

The Relationship Between Maintenance and Service Planning, and Evaluation of Maintenance Strategies (or Should You Run Your EMUs Till They Drop?)

Lexcie Lu • MIT Center for Transportation Studies • lexcie@alum.mit.edu • (617) 436-9896
77 Massachusetts Avenue • Cambridge • Massachusetts 02139-4301

November Draft (version 018) – 14 November, 2003
Word Count: 7,151 Words + 1 Figure + 1 Table = 7,651 Words

Abstract

Preventative or periodic maintenance can improve fleet reliability, but requires substantial ongoing investment in the form of labor, materials, and service downtime. Under simple assumptions, analysis demonstrates transit systems may be overly conservative in setting service levels because they are undervaluing the benefits of service that could be added if vehicles were not held back for scheduled maintenance and spares trainsets. To maximize the benefit of the service, a combined measure should be used to account for these opportunity costs. The paper presents a theoretical framework and an example methodology of what that might look like.

The current paradigm in transit maintenance and service planning does not allow the trade-off between service reliability and frequency of service to be rigorously evaluated. It is mostly standards-based, when a revenue-passenger oriented, evaluatory approach might be more appropriate. The methodology presented in this paper explicitly maximizes social welfare, allowing different operations & maintenance (O&M) plans to be compared in terms of customer impact. Further work could enable the development of a tool to assist planning at a systems level by forcing trade-offs between transportation and maintenance disciplines in quantitative terms. Empirical data on the life-cycle costs of equipment, the customers' degree of perceived disutility towards unreliability, headway degradation, versus their expectation about service frequency and crowding conditions, is necessary before a useful tool could be developed.

Introduction

Railcar maintenance is a complex operation, where successes and failures are difficult to judge. Railcars interact with the infrastructure, customers, and the world at large to create delays. Delays can be created as a result of poor operating plan, unexpected events, but also maintenance failings (in various departments). Longer-term maintenance goals may conflict with short-term service reliability targets. Traditional performance measures such as mean distance between failure (MDBF) and availability can be adversely affected by railcar design and the maintenance regime. For instance, different definitions of 'failure', measurements at the vehicle, married-pair, and consist level, will all yield different answers.

The current research is divided into four major constituent parts: (1) literature review, in which formal treatment of maintenance are described, and the theory behind the design of maintenance regimes are explored; (2) documentation of the state-of-practice, through interviews with shop and other personnel; (3) documentation of the performance measures, and an incentive analysis; (4) development of a methodology

(including a decision support model) to evaluate operations and maintenance (O&M) plan alternatives. The decision support model still requires substantial work, although the methodology is generally valid, and should be explored in greater depth by transit authority analysts.

One of the stated goals of the current research is to find a more suitable performance measure for railcar maintenance operations. This turned out to be an incorrect objective. While the depot-manager types will always pride themselves on maintaining a high MDBF (mean distance between failures), senior managers ought to evaluate O&M from a systems perspective, in addition to maintenance effectiveness indicators. It is more important to understand the limitations of each performance measure and what factors affect each measure (thereby understanding which results are directly comparable), than to find one all-encompassing performance measure. The evaluation framework provides a way to directly evaluate the performance of the transportation system from a customer's perspective, and make service-planning decisions that are sensitive to maintenance realities, which could be more important than evaluating the maintenance effectiveness.

In a well-run maintenance operation, questionable practices that sacrifice the longer-term asset quality to meet short-term performance targets do not generally occur. The primary reason for this is because even though railcars are long-lived assets, the light-maintenance feedback loops are rarely of a long-term nature. Deferred light maintenance usually shows up as an in-service failure within a matter of months. As for heavy-maintenance, such as motor and cam-controller rebuilds, air and brake systems overhaul, deferred maintenance usually shows up in the form of a higher-cost rebuild, 'premature' scrapping, and persistently low reliability on a daily basis. If exceptional reliability is not critical, deferring and accepting a more costly rebuild could be economically sound, especially with high-utilization assets.

Cosmetic or functional deficiencies of railcars observed by the lay-person may result from design issues, and not deferred maintenance. Corrosion could result from poor choice of carbody materials or coating, or oversight in drainage design; poor ride quality could be a combination of inferior truck design and insufficient suspension maintenance. Thus, improving transit service requires three important elements: (1) appropriate vehicle specification, (2) good vehicle design, and (3) adequate maintenance.

Literature Review

Maintenance Programme Design

Haven (1980) pioneered a framework for designing transit vehicle maintenance programmes, which essentially comprises of the following steps: (1) Identifying the goals of maintenance, and the types of maintenance activities; (2) Assessing the costs and benefits of different maintenance activities; (3) Identification of systems, subsystems and parts onboard the vehicle; (4) Identification of the most constraining system – termed the primary system, is the part of the vehicle that requires the most frequent attention; (5) Scheduling preventative maintenance activities to satisfy operating, performance, plant and other constraints. The result, a set of scheduled maintenance activities, is termed the maintenance regime.

Haven characterized the goal of the maintenance department as: “to supply a fleet of safe and reliable vehicles of sufficient size, to meet the needs of the transportation department at the least possible cost”. The goal highlights four major objectives: safety, supply, reliability, and cost minimization. Significantly, Haven suggests that the purpose of the maintenance function is to support operations, and therefore must “meet” all demands, instead of looking at the two functions as competing for the same resource – vehicle-time. This paper differs from conventional analysis by attempting to explicitly measure the benefits of running the railcar, versus the benefit of having it in the maintenance shop.

Little (1991) developed a maintenance regime for freight railcars in North America by studying and documenting existing practices, then improving upon it by applying basic principles of reliability engineering. He observed that North American Class I railroads at that time performed very little scheduled or preventative maintenance, if any. Maintenance scheduling is difficult within the freight railroad environment, but improvement in opportunistic maintenance could lead to productivity improvements. Using a rigorous mathematical approach, Little developed a decision-support heuristic for determining whether parts should be preemptively replaced. Different railcar maintenance strategies were then tested using a simulation model that sought to minimize the costs associated with repair. The model showed that opportunistic policies performed well and were very robust, although far-sighted hard-time policies (where cars are subject to an aggressive replacement of parts at fixed intervals) could also perform well in certain circumstances.

Review of Maintenance Concepts

There are three main types of maintenance activities: (a) scheduled maintenance; (b) repair maintenance; (c) on-condition maintenance. Two types of failures can be distinguished: in-service failure, or an incidental failure. An in-service failure is detected at a time when replacement will cause disruption to services, while an incidental failure is detected while the equipment is already out of use. Severity of failures could be classified into three categories: (a) regulatory failure; (b) engineering failure; (c) catastrophic failure. Four main types of preventative maintenance strategies can be distinguished: (1) no preventative maintenance; (2) age replacement (intervals of length of time, or usage); (3) block replacement, where components are replaced at fixed intervals in addition to failure-related replacement; (4) opportunistic maintenance, where the replacement decisions are based on the state of the rest of the system.

Critically, Little distinguished between constant failure rate (CFR), increasing failure rate (IFR), and decreasing failure rate components (such as certain components subject to a burn-in period). Complex systems of IFR components will exhibit CFR behavior in the long run (Barlow & Proschan, 1965), an important result that favours on-condition policies over hard-time policies.

How is Railcar Maintenance Done?

To understand how maintenance can be scheduled (or to what extent scheduling is really prudent or even possible) and optimized for both efficiency and effectiveness, a thorough understanding of how railcar is maintained on a day-to-day basis is needed. Relying on the author’s experience at the former British Rail

(BR), interviews with a number of American transit agency maintenance personnel, and literature reviews, this section serves to document the current state-of-practice of maintenance regimes and scheduling policies.

Survey of The State of Practice

The former British Rail had developed three different maintenance regimes in its 50-year existence. Prior to 1980, equipment was brought to the Main Works for general overhauls periodically. This is essentially a variable-time “greedy” policy, where the shopping intervals depend on equipment performance and the equipment is subject to an aggressive replacement of parts while in the shop. Certain specialized high-speed equipment was designed with a ‘maintenance through unit-replacement’ regime, where capital spares of most modules were overhauled independently of the vehicle. After the 1980s, BR developed the New Maintenance Policy (NMP), consisting of a series of six to eight tiered examinations that occurred at fixed time or mileage intervals. The NMP bore some relationship to the framework proposed by Haven (1980), which is similar to rules recommended by the American Association of Railroads, and later adopted by the Federal Railroad Administration.

The Massachusetts Bay Transportation Authority (MBTA) Orange Line currently practices an essentially hard-time policy of inspecting railcars every 12,000 miles ($\pm 1,000$ miles) or 90 days (± 10 days), depending mainly on shop workload and if a car comes into the shop for repair close to the scheduled inspection date. The period was determined through an iterative process: initially the cars were inspected every 7,500 miles. As experience accumulated and incremental upgrades are carried out to train-borne equipment, the interval was progressively extended to the current levels. Component rebuilds or major inspections are conducted at variable intervals depending on the life-cycle of the component.

MBTA maintenance appears to have a very risk-averse culture where many components are pre-emptively replaced at the time of inspection, due to the serious service disruptions that could result compared to the cost of parts. The performance indicators stress reliability and availability much more than maintenance costs or the opportunity cost of the asset, which could explain the risk-aversity. The increasing monetary costs of legal liability claims may also contribute. The failure-aversity may not be in the customer’s best interest – one car in the carhouse undergoing pre-emptive repairs is one less on the road.

In some MBTA shops, continual use of data allows the maintenance manager to identify components that fail most frequently, and to identify problem cars. Maintenance and inspection efforts are then targeted in these areas. Maintenance managers interviewed appear to have a strong preference for hard-time policies. The shop has a very limited capacity to perform inspections, thus, transition to a variable-interval inspection regime could cause scheduling conflicts, resulting in railcars ‘tying up the hoist’ when it is required for other repairs (Lu, 2003).

Evaluation of The MBTA Practice

The general problem with hard-time policies (Little, 1991) is that, there may not be a “magic-number” interval which is the best time to do everything. As long as the life-cycle of individual components are not even multiples of the basic interval (a condition that most likely exists), hard-time policies would not be most efficient. Purely on-condition policies are also inefficient as they disregard the costs of failure and potential savings in joint maintenance activities.

The inspection cycle is constrained both by the physical plant and personnel availability at the MBTA. However, some improvements are possible. The reliability of the fleet may be improved by introducing tiered maintenance plan whereby certain components are inspected every other inspection cycle. If reliability could be improved to reduce repair maintenance workload, total costs could be reduced. Productivity may also improve through higher utilization of residual component life, although their costs are not generally significant compared to opportunity costs of the railcar, and the labor.

Although the MBTA keeps track of mileage and failure data for the fleet (MBTA, 2002), it would benefit from collecting detailed cost data in parts and labor, and delay-minutes data by failure mode. Failure mode data is currently kept through dispatcher reports, which can sometimes be inaccurate. Some shop personnel in charge of repair keep records of the precise mode of failure, however, the quality of data is inconsistent and very little analysis is routinely done to identify problem components. These are potential areas of improvement.

Better cost data would enable better cost-benefit analyses. Typically, the cost of failure in-service will greatly outweigh any maintenance savings. If satisfying the market demand requires far fewer railcars than available (such that cars sitting out of service incur zero opportunity costs), significant benefits are most likely to accrue by minimizing failure rates, since the cost of failure is much more significant than the preventative maintenance costs. On the other hand, if a more aggressive preventative maintenance program will result in lower asset utilization, the benefit of reliability gain must then exceed the economic cost of the program.

Performance Measures

There are four basic types of maintenance performance measures, corresponding to the four main objectives of maintenance: (1) Instances of high-risk technical failures (potentially a safety violation); (2) Mean usage between failures (a reliability measure – usage measured in miles, time, or cycles); (3) Availability (a supply measure); (4) Maintenance cost or productivity (a cost-effectiveness measure). The tension between availability and reliability on the road is well known. Some of the measures are adversely affected by variables outside the maintenance manager’s control – apart from maintenance effectiveness, factors such as vehicle design, the weather, and fleet age all affect reliability. The caveats are documented elsewhere, and senior managers should be aware of these issues if chastising maintenance managers. Little (1991) discusses a number of alternate performance measures, such as mean distance between maintenance events, and distance per in-service failure, which are designed to balance preventative with repair maintenance. Some of these could be adaptable to a transit authority environment.

Maintenance Performance – Social Externality Interactions

In literature on maintenance performance measures, the interaction between costs and other performance measures are not often discussed. Traditionally, maintenance has been considered to be a support function that provided whatever the transportation department required. The systems approach towards operations would require costs and benefits of different operations & maintenance plan to be evaluated together quantitatively. The state of practice in evaluating maintenance effectiveness has tended to specify the required availability and reliability, minimized the number of high-risk technical incidents, while remaining within budget. This usually led to an over-emphasis on availability and reliability, while insufficient attention is paid to costs, both direct maintenance costs and the opportunity cost of vehicle time. If a real issue is made of maintenance costs, safety could become neglected over time, as evidenced in Philadelphia streetcars in the 1970s, and privatized British Rail in the late 1990s.

The MBTA case study provides a good illustration of this phenomenon. Anecdotal observation of shop-level initiatives suggest many maintenance managers are very efficient local optimizers – meaning that they are apt at reducing maintenance costs locally while maintaining the same reliability and availability. For example, Riverside Carhouse of the MBTA Green Line pioneered a new method of inspecting a critical part on articulated trolleys, tripling the throughput compared with methods envisaged by the manufacturers. Another initiative to reduce wheel-wear rates is currently being debated. However, often the interaction between opportunity cost of assets and maintenance is neglected, since those costs fall outside the jurisdiction of the maintenance manager (and sometimes the transit authority). For instance, work scheduling is sometimes done to maximize utilization of repairers and physical plant, and not the vehicle, even against a background of chronic car shortages. Maintenance managers sometimes indicated that reduced service levels will lead to greater reliability and lower maintenance costs, without considering the societal costs of spilled or delayed passengers resulting from service reductions. An approach is needed that considers all these costs and externalities within the same framework: direct maintenance costs, societal impacts of unreliability, unavailability, and service reductions.

Average passenger value-of-time and transportation demand are key variables in these frameworks, with the result that very different maintenance regimes could result from different assumptions on these variables. The externalities resulting from unreliability, unavailability, and service reductions are all highly sensitive to transportation demand and alternative methods of transportation available. Although it could be potentially problematic to introduce these assumptions into the performance measures, it could induce goal-directed behaviour where both the transportation and maintenance departments ultimately see their mission as maximizing social externalities. It would also serve to better integrate service planning and maintenance planning.

Coogan (2002) suggested that in transportation management, good practice keeps accountability low and performance measures high in an organization. Social externalities are a high-level measure, representing the customer perspective – which can be broken down into lower level measures that ensure accountability. Using these ideas, a decision support framework is developed, detailed in the next section.

Operating Plan Decision Support Framework

During a discussion with an MBTA maintenance manager, an operating plan was proposed for the Orange Line, which would decrease the peak fleet requirement from 102/120 cars to 84/120 cars. The proposal involves decreasing the rush-hour headway from every 5-minutes to 4-minutes, but running four-car consists instead of six-car. In theory, given the unique nature of Orange Line entryways that are located towards the ends of island platforms, capacity in the middle part of the consist would be better utilized with shorter trains.

How do we quantify the effect of possible overcrowding versus the benefit of increased maintenance? The traditional paradigm for transit service planning (Algorithm 1) do not account for these subtle effects. In effect, the planner assumes that a certain availability is achievable by the maintenance department based on historical experience or vendor specifications. This can occasionally result in unrealistic service plans when new vehicles are introduced, because the reliability characteristics and maintenance requirements of new vehicles are rarely well understood. In the revised framework (Algorithm 2), the performance of the fleet is considered a variable and not an assumed constant. Time spent in operations and maintenance are traded off to explicitly maximize the consumer surplus.

Maximize Service

Subject to {Operating_Budget, Crew, Vehicles, Fleet_Performance, Policy}

Alter (Operating_Plan)

Performance Measure {Dropped_Trips, OnTime_Arrival%}

Maximize Fleet_Performance

Subject to {Maintenance_Budget, Plant, Personnel, Engineering, Safety, Operational_Requirements}

Alter (Maintenance_Plan)

Performance Measure {MDBF, Availability}

Maximize Consumer Surplus

Subject to {Budget, Crew, Vehicle, Policy, Engineering, Plant, Safety}

Alter {Operating_Plan, Maintenance_Plan}

Performance Measure {Average_WaitTime, Headway_Variation}

Algorithm 1: In the traditional paradigm for transit service planning, MDBF and availability are maintenance output measures and are treated as constants in operations planning. Maintenance Department ‘failed’ if a specific level of MDBF was not delivered, while Transportation Department assumed a specific level of fleet availability.

Algorithm 2: In the revised framework, the only output measure the planner is concerned with is the consumer surplus of the riders. Instead of maximizing MDBF, we *optimize* MDBF to generate the best Operations & Maintenance Plan. Customer-oriented performance measures encourage headway adjustments based on a specific day’s actual fleet availability.

To demonstrate how a tool might be developed to implement the revised framework, a decision support model sensitive to operating plan was developed to test the Orange Line hypothesis.

Consist Level Issues (Set MDBF versus Vehicle MDBF)

If railcars functioned independently of each other, the consist MDBF would be a function of its length, and the vehicle MDBF. The vehicle MDBF would be a function of maintenance effectiveness and other external variables. In some failure modes, e.g. automatic train operations (ATO) problem, the failure of equipment in the lead car could compromise the entire consist. With an air leak, failure aboard any vehicle could compromise the whole consist. In other cases (e.g. traction motor cut-out), failure of a single car

would not affect the consist. Nonetheless, it might increase the probability of failure of other cars in the consist, because other motors have to work harder to compensate for the loss of tractive effort. Other interactions are possible in fleets that are semi-permanently coupled.

MBTA failure data treats each two-car set (married pair) as a basic entity. Thus, if one car has a traction motor cut-out, the whole two-car set is considered defective, but not the entire consist. Although the remaining vehicles in the consist that did not ‘fail’ will be returned to service relatively swiftly, service is disrupted, and all the management costs associated with an actual failure would be incurred on the entire consist. At other transit properties, fixed-formation sets of up to five cars are seen, giving rise to unique issues in measuring reliability. There are other subtle effects that may occur due to the way the transportation operation is managed, e.g. six-car consists may run all day with up to two trucks unpowered, resulting in vehicles being dragged in a failed condition, yet continuing to accrue mileage. To assess accurately the failure rates and the consequence of failures, detailed consideration of these effects would be required. At this point, it suffices to say that the current approach ignores such consist-level effects, and considers everything at the set level.

Structure of the Model

An example model might consist of five separate parts of economic evaluations of the rush-hour operating plan: (1) the economic cost of spilled passengers, due to inadequate capacity provided to carry rush-hour loads; (2) the cost of providing operators for the consists, which varies with the operating plan; (3) the cost of unreliability, which depends on the probability of failure and the impact of failures on customers; (4) the cost of waiting time, which depends on the headway; (5) the expected life-cycle cost of the equipment, based on the operations & maintenance (O&M) plan. In the current example, values are enumerated using fairly standard modelling methodologies in some level of detail, with an aim to capture the quality of service as seen from the perspective of the average rush-hour customer. Several O&M plan scenarios were then used to drive the model, resulting in a comparative evaluation from a public-benefit perspective. The current analysis ignores any budgetary constraints, although obviously this shortcoming could be mitigated if only the subset of O&M scenarios that are actually feasible from a budgetary standpoint is fed to the model.

The Passenger Spill Model

The transit passenger spill model was conceptually inspired by airline revenue management spill models developed by Hopperstad (1998), Belobaba and others. In a given time period where transit demand exceeds transit capacity, passenger spill occurs. Unlike airlines where seats are individually assigned, transit capacity is a “soft” number. Spill, or passengers beyond planned capacity, is likely to result in one of three outcomes: (a) overcrowding on the train as passengers pack themselves in beyond design capacity; (b) queueing delays to passengers unwilling to endure overcrowding; (c) loss of transit ridership either temporarily or permanently as passengers seek alternative transportation or residential/employment situations.

In Boston, the Central Transportation Planning Staff (CTPS) provides average passenger boarding and link-flow data at 15-minute resolution. Using the link-flow data, and theoretical capacity provided in each O&M scenario, passenger spill is calculated according to the following expression:

$$\text{Spill} = \sum_{\substack{\text{all time} \\ \text{periods}}} D - \left(\frac{1}{bt} \cdot sc \right) \quad \text{where}$$

- D = Transit link-flow demand
- b = Headway, in hours
- t = Time periods, in periods per hour
- s = Capacity per vehicle
- c = Vehicles per train

Clearly, this model assumes deterministic transit operations (i.e. that no train-bunching occurs), deterministic transit demands within one time period, and that spilled passengers simply disappear without impacting the demand in the following time period. These assumptions are obviously not true, and are likely to underestimate the actual spill that occurs. However, using a conservative number for capacity per vehicle, which represents the nominal capacity that many transit authority declare in their service standards, could mitigate these shortcomings.

The volume of passenger spill could then be methodologically converted into an economic dollar value – the consumers’ willingness to pay to avoid being “spilled” by a particular subway train. Using the economic idea of value-of-time from passenger demand forecasting, the construct here is entirely theoretical, and may not translate into actual monetary values. The spilled passenger is assumed to wait until the next time-period and take a following train, thus the social externality is simply the product of value-of-time and time-period. For simplicity, it is assumed that capacity would be available in the following time-period (if not, the passenger will choose to endure overcrowded conditions). This will tend to underestimate the economic cost, but is realistic since it reflects real-world passenger behaviour. Many refinements are possible, however, this model represents a first-cut effort at trying to explicitly estimate the societal costs of providing inadequate capacity during the rush-hour. Exhibit 1.2 illustrates a particular run of the transit passenger spill model.

The Probabilistic Maintenance Impact Model

The maintenance impact model is designed to explicitly evaluate the consequences of deferring maintenance actions. Assuming safety standards are satisfied, maintenance actions could be treated as a function whose inputs are costs, and whose outputs are some measures of maintenance effectiveness (e.g. MDBF). The trade-off here is one between maintenance costs and probability of in-service failure. This model extends from previous approaches by explicitly considering the opportunity cost of rail equipment out of service due to maintenance actions, in addition to out-of-pocket costs. Theoretically, this model could be used to evaluate the economic consequences of an O&M plan. First, we evaluate the risk of an in-service failure occurring during a given rush-hour period, $p(\text{Fail})$:

$$p(\text{Fail}) = \frac{1}{\text{MDBF}} \cdot m l c$$

- MDBF = Mean distance between failures, for the fleet
m = Rush-hour mileage, per vehicle, per day
l = Daily rush-hour line-up requirements (number of consists)
c = Vehicles per consist

Consideration of the operating plan outside the rush hour is not necessary since small changes in the number of units made available at the beginning of the workday would only affect the service level during the rush-hour. The vehicle requirements were then used to calculate the expected social cost due to in-service failures using the following expression:

$$\text{Cost}(\text{Fail}) = p(\text{Fail}) \cdot d \cdot \frac{1}{2} r v$$

- d* = Average delay per failure
r = One-way rush-hour ridership
v = Average passenger value of time

The MBTA considers all railcar-related incidents that cause more than four minutes' delay a failure. The average delay per failure (*d*) is assumed to be approximately ten minutes, based on January 2002 track circuit data compiled by an automatic computer programme (81 events, average 7.67 minutes, variability 0.48 – Wile, 2002) and the author's experience. The $\frac{1}{2}r$ term is an estimate of the number of passengers impacted by a given incident. Given that the failure will occur at a random time during either the morning or the afternoon rush, on average the number of passengers impacted would be the one-way rush-hour ridership divided by two. This logic assumes that a 10-minute delay to one train will delay all following trains by the same amount, thus delaying on average half of the total rush-hour ridership. It also assumes that repair to the trainset between the morning and afternoon rush is possible, which is the case on the Orange Line. The passenger value-of-time (*v*) was assumed to be \$20 per hour – higher than usual transit practice, chosen to reflect the more onerous nature of unscheduled delays. The result is the expected passenger-minutes wasted due to railcar unreliability, as illustrated in Exhibit 1.1.

The Operating Cost Model

Operations & maintenance (O&M) plan affects traincrew costs in a highly significant way. It costs approximately 50% more to provide service with four-car consists compared to six-car. In this implementation, an operating cost model linear with operator-hours was used, assuming a fully-attributed operator cost-per-hour of \$40. Optimizing models could give more accurate results, but this model is adequate for the present purpose.

The Waiting Time Savings Model

O&M plan also changes the customer level-of-service. Assuming deterministic operations, average waiting time for all passengers is reduced if the headway is reduced. In this model, waiting time is assumed to be $\frac{1}{2}b$. The economic benefit of decreasing the headway is calculated using this expression:

$$\text{Cost}(\text{Wait}) = \frac{1}{2}b \cdot 2r v$$

b	= Headway, in hours
r	= One-way rush-hour ridership
v	= Average passenger value of time

The Life-Cycle Cost Assumption

For the purposes of this demonstration, the operating plan changes were assumed not to affect the life-cycle cost of railcars significantly. The O&M plan changes examined (during the rush-hour) could increase overall fleet mileages by 5~15%, and will have some effect on the replacement cost of consumables. However, these costs are minor compared to labor costs and benefits of service. A more sophisticated tool would include a life-cycle cost model, calibrated for the current and replacement equipment.

Operating Plan Evaluation

For each O&M plan, a partial evaluation of social externalities can thus be given by combining the results from the four simple models and one assumption presented above:

$$\text{Cost}(\text{O\&M}) = \text{Cost}(\text{Fail}) + \text{Cost}(\text{Spill}) + \text{Cost}(\text{Wait}) + \text{Cost}(\text{Traincrew}) + \text{Cost}(\text{Life-cycle})$$

Other terms can be added as deemed appropriate. This is not an absolute cost-benefit analysis of whether the service is economically efficient; it is instead a comparative evaluation tool that could be used to rank O&M plan alternatives in order of social costs. The cost function aims to take into account all costs that vary with different O&M plans, whether they are internal costs to the transit authority, or social externalities imposed on transit riders and the public at large. O&M plans appropriate for the Orange Line are compared in a worksheet illustrated in Exhibit 1.3.

Results from the MBTA Orange Line Study

The current O&M plan in effect on the Orange Line (base case), which calls for an 85% fleet availability in the morning peak, has a combined social and internal cost of about \$65,000 per day given the above assumptions. If the fleet requirements are decreased from 102 vehicles to 68 vehicles (57% availability) by reducing consist length to four-cars, assuming the fleet is 50% more reliable due to increased maintenance actions, the combined cost nonetheless rises to about \$97,000 per day. The excessive passenger spill will outweigh any reliability benefits. If the fleet requirement is held constant, but formed in four-car instead of six-car consists, the combined costs drops to \$37,000. This suggests the benefit of reduced waiting time far outweighs any operator costs. Theoretically, if every vehicle in the fleet were used (100% availability), but MDBF halves as a result of reduced maintenance, the combined costs of \$53,000 is still lower than the base

case. On the other hand, if MDBF reaches 25% of its former value, the combined cost could be double the base case even with improved frequency.

In general, while increased operator costs due to increased frequency are significant, the benefits associated with reduced failures (due to reduced fleet requirements) dominate operator costs. However, in a capacity-constrained situation, the spill costs dominate both. If the reduced wait-time for existing riders is considered a tangible benefit, its value dominates all costs. The resulting operating plan looks essentially like a bus operation – with many operators, very short consists, and high frequencies. The incremental benefits of enhanced maintenance, the costs of unreliability and spill are all swamped by the need for frequency, since frequency changes affect all riders while unreliability or spill only affect a small proportion of riders.

Discussion

Some of these results are exceedingly counter-intuitive, if not controversial. Essentially, this cost-benefit analysis suggests that if riders valued their time with the same dollar value regardless of the circumstance, then reliability really does not matter. Given the current level of technical performance achieved by electric trainsets, since failures are rare, keeping spare trainsets to cover for failures or allow additional maintenance to enhance reliability is a futile exercise. According to the analysis, the trainsets are always better employed on the road, where they are reducing waiting time (through enhanced service frequency) for all riders. Since traincrew costs are insignificant compared to the level of social benefit they generate, the allegedly optimal strategy is to run as many trainsets as you've got every morning, and run 'em till they drop. Spill is almost a non-issue because it occurs only at the peak of the peak.

There are number of problems with this analysis, from a strictly technical perspective. Firstly, the capital costs of acquiring trainsets have not been taken into account. This absurd result may have its origins in trainset acquisition when the fleet size was determined from the peak loadings. If the fleet size was determined by rational analysis rather than the need to cater to every traveller at a time that they choose to travel, it is likely that the marginal trainsets would never have been justified. Secondly, this evaluation is essentially a marginal analysis, thus none of the larger costs and benefits has been taken into account – instead we have assumed they are constant. It is possible to consider a spilled passenger as value lost rather than time lost – the trip might be worth \$10 to consumers who could not get on the train because it was full. Different conclusions might then be reached. None of the larger questions have been considered, such as the use of transit versus automobiles, and congestion pricing.

Equipment Failures, Headway Degradation, and Consumers' Perception of Unreliability

Intuitively, there are other issues with this analysis. Are the values-of-time really the same? Survey of regular subway commuters will likely reveal that in reality, with perfect schedule adherence, there is little difference between headways of five-, four- and three-minutes. On the other hand, there is a big difference between headways of five- and ten-minutes. Thus, all the alleged benefits of increasing frequency disappear, though the costs of operators and spill are real. An unexpected finding of this research is the realization that headway degradation is actually a significant economic cost, and that headway maintenance on a daily basis

(i.e. aggressive holding actions and timely terminal departures) may be more important than equipment reliability in contributing to passengers' perception of the precious commodity "transit service reliability".

Being stuck in a subway train for ten-minutes can be an unpleasant experience, especially for customers on their way to meetings. If offered the choice between ten journeys of seven minutes each, versus nine journeys of six minutes and one journey, chosen at random, of 16 minutes, most consumers will choose the reliable option. In a sense this becomes an equity argument: by keeping the trainsets out of service and running less frequent service, we are in effect distributing the failure delays amongst more people and different days by forcing a longer waiting time, thus the delays are more equitable. This has required a highly risk-averse approach towards failures in service, since they cause inequitable delays and thus generate complaints. Headway degradation may generate many more delay-minutes but receives much less attention. The value of reliability in the mind of the consumers need to be better established before this analysis could be more useful. Such a survey might apply conjoint analysis techniques that explicitly measure the consumer preference along the reliability-frequency trade-off interface.

What is the consumers' response to unreliability? The consumers' value function with respect to transit reliability may be highly non-linear. If long-run on-time performance to within five minutes is more than 99%, individuals might choose a 'just-in-time' train (as observed amongst Metra Electric commuters). At about 95%, individuals might allow one headway as buffer time. If performance is abysmal, a half-hour might be lost for every trip. An 'S-curve' could be calibrated to relate the proportion of passengers who would allow extra 'buffer time' against long-run on-time performance. The model could then be extended to account for lost productivity (a social cost) due to the possibility of being delayed, rather than actual delays. This extension is likely to significantly affect the results, if the O&M plan being analysed causes a level of unreliability that calls for buffer times of about one headway.

Further refinements to this model may require treating transit operations as stochastic – effect of overcrowding on headway reliability should be considered, along with the random nature of train breakdowns and variability in departure times. If overcrowding is not managed on the platform, passengers holding doors and other random disturbances can easily affect the headway to an extent that the benefit of increased train frequencies is negated by headway degradation. The extent of headway degradation is not a simple function, and is highly sensitive to local variables such as number and location of train doors. Further data collection is necessary to calibrate a model. An implicit assumption in the present model is that at any given time, the maintenance shop has perfect knowledge as to which train will breakdown next, and therefore will not send out the marginal train except when absolutely necessary (i.e. in rush-hour service). Without such information, it is necessary to consider the disruption an unreliable trainset may cause in the off-peak hours.

The Effect of Reduced Maintenance on Equipment MDBF

How MDBF would change with the maintenance regime is a function of the proportion of failures that are 'preventable' (i.e. attributable to inadequate or shoddy maintenance) versus 'random' (i.e. effects of unforeseeable events such as weather, previously unknown design failures, etc.) For MDBF to reach 25%

of its former value in a reduced maintenance scenario would require a 75% of the current delay-minutes to be caused by preventable failures. Actual failure data would improve this aspect of the model. Results here are necessarily vehicle-specific and empirical.

Preliminary attempts were made to correlate historical Boston temperature data with failure modes, using MBTA Failure-in-Service System, Vehicle History Report (Orange Line 07/01/02 to 01/23/03). With a dataset containing 687 incidents (including 118 service interruptions), a correlation of 0.64 was found between temperatures of under 25°F, and a moving average of a composite failure index of air, brake, and door components (See Appendix 1). In one interpretation, this suggests 10% of all service disruptions are attributable to inclement weather causing drainage valves to freeze and rubber to fracture, and therefore no improvement is possible here with preventative maintenance. Another interpretation suggests that regular rebuilds of valves and better rubber materials may help. A third source believes even though preventative maintenance may not mitigate all air system faults in cold weather, a combined maintenance/operating strategy may still yet decrease service interruptions. Freight railroads operating in the Great White North have long limited train lengths to 60~70 cars during the winter, to allow for changes in mechanical system performance due to increased leakage and denser air. In addition, wheel and track fractures increase dramatically in cold weather also, giving rise to further sources of unreliability.

Whatever your belief is with respect to cold weather, other failure modes exist for which there is no preventative remedy. Broken windows and slashed seats are the result of vandalism, while wheel flats could result from a faulty brake system, a malfunction in the signalling system, or repeated misuse of the emergency brake by traincrew. However, these only account for 9% of total incidents and 3% of service interruptions. Nonetheless, these are areas of maintenance that most customers perceive as representative of the general state of the system. To what extent maintenance should focus on rectifying mechanical malfunctions versus cosmetic deficiencies, is another topic worthy of research. Artifacts in the MBTA data suggest such cosmetic issues are commonly under-reported. Failures have also not been categorized consistently.

Theoretically it is possible to attribute delay minutes (results of service failure) to one of the following categories: (a) equipment failure related delay, likely triggered by an external condition; (b) other unexplained equipment failure related delay – possibly preventable; (c) delays unrelated to equipment failures – to include infrastructure failures, resource allocation issues, and perturbations of service resulting from the stochastic element of the system (i.e., dwell times and train operator characteristics). Earlier work on freight rail reliability demonstrated that only 30% of shipment delays are related to rail technology (i.e., equipment plus infrastructure). For a detailed review of extent literature, see Kraft (1998). A future study to classify causes for delay may reveal a similarly low loading on total delay minutes of equipment failures, with the implication that even a significantly lower MDBF will not seriously hamper service reliability if other improvements could be made, such as more aggressive headway management. The loading is necessarily a function of route length, headway, and resources available to recover from an equipment failure event – and could conceivably lead to different conclusions for different operating environments.

The Evaluation Framework

Although the framework clearly warrants further refinement, the evaluation methodology is clearly valid and extremely powerful. The traditional view that the maintenance department exists and serves at the pleasure of the transportation department is no longer tenable. Instead, a systems-approach to service planning is required, with the opportunity cost of assets being explicitly traded off between transportation and maintenance disciplines. Evaluation should be in terms of social benefits and costs – both externalities and real costs to the transit authority, such as operator wages.

The spreadsheets shown in the exhibit demonstrates that a reasonably simple decision-support model could be developed without sophisticated computer systems or access to vast arrays of data. Both conceptual and computational or data-based refinements will improve this model. Some assumptions also need to be verified through field exercises and customer surveys. It is the customer who will ultimately judge the success of the transit authority – in both transportation and maintenance.

The moral of the story might be stated as: maximizing reliability, maintenance actions, or service all are not the “right answer”. Instead, a multi-objective analysis is needed to explore alternative maintenance regimes that could improve service without hugely detrimental effects on reliability – for instance, maintenance of trainsets at night while targeting 95% availability.

Conclusions

There are many nuances of maintenance operations and performance indicators that require careful thought before using them to make decisions. The maintenance managers (and line chiefs, responsible for both transportation and maintenance) who are familiar with these caveats may achieve better planning. Operations planning for transportation and maintenance should be done concurrently, using a cost-benefit analysis evaluation framework. Instead of the current standards-driven requirement for availability, minimum service levels, reliability, and other such proxies of service quality, trade-offs between these important variables should be explicitly considered. Assigning monetary values to each of the variables is possible, provided consumer research could back up these values.

MBTA could improve its effective maintenance operations by moving towards a more objective-driven evaluation framework, and perhaps by moving towards a variable-interval maintenance regime. There are many innovations that could be tried, such as maintaining trainsets overnight instead of during the day, tracking costs much more aggressively, and studying the mechanical failure profile of the train components more closely to make informed decisions about when to replace.

In the same way that safety standards of the past has moved towards risk assessment and risk management paradigms, service standards and reliability standards ought to be moving in the direction of service cost-benefit analysis and service management. More research is definitely needed in this area to establish exactly what a cost-benefit analysis would mean in the context of maintenance and transportation management, and market research is required to identify and act on customer preferences. Although MBTA may be able to

obtain a more efficient maintenance operation through changes in the maintenance regime, at some level the existing maintenance regime is functional and supports operations well. Identifying and quantifying (with a dollar value) the transportation department's needs in terms of customer preference should precede any attempt to change the maintenance regime to achieve either higher reliability or availability – perhaps high reliability or availability may not be economically efficient.

Acknowledgements

The author wishes to acknowledge Bob Doyle, Jack McLaughlin, Brian Dwyer of the MBTA for assistance with data collection, Erik Wile of MIT for processing data, Andreas Aepli of Reebie, Pat Little of Harvey Mudd College for stimulating thoughts, and Prof. Nigel H.M. Wilson for his wonderful guidance as well as all of the above.

Appendix 1: Algorithm to calculate correlation between air system failure and inclement weather

Variable	Explanation
AirFault =	Number of air system related incidents today (Air Leak, Low Air, Compressor, Train Dumping)
AirIntrp =	Number of air system related service interruptions today
BrkIntrp =	Number of brake related service interruptions today
IsHalfDoor =	Number of reported incidents of 'Half-Door Shut Off' today
IsSideNotOpen =	Number of reported incidents of 'Side Door Will Not Open' today
FailIndex =	$0.2 * \text{AirFault} + 0.8 * \text{AirIntrp} + \text{BrkIntrp} + 0.8 * \text{IsHalfDoor} + 0.2 * \text{IsSideNotOpen}$
FailIdx7 =	Moving Average of FailIndex over the past seven days, including today
MinTemp =	Minimum recorded temperature today
Freeze =	If $\text{MinTemp} < 25^\circ\text{F}$, Then $\text{Freeze} = 25^\circ\text{F} - \text{MinTemp}$
Correl =	Correlation between (FailIdx7) and (Freeze) = 0.64 for this dataset

References

Little, Patrick. *Improving Railroad Freight Car Reliability Using a New Opportunistic Maintenance Heuristic*. MIT Thesis (1991).

Haven, Paul. *Transit Vehicle Maintenance: Framework for Development of More Productive Programs*. MIT Thesis (1980).

Morrison, Gavin. Chapter 5, Maintenance Under The New Maintenance Policy, *The Class 47*. Ian Allen, London (1999).

Coogan, Matthew and Stanley, Robert. *New Paradigms for Local Public Transportation Organizations, TCRP Research Results Digest 55 (TCRP Project J-08B)*. Transportation Research Board, Washington D.C. (2002).

Wile, Erik. S. *Using Automatically Collected Data to Enhance Transit Operations: Data Source*, Proprietary Database. Massachusetts Institute of Technology, Cambridge, Mass. (2002).

Central Transportation Planning Staff. *Passenger Counts for the Massachusetts Bay Transportation Authority: Heavy Rail Subway and Light Rail Surface Lines*, Boston, Mass. (1995, 97).

Macchi, Richard A. *Expressing Trains on the M.B.T.A. Green Line*, MIT Thesis, Cambridge, Mass. (1990).

Deckoff, Anthony A. *The Short-Turn as a Real Time Transit Operating Strategy*, MIT Thesis, Cambridge, Mass. (1990).

Kinki-Sharyo Company. *Maintenance Handbook for Massachusetts Bay Transportation Authority No.7 Light Rail Vehicle*. Osaka, Japan (1986).

Massachusetts Bay Transportation Authority. *Heavy Rail Equipment Maintenance Department – Fiscal Year End Report*, Boston, Mass. (1991).

Massachusetts Bay Transportation Authority. *MCRS System Daily/Monthly Operating Reports (March)*, MBTA Proprietary, Boston, Mass. (2003).

Lu, Alex. *Comparative Analysis of Maintenance Practices, Proceedings of 1.259 Term Papers*, Massachusetts Institute of Technology, Cambridge, Mass. (2003).

Soeldner, David W. *A Comparison of Control Options on the MBTA Green Line*. MIT Thesis, Cambridge, Mass. (1993).

Kraft, Edwin R. *Rail-Related Literature Review (Ch.2)*, from *A Reservations-Based Railway Network Operations Management System*. Ph. D. Dissertation, Department of Systems, University of Pennsylvania, Philadelphia, Penn. (1998).

Exhibit 1.1 – Railcar Maintenance Delay Cost Model

Railcar Maintenance Evaluation Model											
Lexcie Lu, MIT Center for Transportation Studies, 11/05/03											
Delay Costs of Unreliability				Same Reliability as Base Case				More Reliable than Base Case by			50%
Base Case				4-car	4car, 4min	4car, 3.4m	6car, 4min	4-car	4car, 4min	4car, 3.4m	6car, 4min
Route Miles (one-way)			11.16 miles	11.16	11.16	11.16	11.16	11.16	11.16	11.16	11.16
Trip Time (one-way)			33 minutes	33	33	33	33	33	33	33	33
Recovery Time			9.5 minutes	9.5	9.5	9.5	9.5	9.5	9.5	9.5	9.5
Vehicle Cycle			85 minutes	85	85	85	85	85	85	85	85
Headway		every	5 minutes	5	4	3.4	4	5	4	3.4	4
Sets on the road			17 sets	17	21.25	25	21.25	17	21.25	25	21.25
Morning Rush between	6.25	10.25	= 4 hours	4	4	4	4	4	4	4	4
Afternoon Rush	2	7.25	= 5.25 hours	5.25	5.25	5.25	5.25	5.25	5.25	5.25	5.25
Rush hour mileage (per set, per day)			145.7 miles	145.7	145.7	145.7	145.7	145.7	145.7	145.7	145.7
Fleet set-mileage (per day)			2,478 miles	2,478	3,097	3,643	3,097	2,478	3,097	3,643	3,097
Units per set			3 units	2	2	2	3	2	2	2	3
Fleet unit-mileage (per day)			7,433 miles	4,955	6,194	7,287	9,291	4,955	6,194	7,287	9,291
Vehicles per unit			2 vehicles	2	2	2	2	2	2	2	2
Vehicles Required			102 cars	68	85	100	127.5	68	85	100	127.5
Rush hour ridership (per day)			126,606 paxs	126,606	126,606	126,606	126,606	126,606	126,606	126,606	126,606
Northbound am			27,950 paxs	27,950	27,950	27,950	27,950	27,950	27,950	27,950	27,950
Northbound pm			37,786 paxs	37,786	37,786	37,786	37,786	37,786	37,786	37,786	37,786
Southbound am			30,619 paxs	30,619	30,619	30,619	30,619	30,619	30,619	30,619	30,619
Southbound pm			30,251 paxs	30,251	30,251	30,251	30,251	30,251	30,251	30,251	30,251
MDBF gained as a result of smaller fleet				0	0	0	0	13,869	13,869	13,869	13,869
Failure Rate (unit MDBF)			27,738 miles	27,738	27,738	27,738	27,738	41,607	41,607	41,607	41,607
Expected rush-hour failures per day			0.268 failures	0.179	0.223	0.263	0.335	0.119	0.149	0.175	0.223
Delay per failure (minutes)			10 minutes	10	10	10	10	10	10	10	10
Expected passenger impacted per rush-hour incident			31,652 paxs	31,652	31,652	31,652	31,652	31,652	31,652	31,652	31,652
Value-of-Time (average rush-hour pax)			\$20 per hour	\$20	\$20	\$20	\$20	\$20	\$20	\$20	\$20
Delay pax-minutes due to unreliability (per day)			84,812	56,541	70,677	83,149	106,015	37,694	47,118	55,433	70,677
Delay cost due to railcar unreliability (per day)			\$28,271 per day	\$18,847	\$23,559	\$27,716	\$35,338	\$12,565	\$15,706	\$18,478	\$23,559

Exhibit 1.2 – MBTA Passenger Spill Data, 1997

Passenger Load Factors, MBTA Orange Line Northbound					
Data from Central Transportation Planning Staff, 1997					
Nominal seating capacity	48			Results	
Nominal standing capacity	36				
Cars per unit	2	Total AM Rush Pax Spill =	907	Vehicle cycle	85 minutes
Units per set (consist)	3	Total PM Rush Pax Spill =	1,550	Sets required	17 sets
Capacity per set	504	Spill Pax Q Time =	10	Vehicles required	102 vehicles
Headway (minutes)	5	Spill Pax \$ of Time =	\$20	Fleet	120 vehicles
Frequency (tph)	12			Operator cost	\$40 per hour
Seats per 15-min	1,512	Spill Cost =	\$8,190	Total cost	\$1360 per hour

Note: The 'D' term, transit link-flow demand at the 15-minute resolution, shown on a separate portion of the spreadsheet, is used in calculating the passenger spill numbers, and is not reproduced here. Interested readers should refer to the data source (CTPS, 1995, 97).

Exhibit 1.3 – Operating Plan Alternatives Evaluation

Passenger Spill Cost Analysis, MBTA Orange Line					
Lexcie Lu, MIT Center for Transportation Studies, 12/5/03					
Spilled Pax Q Time =	10	minutes			
Pax Value-of-Time =	\$20	per hour			
			Same Reliability as Base Case		
	Base Case	4-car	4-car,4min	4-car,3.4min	6-car,4min
Units per set (consist)	3	2	2	2	3
Headway (minutes)	5	5	4	3.4	4
AM Rush Spill Northbound	907	3,402	1,897	996	8
PM Rush Spill Northbound	1,550	4,107	2,514	1,639	502
AM Rush Spill Southbound	2,935	6,272	4,447	3,113	1,103
PM Rush Spill Southbound	0	1,262	506	0	0
Total Cost of Spill (per day)	\$17,973	\$50,143	\$31,213	\$19,159	\$5,377
Seats provided per 15-min	1,512	1,008	1,260	1,482	1,890
Vehicles Required	102	68	85	100	128
Operator cost (per hour)	\$1,360	\$1,360	\$1,700	\$2,000	\$1,700
Operator cost (per day)	\$9,520	\$9,520	\$11,900	\$14,000	\$11,900
Cost of Unreliability (per day)	\$28,271	\$18,847	\$23,559	\$27,716	\$35,338
Spill+ Operator+ Unreliability \$/day	\$55,764	\$78,510	\$66,672	\$60,875	\$52,615
ΔCost (per day) – compared to Base	\$0	\$22,746	\$10,908	\$5,112	-\$3,148
Assumptions:					
* Fully allocated operator cost is \$40 per hour					

ENDS