Decision Factors In Service Control on a High-frequency Metro Line and Their Importance In Service Delivery

Revised paper submitted for presentation at the 89th Transportation Research Board Annual Meeting (2010) and possible publication in Transportation Research Record

Nov. 14, 2009

André Carrel (corresponding author)
University of California at Berkeley
109 McLaughlin Hall
Berkeley, CA 94720-1720
acarrel@berkeley.edu
(857) 919-6351

Rabi G. Mishalani
The Ohio State University
2070 Neil Ave., Rm 470
Columbus, OH 43210
mishalani@osu.edu
(614) 292-5949

Nigel H.M. Wilson
Massachusetts Institute of Technology
77 Massachusetts Ave., Rm 1-238
Cambridge, MA 02139
nhmw@mit.edu
(617) 253-5046

John P. Attanucci
Massachusetts Institute of Technology
77 Massachusetts Ave., Rm 1-274
Cambridge, MA 02139
jattan@mit.edu
(617) 253-7022

Adam B. Rahbee
London Underground Ltd.
55 Broadway Rm 290
London SW1H 0BD
arahbee@alum.mit.edu

Word Count: 8456 (text) + 750 (figures) = 9206
Decision Factors In Service Control on a High-frequency Metro Line and Their Importance In Service Delivery

André Carrel, Rabi G. Mishalani, Nigel H.M. Wilson, John Attanucci, and Adam Rahbee

ABSTRACT
Service 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. Following a discussion on the role and importance of service control in service delivery, previous research is reviewed and their shortfalls are noted. A research framework intended to remedy some of these shortfalls is then proposed. An important element of this framework, and the greater focus of this paper, is the description of the full decision environment in which service control takes place. Based on insights gained from extended visits to a control center, the reliability of the system is found to depend on many endogenous factors that were not previously recognized in a comprehensive manner by either researchers or practitioners. Aside from the objective of maintaining adequate levels of service from an operations perspective and minimizing the impact of schedule deviations on passengers, considerations relating to management of crew and rolling stock, safety, and infrastructure capacity have a major influence on service control decisions. Also, given the uncertain environment in which service control operates, a strong preference was observed among controllers for manageable and robust control strategies. An illustrative example is discussed where service controllers react to two similar disruptions with different recovery strategies mainly due to crew management considerations. This result further demonstrates the importance of developing a comprehensive understanding of the objectives and constraints faced by service controllers in daily operations. Finally, specific recommendations regarding future research are presented in light of the findings of this study.
Decision Factors In Service Control on a High-frequency Metro Line
and Their Importance In Service Delivery

André Carrel, Rabi G. Mishalani, Nigel H.M. Wilson, John Attanucci, and Adam Rahbee

INTRODUCTION

This paper highlights the importance of studying service control to better understand the interactions between the operations plan and service control, and ultimately to improve both the service control and the operations planning process. Specifically, it describes the decision environment faced by rail service controllers and shows why previous research approaches, which have been strongly focused on mathematical modeling and have involved significant simplifications to keep the models tractable, have had very little application. The description of the service controller’s decision environment is intended to point out these gaps and suggest what might be done to fill them. In addition, an illustrative example based on an actual service control situation is presented to demonstrate the importance of some of these gaps and further motivate the need to address them from a practical perspective. Based on the findings, specific recommendations regarding future research are presented.

The basis of operations on a transit line, which needs to be understood well before studying service control, is the operations plan. The operations plan is, generally speaking, the set of plans that fully describe the utilization of transit agency resources to bridge the service plan with daily operations. It typically consists of a working timetable (including movements of trains that are not in passenger service), a crew schedule, a vehicle assignment plan, and a crew roster. As described by Moore (1), it represents the ideal operational procedures, crafted in advance. It is designed to deliver the service policy requirements resulting from earlier phases of the service planning process, while complying with crew work rules, vehicle management, infrastructure capacity, and maintenance requirements. The operations plan serves as an input into service control, the function of which is to implement that plan under everyday conditions, and to modify it in real-time to deal with disruptions (which are discussed in detail subsequently), significant variation from expected demand, and inherent variability in dwell times and running times experienced by vehicles. These sources of service unreliability affect passengers negatively and are of major concern to most transit agencies. However, although unreliability can be observed in the variability of passenger travel times and in passenger complaints, its root causes lie predominantly on the supply side and cannot be effectively addressed without building a good understanding of the operational processes on the transit line of interest. This understanding can be achieved only by studying daily operations on the system, for which service control is a very important component. This holds especially true on a high-frequency, high-demand metro line where non-ideal (“disrupted”) conditions occur frequently.

Deviations from the operations plan can be due to either congestion during the peak hours, which causes longer running times and dwell times than were scheduled, or service disruptions. A service disruption is defined as a single, unforeseen event that prevents one or more trains from completing their trips as scheduled. The cause of a service disruption can be controllable or uncontrollable (at least in the short run) by the transit system. Uncontrollable reasons are, for example, a passenger activating the emergency alarm, an object on the tracks, or weather (on open track sections). Controllable disruptions, on the other hand, can be influenced by the transit agency's maintenance procedures, accountability structures, and employee discipline policies. Examples of controllable causes are defective trains, infrastructure problems (such as signal failures), staff communication errors, or the unavailability of a driver or train.
A disruption can result in a gap in service (possibly followed by bunched trains), an incorrect sequence of trains, general train lateness, or some combination of these effects. It is important to note that the effect of a disruption on the service is likely to be viewed differently by the agencies and by the passengers, especially when considering high-frequency metro lines where passengers generally arrive randomly, without referring to a timetable (2). In this case, service quality variables that are important to the passenger include platform waiting time and on-train travel time, but not necessarily schedule adherence. However, from the agency's perspective, the degree of adherence to the operations plan (and thus to the crew and vehicle schedules) is of critical importance. Disruptions can cause both a deviation from the schedule (i.e., train or driver lateness) and a deviation from service quality standards. The two effects, although correlated, are not necessarily causally linked. For example, it may be possible to maintain the scheduled headways on a line section despite trains running late due to an earlier disruption. That would be a deviation from the schedule, but waiting passengers would experience service at the expected headways, and there would be no deviation in service quality with regard to waiting times. Therefore, it is important to consider both perspectives and to understand the interactions between them in studying service control, as will be further discussed in this paper.

SERVICE DELIVERY PROCESS AND MOTIVATION OF STUDY

In managing the effects of a disruption or congestion-induced delay and in recovering from them, controllers work with real-time information on the state of the system and resort to a menu of changes that can be made (i.e., service control interventions) to achieve an ultimate target state. Those changes include train related interventions such as holding a train at a station or dispatching it early, expressing it (skipping scheduled stops), canceling an entire train trip or adding an unscheduled trip, short-turning a train, extending it beyond its scheduled destination, withdrawing it from passenger service prematurely, or, if the line has branches, diverting it to a different branch than it was originally scheduled to serve. A further set of important service control decisions is related to train priorities at junctions and terminals, which can influence service quality throughout the line. While the train-related interventions affect trains and crews alike, there are also a number of crew-related interventions that only affect crew usage, such as substituting a spare driver or changing the location or time of a crew relief.

In order to achieve a certain outcome (e.g., put a late train back on schedule), there are often many different possible interventions from which a controller can choose. The controller’s choice of intervention is driven by a set of objectives and priorities, which may be defined by agency policy or, more commonly, by the controllers. Despite recent advances in vehicle and signaling technology and improved control systems, the field of service control remains heavily reliant on human judgment and on informal, undocumented practices. Owing to the complexity of the field, the managers and planning staff in many transit agencies are often not well informed of the problems faced by service controllers as they try to implement the operations plan on a daily basis, or of the decisions that result. This situation has several implications. On the one hand, management decisions and agency policies aimed at improving service control may be, despite the best of intentions, unrealistic or inapplicable in the real-world context. On the other hand, in the absence of an understanding of the role of service control, planners may have difficulty interpreting performance metrics correctly and verifying whether assumptions and models used in scheduling were in fact correct.

One can conceptualize a transit service as a business process, as shown in FIGURE 1. The overall service policy and plan, such as span of service, frequency, and routing are determined at the management levels of the transit agency. These decisions are usually based on expected or actual demand, network connectivity, financial constraints, and political considerations. The service policy and plan are then used as input by the planning department, which is responsible for developing an operations plan. The last
piece within the transit agency comes together in the form of daily operations, which involve all front-line staff (e.g., train operators and doormen) as well as vehicle and infrastructure maintenance divisions, engineers, and operational support personnel. Service control, which is an essential component at this operational level, oversees and coordinates the implementation of the operations plan and modifies it in order to cope with unforeseen events and the resulting short-term infeasibilities. The service delivery process results in the daily operations provided to passengers. It is important to recognize that passenger experience is heavily influenced by service control. That is, passengers do not always experience the service as it was planned; they experience the actual operations.

In FIGURE 1, decisions flow from left to right with the high-level, strategic policy and service planning decisions taken at the management level. Then, the operations planning decisions made at the planning level determine how to implement the policy and service plan in practice. At the operational level, falling at the end of the decision chain, the resulting operations plans should ideally be carried out directly, but in reality they often need to be fine-tuned. In order to make informed and realistic decisions, the higher levels rely on information about daily operations and the performance of the system, which is provided by the operational level to the planning level and then on to the management level. This feedback often represents the only direct communications link between these three levels. In addition to cross-divisional communication, some feedback is also provided by passengers in the form of complaints or responses to surveys.

![FIGURE 1 Transit service delivery as a business process.](image)

It can safely be said that service control is one of the most poorly understood aspects of rail transit. It is often a proverbial “black box” to managers and planners alike, despite its crucial role in service delivery. Therefore, the flow of information from the operational to the planning and management levels of a transit agency, as shown in FIGURE 1, is often not as strong as it should ideally be. In addition, as Rahbee (3) points out, service control techniques and objectives are typically passed down by word-of-mouth, and they vary across lines as well as agencies. Given the level of influence that service control has over the functioning of the line, it would be desirable for transit agencies to move towards more unified service control policies. The study of service control can not only help an agency with these efforts, but also improve understanding of daily operations and the types of deviations from the service plan that an analyst can observe in performance metrics at an aggregate level. More specifically, only the study of
service control will help build an understanding of the nature of, and the reasons for, these deviations, and any research effort in this area is greatly facilitated if the analyst can build on prior knowledge about the objectives and constraints of service control.

In the following section, the literature relating to service control is reviewed to provide further context for a comprehensive examination of service control based on field observations, which sets the stage for the overall analysis framework adopted in this study. This presentation leads to the description of service control observations on a specific metro line, forming the basis for a comprehensive view of the service control environment, the possible objectives, and the various constraints faced by controllers. Based on this description, the gaps in the service control literature are highlighted. Finally, an illustrative example is discussed whereby the developed comprehensive view allows for interpreting service control decisions and their impacts on passenger experience in a more informed manner than previously possible due to the gaps in the service control literature.

SERVICE CONTROL LITERATURE AND PROPOSED ANALYSIS

FRAMEWORK

Although research on service control 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 (4), Koffman (5), and Abkowitz (6). Since then, much work has been done on the holding problem without real-time information, minimizing unreliability (and thus passenger delays) by optimizing the location of control points and slack times allocated in the timetable with the help of simulation, analytical methods, and dynamic programming models. A good overview of this research can be found in (7). Virtually all of these studies were limited in scope to routine holding and dispatching techniques.

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 service control interventions such as short-turning, expressing, and deadheading. This development creates a strong tendency to use optimization models, calibrated with real time data, which seek to find the optimal control strategies under disrupted conditions. Some important work in this regard was done by O’Dell (8) and Shen (9). O’Dell developed a real-time decision support tool for service recovery on a rail line after a disruption. She derived a linear programming model, which included holding and short-turning as feasible strategies for responding to a disruption under a certain set of constraints, with the overall objective function being to minimize passenger waiting time. Shen followed a similar path in model development as O’Dell, but generalized it to allow any train to be held anywhere, and combined this with expressing and short-turning strategies. The resulting model was a deterministic, mixed-integer program with a dwell-time sub-model. Again, the objective is to minimize passenger delays, but passenger delays both at stations and on trains were considered, and the trade-off between them was examined. Other research on real-time service control was done by Eberlein (10), Adamski and Turnau (11), and Puong (12). Most recently, Walker (13) studied the simultaneous recovery of a train timetable and of the crew roster after a single disruption on a simple rail line. The problem was solved with an integer program in which train departure times could be modified and crews could be swapped across platforms.
In air transportation research, the recovery of aircraft and crews with respect to schedules and assignments, respectively, has generally received more attention than in rail transportation (see, for example, (14) or (15)). However, the models that have been developed to solve these recovery problems, though similar, are not directly applicable to urban rail transportation. This situation is mostly due to the fact that many airline recovery models and formulations make use of specific system characteristics to simplify the problem and make it tractable and, as a result, they are not readily transferable to the urban rail operations recovery domain. Nevertheless, some concepts and principles are pertinent and should be considered in developing systematic recovery strategies for the urban rail case as is discussed in the conclusions section of this paper regarding future research.

The above and practically all other research on rail service control assumed that the only objectives in service control are related to passenger travel time or passenger wait time (with the exception of Walker (13), who did not consider passengers), and focused heavily on headway regularity. This narrow focus is problematic, as it does not consider other important decision factors in service control, which may be seen as objectives or constraints limiting the options available to a service controller. To date, there are only two pieces of research that recognize the complexity of a controller’s decision environment, though in different ways: Rahbee (3) and Froloff et al. (16). Rahbee examined the problems associated with the operation of rail transit lines. His research was not strictly limited to service control, but rather took a comprehensive approach towards improving service quality on rail transit lines. With respect to service control, Rahbee stated three main investigative goals that an analyst can pursue:

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

Froloff et al., on the other hand, wrote a manual for the RATP (The Autonomous Operator of Parisian Transports), which builds strongly on the practitioner’s point of view. It focuses on bus service and was originally written in an effort to identify and systematically categorize objectives, constraints, and techniques for bus service control in preparation for the design of a simulation system for controller training. One of the most important elements of the manual is an exploration of the relevant decision factors, objectives, and constraints in bus service control.

The goal of this study is to take a new approach to the study of service control by using automatically collected operations and passenger travel data, which are increasingly available to transit agencies, coupled with actual observations of a service control environment allowing for an informed interpretation of the analysis of the data. Specifically, operational procedures and service control interventions are analyzed using data combined from several sources, thus allowing a multi-perspective approach. To do so, an integrated framework, shown in Figure 2, is developed by Carrel (17) for studying service control interventions recognizing the presence of multiple factors relating to operating and demand conditions that need to be considered by service controllers when making intervention decisions in the presence of disruptions. Operational procedures along with service control interventions, and their impacts can be analyzed and understood most comprehensively by examining data from multiple sources reflecting service control, operating level of service, and the passenger experience. Thus, a set of measures are proposed to capture these three main ingredients of operations on a line. This framework was applied in a series of cases of the London Underground Central Line to demonstrate its value in a practical setting, gain insights into three common service control strategies, and assess the impact of a timetable change on the way service control is performed on the line (17).
One of the most important elements of the framework is a description of the service controller’s decision environment with all of the objectives and constraints that were observed during extended visits to the control center of the Central Line. That description, presented subsequently in this paper, is intended to help researchers, practitioners, and analysts understand a service controller’s decision environment, with its objectives and constraints. The intent is to improve cross-divisional communications within an agency by defining a common knowledge base and by giving management and planning staff an understanding of what causes service controllers to make the dispatching decisions that are observed in daily operations. Furthermore, it can help researchers move towards more realistic models, which could serve as the basis for strategic planning, operations planning, and control decision-making. The remainder of this paper focuses on the observations of a service control environment and the development of the factors which drive the process, followed by the application of the above framework to one service control example for the purpose of illustrating the importance of developing such a comprehensive understanding.

CONTROL ROOM OBSERVATIONS AND IMPORTANT DRIVERS OF SERVICE CONTROL

The visits to the control room took place over two weeks in January 2009 to observe how service controllers managed the line and to discuss their actions with them to understand why a certain strategy was chosen over alternatives. These activities reflected a largely informal and unstructured process. Beyond observing daily service management on the Central Line, these visits also explored the response to past disruptions using the automatically collected operational data and service controllers’ input to reconstruct the context and the response. Hypothetical situations as well as observed strategies were also discussed with the service controllers. The results were distilled into a set of decision factors, which were presented to service managers\(^1\), asking them for their input and assessment.

As previously discussed, the two main components of the operations plan are the train timetable and the crew schedule. A significant deviation from those plans can trigger a service control intervention, either to correct that deviation or to avoid a conflict. However, aside from the elements of train and crew management, several other factors (described subsequently) have an influence on service control decisions. Generally speaking, the constraints on a service control decision can be related to these factors

\(^1\)The service manager is in charge of a shift of service controllers.
or to provisions in the operations plan, and there are complex interactions between these decisions, constraints, and provisions. In order to understand the rationale behind service control interventions, a recognition and understanding of these factors and the interactions is necessary.

In general terms, service control needs a set of simple, real-time performance measures through which the operations on a line are evaluated and compared to the target state defined by the service and operations plan. These performance measures largely define the decision rules and priorities for service regulation. Common supply-centric measures can be headway regularity, lateness of service with respect to the timetable, or total missed trips. Another supply-centric measure is adherence to crew schedules, however, it is not currently tracked by any major transit agency to the best knowledge of the authors. Passenger-centric measures might include total passenger delays or travel time reliability metrics, but they are generally much more difficult to calculate in real-time as well as to relate to the service variables that controllers can influence.

This section presents the set of decision factors that were observed to influence the decisions of service controllers. While efforts have been made to generalize the results to other metro lines and systems, the degree of importance of the individual factors will vary across metro lines.

**Manageability and Uncertainty**

While it is very difficult to identify the direct influence of these two factors on how train service is restored after a disruption, one can say that they define the overall approach controllers take to managing a problem. When a disruption occurs, controllers must react under severe time pressure and with uncertainty about the duration of the incident and what other problems might occur in addition to those immediately apparent. Therefore, in deciding between different recovery strategies, several controllers stated that they would try to avoid timetable changes, which they knew from experience might be misunderstood, disregarded by drivers, or not have the desired effect for some other reason. It was also observed that service controllers often tend to be self-regulating in terms of workload. During “quiet” times they may have more time to dedicate to isolated interventions, for example by communicating with the drivers to discuss crew relief issues. However, when more significant disruptions occur, particularly during peak hours, controllers need to manage their workload efficiently in order to tend to all needs while preserving the capacity to respond to unforeseen events or new information regarding the ongoing disruption. The result is that controllers will have a tendency to choose “simpler” intervention strategies, and narrow the scope of line management during a disruption to “keep the service running,” as many controllers described it. This strategy can be understood as saying that the objective is to meet the level of service requirements (as outlined in the following section) over the relatively short time span that a disruption is affecting the line, and that other considerations or a longer-range view may not enter the picture until the recovery phase. As it is often necessary to act quickly when an incident occurs, controllers may not have time to think through multiple permutations of possible solutions. Their choices will therefore not necessarily reflect the optimal solution as could be determined by post-analysis, but they will generally choose a solution which meets all current constraints, is feasible under time pressure, is manageable, and is flexible, to deal with under various uncertainties.

**Level of Service**

Providing an adequate level of service on the line is probably the most important objective of service control. Yet, exactly how this is defined depends on agency policies, the layout of the line and the information available to the controllers via the operations control system. As previously mentioned, service controllers need to work with a set of real-time performance measures. The most important of these measures is schedule adherence, which relates to the following two objectives:
• operating as many of the scheduled trips as possible, and
• operating those trips with as little lateness as possible compared to the timetable.

These objectives are based on the premise that the timetable is the optimal output for the system, even in the case of disruptions. Further examples of supply-side criteria are headway regularity and, on lines with several branches, the sequence of destinations of trains on the trunk section.

On low-frequency line sections with published timetables, passengers obviously care about schedule adherence, but on high-frequency sections 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 a substitute for schedule adherence, which is very important for the efficient management of assets (e.g., trains, infrastructure capacity, and staff). Control room observations revealed that conflicts between multiple performance measures and the dominance of schedule adherence as a criterion can cause controllers to perform interventions that are solely focused on improving schedule adherence, thus potentially introducing irregularities and unreliability into the service. Furthermore, the observations also revealed that in the absence of an official agency policy, controllers will define their own understanding of what constitutes “good” service in terms of schedule adherence, headway regularity, or traffic patterns, but not always taking into account the impact of those assumptions on passengers.

Crew Management

Without a doubt, crew management is one of the most complex aspects of service control. It has a direct impact on how the train service can be operated and is governed by numerous rules and regulations, which are agency-specific since they usually depend on labor laws and agreements with unions, but generally they include provisions about maximum driving times (per day or period) and minimum layover and break times. These constraints are all accommodated in the original crew scheduling process, but they can be challenging when service is disrupted and trains need to be rescheduled.

Inclusion of Crew Management Considerations

Crew management can enter the controller's decision as a constraint or as an objective. Drivers cannot exceed their maximum driving time, and any driver stepping off a train at a relief point needs to be met by a relief driver. Furthermore, some drivers have a firm time constraint regarding when they must step off in order to meet other obligations. The result is that driver lateness is a large concern, and since drivers are tied to vehicles and both are assigned to schedules, driver lateness is directly linked to train lateness. To illustrate the importance of crew management, one can think about how it affects various service control strategies. Holding trains often means transitioning them into a delayed state. In the case of trains upstream of a blockage, holding exacerbates the delay resulting from the blockage. Trains held downstream of a blockage are added to the pool of late trains, assuming that they were on schedule before the disruption. Thus, after the blockage has cleared, controllers need to deal with a larger number of late crews. Short-turning trains also causes them, and more importantly, the drivers, to be out of sequence. A driver on a short-turned train may pass the crew relief point either too late or too early, in which case no relief driver may be available.

Since crew management problems generally arise only when the service deviates from the operations plan, crew and train lateness often occur simultaneously, and service control interventions cannot be attributed to only one of those problems. That is, since a change to a train’s trajectory is also always a change to the trajectory of its driver, the strong linkage has an influence on how controllers restore the service after disruptions. The essence of many conversations with controllers was that the choice of train on which to perform an intervention, and when to do it, is strongly driven by when and where crew reliefs are scheduled to take place.
Two additional crew-related problems are worth mentioning. First, many operations control systems do not report crew lateness, and it can be time consuming for a controller to get critical information on driving time constraints and lateness of crews on the line, especially if doing so has to be achieved through direct contact with drivers and crew managers. In situations where decisions must be made under time pressure, controllers may not be able to go through such communications and may instead choose recovery strategies in which as many crew constraints as possible are met, even those which would not have been binding. Second, at every crew relief there is a risk that something will go wrong due to a misunderstanding or staff error. In a worst case scenario, at relief points with sidings, a train can be withdrawn from passenger service if there is no relief driver available. However, at relief points without sidings, any problem with crew relief can easily cause delays to the service, and therefore controllers may be reluctant to make any changes to crew reliefs scheduled at those locations.

**Utilization of Spare Drivers**

Generally, transit agencies have a pool of spare drivers available to cover for absences and late drivers who need to step off due to a hard time constraint. Having this pool reflects a large cost component for agencies, which they naturally attempt to minimize. The availability of spare drivers is a function of the specifics of the non-spare duty that needs to be covered and is also dependent on how much driving a spare driver has already had to carry out during his or her shift. Moreover, discussions with controllers revealed that the utilization of spares could be a function of the number available. The larger the pool of spare drivers, the more controllers are inclined to utilize them for interventions or reliefs, thereby reducing the need for service curtailments (such as short-turns or cancellations) because of late drivers.

**Rolling Stock Management and Maintenance**

If a disruption occurs along the line, as opposed to at a terminal, it is often the case that trains are blocked in only one direction. However, in the case of a prolonged blockage in one direction, controllers may need to stop the service in the opposite direction as well to maintain rolling stock balance, which is important because an imbalance in the distribution of train units on the line presents large problems for service recovery. On a line with multiple depots, rolling stock imbalance is also of concern due to both limited depot capacity and the fact that the number of trains in a depot overnight should equal the number of scheduled pullouts from that location at the beginning of service next morning. A further trigger for service control interventions can be the rolling stock maintenance schedule, especially if the different depots specialize in different types of maintenance work.

**Terminal Capacity**

The capacity of terminals and reversing points has a direct influence on the propagation of delays throughout the line, since the layover time scheduled at terminals helps modulate headways and puts trains back into sequence or on schedule. Maximum layover time is a function of terminal capacity and the train arrival rate. Hence, if the scheduled arrival frequency at a terminal is increased or if trains become bunched due to delays, the maximum possible layover time is reduced, and controllers potentially need to perform interventions (such as short-turning trains before they reach the terminal) in order to avoid terminal congestion. Terminal congestion happens most often when high frequencies are scheduled (i.e., during peak hours). If many crew reliefs are scheduled after the end of the peak, it is likely delays will carry over into the crew relief period. These delays from the peak hours may then trigger control interventions, thereby negatively affecting the level of service in the off-peak and lengthening the recovery process.
As already discussed, the academic literature often focuses on minimizing passenger travel time or waiting time as the main objective of service control. However, conversations with service controllers revealed that passenger impact is a relatively fuzzy concept in daily operations management. Performance metrics such as excess travel time and average wait time are not available to controllers in real-time, let alone with high enough precision to evaluate alternative service control interventions. Therefore, controllers must work with assumptions and past experience. Observations revealed that the impacts of disruption management on passengers during the recovery phase generally enter into the decision process as a set of constraints in the following areas:

- **Crowding and congestion:** Overcrowding of platforms and trains can cause operational problems and disruptions. Therefore, holding trains downstream of a disruption can be problematic as it leads to crowded trains, and trains cannot be short-turned or diverted at stations lacking the capacity to absorb the additional number of passengers who are forced to alight.

- **Passenger complaints:** One of the shortfalls of the aggregate measures of passenger impact is that they generally do not account for the fact that additional waiting time and inconveniences such as crowding can be valued very differently by passengers depending on the individual situation. Although there is no easily deployable measure of “passenger (un)happiness”, one can use the frequency and reasons for passenger complaints as a proxy. Controllers often have a strong (but subjective) “gut feeling” about how different types of service control interventions cause customer discontent.

- **Availability of alternatives:** On a line with multiple branches, where successive trains are often bound for different destinations, the availability of alternatives to passengers affected by a change in service forms part of the consideration when controllers need to divert, short-turn, or withdraw trains.

- **Total excess journey time:** In general, a controller can try to minimize the impact on passengers by performing interventions that are not time-critical (i.e., not in direct response to an incident) outside peak hours and on low-demand sections of the route, and, if there is a choice, by selecting the least-loaded trains for time-critical interventions.

Two safety-related decision factors in service control were observed:

- **Passenger safety:** In-tunnel holdings are the second point of concern aside from platform crowding. Because evacuating passengers from a train is significantly more problematic when the train is held in a tunnel rather than in a station, and because the temperature inside a crowded train with closed doors can quickly reach critical levels, holding trains at stations rather than in the tunnel upstream of an incident is preferable. This has an influence on the nature of the timetable deviation observed after the incident and on the interventions necessary to “repair” the service.

- **Maintenance crew safety:** Depending on agency policies and procedures, certain train trips may be published for track maintenance crews as the last trips over certain line sections and must therefore imperatively be operated as scheduled.

The effect of infrastructure maintenance requirements on daily train service is generally foreseeable, as maintenance work is usually planned in advance, allowing time to provide customer information and to
develop an alternative operations plan. However, during the autumn and winter months and on lines with open track sections, sandite application to prevent leaf buildup and track de-icing may be required, both of which have an unplanned impact on service and need to be coordinated by controllers in real-time.

**Energy Management**

A further issue is energy management. Energy costs generally form a large enough part of a line’s operating cost to be of concern to the transit agency. The most relevant decision a service controller would need to make in this regard would be about an early or late shut-off of traction current on a line section.

**Synthesis**

In summary, it can be said that many of the aforementioned decision factors for service control interventions (such as crew and rolling stock management) are directly related to passenger service. That is, there is evidence that the reliability of the system depends on factors that have barely been recognized so far, let alone monitored or modeled. It was observed that in performing interventions to account for these issues, controllers could actually cause gaps in the service and delays, which would likely require remedies at a later stage. In other words, service control not only “repairs” unreliable service, it can also cause unreliable service.

Observations on the London Underground Central Line, which has a fairly complex line layout, suggested that controllers tend to emphasize adherence to the operations plan. This tendency is likely to be due to the limitations of human processing capability in dealing with the complexity of crew and rolling stock management tasks which increases with increasing line length, running times, and the number of crew and train depots along the line. Managing a long and complex line when, for example, every train is running either off schedule or on an unscheduled trajectory would hardly be feasible for controllers, even if the service to passengers (in terms of headways and destinations served) were very good.

Finally, the more complex a line’s layout, the more options are available to controllers for recovering service, which tends to cause controller variability in the way different controllers deal with the same disruption or timetable deviation – even in the presence of the tendency discussed above.

**ILLUSTRATIVE EXAMPLE**

The following section presents a case observed on the Central Line, illustrating how crew management considerations can influence service recovery after disruptions. The strategy of short-turning trains was selected as a focus since it is a very common disruption recovery strategy on the Central Line, and because of all service control interventions it is probably the one that is the most frustrating for passengers and leads to the largest number of passenger complaints.

**Description and Analysis Results**

The focus of the analysis is two disruptions that occurred on the Central Line on April 3 and November 12, 2008. Both disruptions caused very similar delays, but there were significant differences in the way service was restored. Before describing these events in detail, some information on the line layout is necessary. The part of the line of interest is the trunk section, which has a total of 22 stations between Leytonstone in the east and North Acton in the west. West of North Acton, the line has two branches: a short one to Ealing Broadway and a long one to West Ruislip. There are two crew relief points on this section of the Central Line: one at the eastern end at Leytonstone, which has no sidings, and one at the
western end at White City (two stops before North Acton), which is adjacent to White City depot and has storage capacity for trains. FIGURE 3 shows a line schematic (not all stations are shown). Both disruptions occurred on the trunk section westbound, as summarized below:

- **April 3, 2008, 10:27 AM**: A train sat at Queensway station (6 stops before the western end of the trunk section and 3 stops before White City) for 14 minutes resulting from a passenger emergency alarm. When the disruption cleared, there was a cluster of delayed trains upstream. For recovery, a total of 9 branch-bound trains were short-turned at the end of the trunk section or diverted from the West Ruislip to the Ealing Broadway branch in order to save cycle time.

- **November 12, 2008, 08:28 AM**: A train suffered a delay of approximately 17 minutes on the approach to Liverpool Street station (16 stops before the western end of the trunk section) due to technical problems. Again, when the disruption cleared, there was a cluster of delayed trains upstream. For recovery, a total of 10 branch-bound trains were short-turned in order to compensate for the delays.

Despite a similar number of short-turns on the two days, the patterns were different. On April 3, practically all trains after the initially blocked train were short-turned or diverted, and all of those trains were in a compact group. An observer standing downstream of Queensway station would have first seen the gap caused by the disruption, followed by the train that was originally blocked, and then the group of trains that were later short-turned and diverted. Approximately half an hour passed between the first and the last train of that group, after which the service returned to normal. On the other hand, on November 12, the short-turned and diverted trains were much more dispersed, with a maximum of three trains in sequence being short-turned. To the observer at Queensway, the short-turned and diverted trains passed over a period of approximately one hour in small groups among trains that were run to their original destinations.
Analyses of data from Transport for London’s smart card system, the Oyster card, showed that delays to passengers traveling to and from the western branches were more severe on April 3 than on November 12. Not only were the average travel times higher, but service on April 3 was also less reliable than on November 12, as could be seen from the travel time distributions on the affected OD pairs\(^2\). On November 12, passengers on many of those OD pairs experienced a 95th percentile travel time between 1 and 7 minutes higher than on a “typical” (i.e., non-incident affected) day, but on April 3, passengers on some OD pairs experienced 95th percentile travel times of up to 18 minutes, or even in one case, 22 minutes higher than that of a “typical” day. The methodology of defining these metrics and conducting the calculations is described in detail by Carrel (17).

**Crew Management as a Possible Cause**

In both cases, the disruption caused a gap in the service and a cluster of late trains. This warrants a set of interventions in order to improve the level of service on the line. However, controllers performed these short-turns much more aggressively on April 3 than on November 12. Crew management provided a very likely explanation for these differences. The term “likely explanation” is used here because this could not be verified with the service controllers, although the data strongly suggest it. On April 3, of the nine trains that were diverted or short-turned, six were on their last or penultimate trip before a crew relief. Three of the crew reliefs were scheduled to take place at Leytonstone eastbound, two at White City eastbound and one at White City westbound. Furthermore, one driver was later replaced by a spare driver at a station that was not a regular crew depot, thus suggesting he or she was at risk of breaching a driving time constraint.

It is understandable that service controllers become nervous at the idea of a train arriving late for a crew relief at Leytonstone, since a problem with a crew relief at that station can create a significant blockage, as there is no possibility of storing trains due to the lack of sidings at Leytonstone.

On November 12, on the other hand, only four of the ten short-turned and diverted trains had an impending crew relief, and only one was located at Leytonstone eastbound, while three others were White City reliefs (1 westbound, 2 eastbound). Given the concentrated distribution in time of the short-turns and diversions, it appears that on April 3, the larger number of impending crew reliefs was the cause, whereas on November 12, the crew relief constraints were not as binding, resulting in better spacing between trains selected for short-turns or diversions. The difference in the tightness of crew relief constraints was because the disruption on November 12 occurred at around 08:30 AM, while the disruption on April 3 occurred at 10:30 AM, when many of the drivers who had stepped on for the morning peak hour were about to be relieved.

Moreover, upon inspection of the trains that were not short-turned or diverted, one finds that controllers appear to have used spare operators on November 12 to avoid having to short-turn many of the trains following the gap despite some impending crew reliefs, but did not do the same on April 3. Although the data do not allow a complete reconstruction of crew movements and do not offer insights into the availability of spare crews, this raises the question of why spare crews were not used on April 3 in a similar manner. One realistic hypothesis is that the controllers were under much higher time pressure on April 3, since the blockage happened only 3 stations before White City (where trains could be short-turned and crew pickups were scheduled), and that controllers did not have the time to focus on driving time constraints and spare driver availability. Therefore, they may have decided to “play it safe” and short-turn all trains with impending crew reliefs. On November 12, the time lag between the clearing of the disruption and the arrival of the first trains at White City was larger, thus allowing the controllers to dedicate more time to crew management issues.

\(^2\)The sample was comprised of the OD pairs between the trunk section and the western branches in both directions.
In light of the above possible explanations of the differences between the two disruption cases, crew management could play a dominant role in conceiving and implementing an intervention plan. And, such dominance could easily result in the overriding of other considerations commonly assumed as the primary objectives of disruption recovery, namely passenger wait and on-board travel times.

CONCLUSIONS

Previous research into service control has usually assumed that the primary objective of service control is the minimization of passenger travel time in the face of unreliability caused by external factors. While that may be an important aspect of service control, it by no means constitutes all the objectives and constraints that face a service controller. As shown in this paper, the set of decision factors which cause service controllers to perform interventions or which influence interventions performed for other reasons is appreciably larger. The main drivers of service control are considerations about the level of service to passengers, crew management, rolling stock management, safety, and infrastructure maintenance.

Aside from these considerations, virtually all decisions are influenced by uncertainties regarding the outcome of an intervention and concerns about the manageability of the service. It was seen that the reliability of the system depends on many endogenous factors that previously may not have been recognized. In the absence of official policies or effective decision support, the management of these factors is often governed by rules of thumb. In fact, service control not only works to manage unreliability caused by exogenous events but can also be the cause of unreliability as controllers work to meet other objectives and constraints. There is a conflict between service quality perceived by the passenger (e.g., regularity of headways on a high frequency service and reliable travel times) and the other objectives and constraints that service controllers have to satisfy, such as crew management.

These conclusions imply that any effort to improve service control (and thus, operations in general) on a specific metro line must build on a solid understanding of how that line operates. Thus, there is significant value in directly studying service control interventions on a line in order to identify the most promising aspects for improvement. Furthermore, building a better understanding of the interaction between service control, the line characteristics, and scheduling variables can help improve the scheduling process since it shows where the provision of operational flexibility in the operations plan best helps controllers maintain a stable and reliable service. Last but not least, it would be desirable for performance metrics to take into account not only passenger-centric variables (such as wait time and travel time), but also supply-side variables such as crew lateness and the number of changes that service controllers need to make to a timetable in order to deal with infeasibilities in daily operations. Thereby, an agency can acknowledge the tradeoff between the provision of spare capacity and the need for service control interventions, and the performance of service controllers can be discussed in light of these diverse facets of their task.

This research is only a first step, and naturally more research is needed to explore the area of service control and its interactions with planning and performance monitoring on a metro line. However, given the strong dependence of service control on agency-, system-, and line-specific variables such as crew work rules and the location of reversing and storage facilities, the possibilities for studying service control at an abstract level (i.e., not focused on a particular agency or line) are limited, and it would mostly be up to individual transit agencies to take the initiative and dedicate the necessary resources to better understand service control on their network. Although starting up such an internal research program may be difficult at first, the results would be well worth the effort.

Important research topics that need to be addressed in this regard are related to the improvement of performance metrics, the design of recoverable (or robust) operations plans and the provision of
information to service controllers. Among these issues, the extension of traditional performance metrics to include a broader range of undesired events (e.g., delays to passengers, drivers and rolling stock imbalances) should be studied first since this would allow researchers and practitioners to develop a better notion of what constitutes a robust or recoverable operations plan, and the development of better design criteria could build on the insights from such studies.

Regarding research conducted at a more abstract level, it is desirable that operations researchers focus on expanding models to include more decision factors beyond a simple consideration of total passenger delay. In doing so, drawing upon concepts and principles developed in other sectors with similar problems, such as aircraft scheduling and crew assignment recovery techniques in the airline industry, would be valuable. Nevertheless, unlike previous studies in the urban rail domain, which have mostly presented the models as tools to automate disruption recovery on metro lines without service controllers’ involvement, it would be helpful to consider the models as decision support tools for service control staff. By acknowledging the strengths as well as the limitations of optimization models, and by including considerations on the most effective ways of providing and displaying the information to service control staff, researchers can ensure that their findings have an impact on the everyday practice of rail service control. In practical terms, this could include the design of interactive decision support systems which allow controllers to make use of proposed optimal recovery strategies on certain parts of the line while interactively specifying desired boundary conditions or constraints and updating them as the system evolves.

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.
REFERENCES


