disruptions and they are more or less directly proportional to the number of changes that were made to train timetables, potentially affecting passengers. Last but not least, since controller workloads can vary considerably across days, total intervention counts can give a sense of their workload over a longer period of time.

3.4. Measures of impact

Following the identification of operations control interventions, measures are needed to describe the impact of these interventions. There are two main reasons why a transit agency should be interested in measuring the impact of operations control. Firstly, to determine the amount of passenger delay caused by recovery actions and operations control interventions, and to separate those from the initial delay caused by the disruption. Secondly, as Rahbee (2001) points out, operations management techniques are often passed down by word-of-mouth to new controllers. Any initiative to improve operations controller training should focus on choosing among the different “tools” for restoring the service and would need to be based on an understanding of what interventions are most effective in any given situation. The specification of impact measures allows for formalizing such protocols.

An important issue is that the impact of operations control interventions is typically limited to specific segments of the line and to small time windows. In choosing between impact measures, the analyst should be aware of the resolution trade-off in effect. Highly disaggregate data allow for a high-quality analysis of intervention strategies but make it challenging to analyze operations control over longer periods of time due to the associated high degree of variability from situation to situation. Aggregate data, on the other hand, allow for a long-term assessment but may hide the effects of operations control interventions. Finding the right balance between these two extremes is critical.

Level of service by line segment

All interventions that cause a train to be rerouted or canceled can be characterized in terms of their impact on individual segments of the line. For example, a canceled or short-turned train trip means that part or all of the line segments it was scheduled to serve were served by one train less. While the addition or removal of train service at individual stations would be the most precise way of examining the impacts, the data become more manageable without much loss of information if they are aggregated at a segment level. The segments should be defined such that the scheduled level of service is approximately constant throughout the segment and such that within the segment there are preferably few, if any, reversing tracks that are known to be used by operations controllers.

Table 1 shows a sample of how operations control interventions and the effects on different segments of the line can be displayed in the form of a matrix. This example relates to the London Underground Central line, for which a more in-depth example is presented in section 4. For convenience of presentation, only the three westernmost segments are shown. The top part of the table shows the types of interventions by hour of day, and in the lower part, which is divided by segment, the net effects of the operations control interventions are shown either as added service on that segment (e.g., "1" indicates that aside from the scheduled service, one additional train traveled through that segment) or a reduction in service (e.g., a "-1" indicates a train reduction with respect to the scheduled service).

Central Line Intervention Matrix - Wednesday, November 12, 2008																					
Westbound Time:	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***21***	***22***	***23***	***0***	***SUM***
Short-turns					8	3											1				***12***
Diversions					2																***2***
Extensions				1	1							1			1	1	1				***6***
Canceled trips			1	2							1	2	1	2							***9***
Unscheduled trips					1																***1***
Withdrawn trains				2														1			***3***
Total westbound																					***33***
Segment 1 (branch): Ealing Broadway to North Acton																					
Reduced service					-3	-2							-1				-1	-1			
Additional service					2												1				
Segment 2 (branch): West Ruislip to North Acton																					
Reduced service			-1	-2	-5	-1						-2		-2							
Additional service				2	1							1			1	1					
Segment 3 (trunk): North Acton to Liverpool Street																					
Reduced service			-1	-3	-6	-3					-1	-2	-1	-2			-1	-1			
Additional service				1	2												1				

Table 1: Intervention matrix with effects on line segments

The measures displayed in the matrix should not be confused with the number of trains per hour (tph) serving a line segment. The reason is that tph (or average headway) are good measures across a specific point on the line (typically stations), but for a line segment, trains may only serve portions of that segment within a given time window, and the measures would need to be averaged over multiple stations where they are not necessarily equal. The ratio of scheduled to operated service kilometers at an hourly and segment level (i.e., measured in units of train-km/h for each segment) is proposed instead. This measure allows for the over- and under-provision of service to be pinpointed to individual segments of the line and time periods, thus capturing the net effect of an intervention strategy. For instance, if one train is short-turned but the one behind it is extended to cover its trip out to the branch terminal, this will show up as better service than if the first train had been short-turned without extending the following one. This distinction is important since only looking at the intervention counts may convey the impression that larger absolute numbers of interventions always result in larger numbers of service curtailments. As the previous example shows, this result is not necessarily the case.

Passenger impact

The increasing use of smart card technology in the past decade is beginning to provide transit agencies with easily usable information about passenger flows on their network. In cases where passengers are required to tap their cards upon their entry to and exit from the system, or when they transfer, smart card data allow the direct measurement of passenger time spent in the system. From the measurement of this total travel time, it is also possible to draw inferences about the impact of disruptions and service changes on passengers. In the following discussion, it is assumed that passenger data include the entry time and location (station), exit time and location (station), and that a unique card number allows the entry and exit to be linked. Furthermore, it is assumed that transfers are not recorded.

Not all operations control interventions have the same impact on passengers, and different passenger groups will experience the effects of an intervention differently. One challenging task in assessing passenger impact is the projection of the (dis-)benefits into the future, due to the following reasons:

- In daily operations, dispatching decisions are made continuously. Some interventions are performed in order to mitigate the negative impacts of earlier interventions, so the impact of a single intervention can often not be assessed without considering the overall strategy.
- Due to the stochastic nature of the environment in which the service operates, the inherent variability of dwell times and running times and unforeseen events (e.g., additional disruptions) could quickly obfuscate the effects of an individual intervention.

In brief, when quantifying passenger impacts, it is advisable to limit the analysis to passenger delays (or travel time improvements) that can be clearly attributed to an intervention. The impacts of operations control interventions on passengers include:

- Increased or decreased waiting time at the passenger trip origin
- An additional train transfer and waiting time during the journey.
- Increased in-vehicle travel time, e.g. due to holding of trains in stations.

In this study, the focus is on disruptions and operations control interventions with a negative impact (i.e., delays to passengers). While it is important to bear in mind the different types of passenger impacts, identifying them individually with smart card data is difficult if not impossible. Therefore, this analysis is restricted to the study of a sample of passenger total travel times (from the time of entry at the initial origin station to the time of exit at the final destination station) aggregated over a certain time period and pertinent OD pairs. The resulting travel time distribution captures all passenger trips that started within that time period, distributed over all trains that served the respective OD pair. A travel time distribution is usually skewed to the left since there is a minimum possible travel time. The distribution's right "tail" includes riders who either experienced the worst service during that time period or whose travel times include activities other than traveling. For a more in-depth discussion of the travel time distribution, the reader is referred to Uniman (2009) and Uniman et al. (2010).

Two measures are used to summarize the travel time distribution and thus the nature of the impact of interventions on passenger travel times. The average travel time, either for every origin-destination (OD) pair or averaged over several OD pairs, and a measure of the spread in the travel times for each OD pair, which is introduced subsequently in this section.

Generally, it is not straightforward to distinguish whether the delays reflected in a travel time distribution are caused by disruptions or by operations control interventions; gaps in the service can be the consequence of either of those two events. As a matter of fact, one could argue that passengers may not even care about the origin of the gap, but to the transit agency, the origin is of great interest. Addressing the problem of determining the origin of a delay is highly specific to the context of the situation under examination. Therefore, addressing this problem in a practically meaningful fashion is reserved to the application discussed in detail in section 4.

In the specific case of a high-frequency metro line, the temporal aggregation of passenger journeys captures passengers of multiple trains, but oftentimes, only a few trains are affected by a delay. For instance, of 12 scheduled trains per hour one or two may be short-turned. For the purpose of illustration only, if one assumes even demand, only $\frac{1}{12}$ to $\frac{1}{6}$ of the passengers traveling to destinations beyond the short-turning point would be affected, either because they had to alight and wait for a later train or because they experienced a gap of one or two additional scheduled headways. Although the average travel time during this time period is an important measure, it is clear that the average is insufficient for capturing the effects of small disruptions and operations control interventions, as it has a limited sensitivity to the (often small) subset of passengers who experience worse service. However, one should expect to see a change in the shape of the travel time distribution, as said subset of passengers experiencing worse service is pushed to contribute to the right side of the travel time distribution and as a result the distribution becomes wider. The benefits of considering the width of the travel time distribution in addition to average travel time are even more important in real-world applications, where the analyst does not necessarily know how demand was distributed among the set of trains within the time window that is being analyzed.

The above effect relates directly to the issue of *service reliability*, which is of great concern to transit agencies. More specifically, before making the trip, a single passenger will not know whether his/her train is the one which is going to be short-turned or otherwise delayed; i.e., his/her travel time might fall anywhere within the travel time distribution. The wider the distribution, the more unreliable the service is from the perspective of the passenger. This acknowledgement calls for an additional measure (or measures) capturing the width of the distribution and, thus, the unreliability in travel time. The *Reliability Buffer Time (RBT)* is such a measure. Various authors have discussed this type of measure, among them Furth (2006), Chan (2007), Uniman (2009), who introduced the term Reliability Buffer Time, and Uniman et al. (2010). Uniman defines it as:

$$\textit{Reliability Buffer Time} = (N^{th} \textit{ percentile travel time} - M^{th} \textit{ percentile travel time}) \qquad (1)$$

Uniman proposes M = 50th percentile, which is the median and is an indicator of the typical travel time a frequent traveler on the system would anticipate (the mean is not proposed to capture the typical travel time because it is more sensitive to outliers than the median). Furthermore, Uniman states: "*The variable N is an indicator for the threshold of certainty for reliable service".* In other words, if a passenger wishes to reach his or her destination by a certain time with 90% probability, he or she must budget the travel time given by the 90th percentile of the cumulative travel time distribution. Thus, the RBT is the additional time to be budgeted over and above the typical travel time by passengers who wish to arrive at their destination by a certain time with some degree of confidence. For those passengers who budget this additional time but arrive before their target arrival time, the resulting extra time at their destination could be spent unproductively.

While the travel time distribution is a powerful way of representing how passengers experience service on a line, it has some limitations. Any action of a passenger which is not directly related to the originally intended trip – for example, taking the wrong train – enters into the travel time distribution despite its irrelevance to the quality of service. Those passengers are likely to show up in the higher percentiles of the distribution, giving a false impression of service unreliability. Unfortunately, there is currently no way of filtering out such passengers, so the analyst depends on the assumption that either any inexplicable passenger behavior results in travel times beyond the 90th percentile or that this subset of passengers is sufficiently small and does not severely affect the summary statistics.

Both the average travel time and the RBT can be expressed as totals (i.e., total travel time and total RBT), weighted by passenger origin-destination demand, which is helpful in putting the numbers into context. The calculation of total travel time can be performed at a line segment and time period level. The total RBT is most meaningful when compared to the total travel time, as it represents the total time which all passengers would need to budget in addition to their expected travel time in order to arrive at their destination by their desired arrival time with 90% certainty.

4. ILLUSTRATIVE EXAMPLE

The presented illustrative example is intended to show how the various elements of the framework are integrated and how it can be applied in a real-world setting. As already mentioned, the observations are made on the London Underground Central Line. To facilitate the understanding of this example, Figure *2* shows a line diagram, along with the stations referred to in this example (in total the Central line serves 49 stations).

This example shows how a retrospective analysis of the way a disruption was managed by operations controllers can help identify problematic operations management techniques. The insights from such an analysis can, for example, be used to establish agency guidelines outlining priorities to operations controllers and to consequently improve operations controller training.

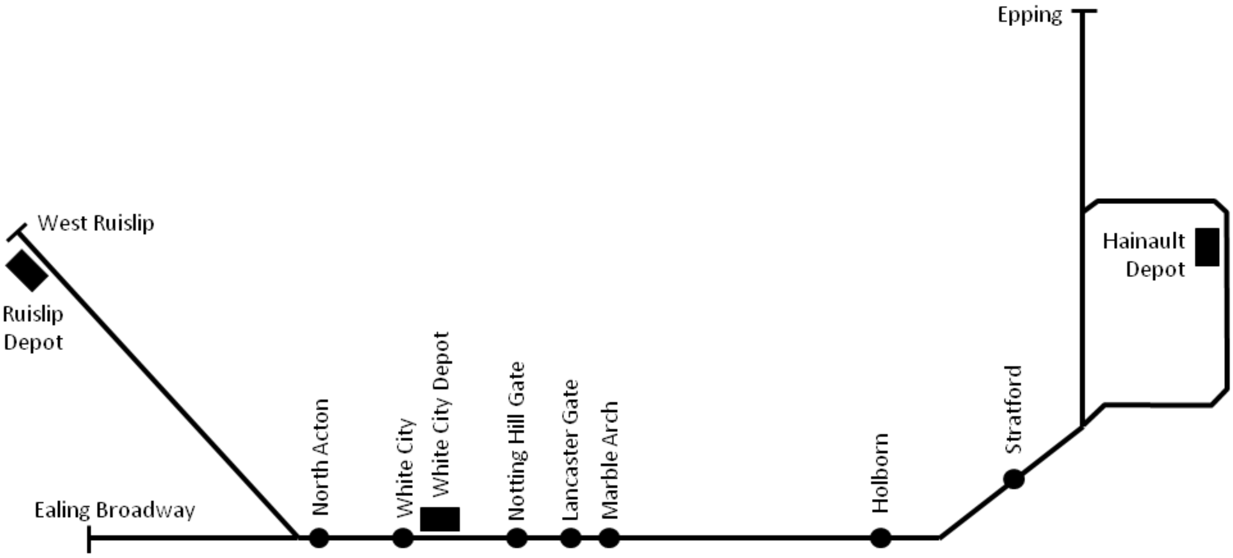

Figure 2: Central line diagram

The focus of the analysis is a disruption that occurred on the Central Line on April 1, 2008. The disruption can be summarized as follows. At 08:29, eastbound train #141 experienced an in-tunnel delay between Lancaster Gate and Marble Arch station due to a track circuit failure[2]. The disruption was cleared at 08:35, and the affected train departed Marble Arch station at 08:36. This is the only information given by the Underground's incident log, and to understand what happened in the wake of the incident and how it affected passengers, an analyst needs to draw together information from multiple sources, as outlined in the framework presented in this paper. First, one may want to consider the demand and operating conditions during this time. Demand on the London Underground has a large peak-to-base ratio, with demand peaks between 08:00 and 09:00 as well as between 16:00 and

[2] A track circuit is used to communicate operational commands of the automated signaling system to the on-board computer of the train unit. In the event of a track circuit failure, trains cannot operate normally over the affected segment of the line.

17:00. Furthermore, April 1 being a regular weekday, demand is highest on the trunk segment of the line, roughly between Notting Hill Gate and Stratford. Hence, this incident occurred at one of the most congested times and locations on the line.

The next step was to identify operations control interventions that occurred on the Central line eastbound at or shortly after 08:29, the start time of the disruption. The result of the analysis based on the methodology described in section 3.3 was that operations controllers held all trains upstream of the disruption between Lancaster Gate and White City in stations. White City is a depot and an important crew relief location; to the east of White City, the Central line runs below ground through central London, while to the west of White City, the line is above ground. On April 1, no holding could be detected to the west of White City, as trains continued passing those stations with average-length dwell times. Furthermore, two cancellations of consecutive eastbound trains out of White City occurred. One of those trains (#114) was scheduled to reverse at that station, while the other one (#105) had arrived from West Ruislip. The canceled trains had arrived at White City at 08:33 and 08:36.

The knowledge of operating and demand conditions, the nature of the disruption and the actions taken now allows the focus of the analysis to be moved to the reasons for those interventions. First, the holding of trains in stations in the tunnel segments is known to be performed for safety reasons (to facilitate train evacuation if found to be desirable or necessary). Second, based on the explanations given by controllers, as described in section 3.2, it can be assumed that the strategy of canceling trains during disruptions is common. There are two specific reasons for which these cancelations are likely to have been performed: to avoid train congestion build-up upstream of White City and to remove trains from the pool of late trains and drivers, thus improving the manageability of crew lateness. These hypotheses could not be verified directly with controllers, but conversations with TfL engineers confirmed that they are highly likely.

The two remaining elements of the framework to be explored are the impacts of these decisions and resulting interventions on the level of service and on passenger travel times. According to the operational data, the result of the two train cancellations was that there was a gap of 7:20 min in the service after the disruption had cleared. Applying the operated train-kilometers to scheduled train-kilometers ratio measure proposed in section 3.4, only 81% of scheduled service (measured in train-kilometers) was operated between 08:00 and 09:00 on the eastbound trunk segment of the line.

Since the Central line OCS continuously records train loading as a percentage of crush load, ranging from 0% to 100%, it could be confirmed from the data that train capacity did become a problem due to peak hour demand. The removal of two trains from service during the peak hour reduced throughput capacity by approximately 2100 spaces in that short period of time, leaving very little additional capacity for downstream passengers. Presumably many passengers were not able to board the crush loaded trains, causing them to be left behind and experience longer waiting times than could have been attributed to the gap alone.

To verify this finding and assess the passenger impact of the disruption and intervention, individual passenger travel times from smart card data are analyzed following the methodology described in section 3.4. A high variability in travel times is observed for different passengers who had entered the system shortly before the time when the crush loaded trains passed their station. This presumably reflects the fact that only a fraction of passengers at stations where the trains were at or near crush load were able to board, while others waited for the arrival of less loaded trains. The specifics of the corresponding analysis are outlined in what follows.

Only eastbound passengers on the trunk segment are considered as they were the ones unambiguously identified as affected by the disruption and the following gap. Two time bands are defined for the analysis: time band 1 includes the 30 minutes before the end of the disruption, including the disruption itself, while time band 2 captures the 30 minutes immediately after the disruption had cleared. Both time bands are defined to relate to passengers depending on the stations they are boarding at. For instance, passengers accessing the service at a station downstream of the disruption were attributed to the first group if they tapped in before the passage of train #141 at that station and to the second group if they entered the station afterwards.

To get a sense of the magnitude of passenger travel time changes due to the disruption and intervention, passenger travel times were also sampled during the same time periods on November 21, which had very good, undisrupted service and which served as a "baseline day". By comparing travel times on the disrupted and the undisrupted day, conclusions could be drawn about the passenger impact of the disruption and, more importantly, of the recovery strategy.

The results for both time bands and both days are shown in Table 2. The comparisons clearly show the impact of the disruption management strategy in comparison to the delays caused by the disruption itself. Firstly, the travel times per passenger calculated based on the total travel time hardly differ between the two time bands, and both are higher than the corresponding figures for the baseline day, November 21. As a matter of fact, passengers after the disruption effectively experienced the same service degradation with respect to November 21 as during the disruption, which is evident from the travel time differences per passenger.

	Time band 1 (before and including time of disruption on April 1)		**Time band 2 (after time of disruption on April 1)**	
	1-Apr	21-Nov	1-Apr	21-Nov
Average travel time per passenger [min]	19.04	16.55	19.69	17.21
Average RBT per passenger [min]	7.35	3.96	9.22	3.65

Table 2: Comparison of passenger impact of the disruption on April 1, 2008 with November 21, 2008 (undisrupted baseline day)

Secondly, in terms of the unreliability of the service as measured by the RBT, time band 1 does not show a large increase compared to November 21, amounting to slightly more than 3 minutes per passenger. In other words, passengers during the disruption experienced longer overall travel times than under normal peak-hour conditions, and the service was somewhat less reliable than usual. However, the situation is more severe after the disruption, where the average additional RBT amounted to slightly more than 5.5 minutes per passenger or 32% of the average travel time under undisrupted conditions.

Therefore, not only did the interventions cause additional waiting times of the same magnitude per passenger as the delays caused by the disruption itself, but the service became considerably less reliable during the recovery than during the disruption. This difference in impact on reliability is presumably due to the two gaps left by the canceled trains as well as to passengers left behind by the crush-loaded trains mentioned earlier. This result is a clear example of how intervention actions taken by operations controllers, however well justified from their perspective, can lead to increased unreliability directly experienced by passengers.

Aside from the delays to eastbound passengers on the trunk segment that are quantified in this analysis and reported in Table 2, there were additional passenger delays not captured by this analysis that occurred later in the westbound direction since the two canceled trains were missing for an entire round-trip and, therefore, the quantified delays are lower bounds. While the fact that westbound trips were omitted for this analysis is certainly a limitation, further research would be needed to be able to define and capture the affected passenger groups in the westbound direction. The westbound level of service cannot be linked to the disruption as clearly as the eastbound level of service, mainly for two reasons: Controllers have the ability to alter train sequences through interventions such as short-turns, and there is an important time lag between the disruption and the reversal of the trains during which other unforeseen events can occur.

A more complete discussion of the analysis, these results and questions that cannot be answered with the passenger travel time data alone is presented by Carrel (2009). In summary, analyzing this disruption by applying the proposed framework and the corresponding methods associated with each of its elements leads to meaningful insights on operations control and on operations. In this particular case and as already discussed, two possible reasons that are suggested for the controllers cancelling the two trains are to avoid train congestion build-up and to manage crew lateness. In the case of the former, an alternate strategy of holding additional trains upstream of White City would have resulted in lower passenger delays after the disruption. In the case of the latter, it is questionable whether controllers were following the optimal priorities, as the tradeoff between passenger delays and driver delays appears to be strongly out of balance.

As a next step, it would be helpful to discuss the findings with controllers directly in order to verify that there were no other operational issues that the controllers might be aware of and that were not revealed in the data used in the analysis. Furthermore, it would be worthwhile to discuss with controllers possible strategies to address their concerns about congestion and crew lateness while causing less passenger delays.

5. CONCLUSION

Research into operations control has so far relied heavily on modeling and has mostly been focused on individual operations control strategies. Owing to the complexity of the field, researchers have often made simplifying assumptions on the functioning of metro lines and on the objectives and constraints of operations control. Within many transit agencies, on the other hand, there is a need to improve the feedback of information regarding everyday operating conditions on the network to inform the development of operations plans that take such conditions into account.

The contributions of this paper are fourfold. First, it is shown that automatically collected data, which are increasingly available and accessible to transit agencies, can be of significant value for improving the information feedback loop in a transit agency, allowing managers and planners alike to make more informed decisions and to validate assumptions and models used. The only cost associated with the use of these data consists of the effort needed to extract, process, integrate and analyze them. Second, it is argued and demonstrated that any effort to analyze and build a good understanding of operations on a line must include operations control. Third, a framework is developed to integrate automatically collected data from multiple sources and, based on these integrated data, analyze operational questions from various perspectives in an operations control centric manner. Fourth, the developed framework is applied demonstrating how the various elements are integrated to arrive at clear conclusions and insights regarding the actions and impacts of operations control.

6. ACKNOWLEDGEMENTS

This research was made possible by the generous support of Transport for London. The data and expertise provided by TfL management and technical staff are greatly appreciated. The authors would also like to thank the two anonymous reviewers of this paper for their helpful comments.

7. REFERENCES

Abkowitz, M. (1986). *Optimal control of headway variation on transit routes*. Journal of Advanced Transportation, 20, pp. 73-88.

Adamski, A., and Turnau, A. (1998). *Simulation support tool for real-time dispatching control in public transport.* Transportation Research, 32A, pp. 73-87.

Carrel A., Mishalani, R., Wilson, N., Attanucci, J., and Rahbee, A. (2010). *Decision Factors in Service Control on a High-frequency Metro Line and their Importance in Service Delivery*. Transportation Research Record (forthcoming).

Carrel, A. (2009). *Diagnosis and Assessment of Operations Control Interventions: Framework and Applications to a High Frequency Metro Line*. MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Chan, J. (2007). *Rail OD Estimation and Journey Time Reliability Metrics Using Automated Fare Data.* MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Dixon, M. (2006). *Analysis of a Subway Operations Control Database: The MBTA Operations Control System.* MS Thesis. Northeastern University, Boston MA.

Eberlein, X. (1995). *Real-Time Control Strategies in Transit Operations: Models and Analysis.* PhD Thesis. Massachusetts Institute of Technology, Cambridge MA.

Froloff, E., Rizzi, M., and Saporito, A. (1989). *Bases et pratiques de la régulation*. RATP, Direction du Réseau Routier RC/MSE.

Furth, P., Muller, T. (2006). *Service Reliability and Hidden Waiting Time: Insights from Automatic Vehicle Location Data.* Transportation Research Record, No. 1955, pp. 79-87.

Koffman, D. (1986). *A simulation study of alternative real-time bus headway control strategies.* Transportation Research Record, No. 663, pp. 41-46.

O'Dell, S. (1997). *Optimal Control Strategies for a Rail Transit Line.* MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Osuna, E., and Newell, G. (1972). *Control strategies for an idealized public transportation system.* Transportation Science, 6, pp. 57-72.

Puong, A. (2001). *A Real-Time Train Holding Model for Rail Transit Systems.* MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Rahbee, A. *Rail Transit Operations Analysis: Framework and Applications.* MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Shen, S. (2000). *Integrated Real - Time Disruption Recovery Strategies: A Model for Rail Transit Systems.* MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Uniman, D. (2009). *Service Reliability Measurement using AFC Smart Card Data – A Framework for the London Underground*. MS Thesis. Massachusetts Institute of Technology, Cambridge MA.

Uniman, D., Attanucci, J., Mishalani, R., and Wilson, N. (2010) *Service Reliability Measurement Using Automated Fare Card Data: Application to London Underground.* Transportation Research Record (forthcoming).

Walker, C., Snowdon J. and Ryan D. (2005). *Simultaneous disruption recovery of a train timetable and crew roster in real time.* Journal of Computers and Operations Research, Vol. 32, pp. 2207-2094.

Wile, E. (2003) *Use of Automatically Collected Data to Improve Transit Line Performance.* MS Thesis. Massachusetts Institute of Technology, Cambridge, MA.