ISSN 1046-1469
Indexed by PAIS

JOURNAL OF THE TRANSPORTATION RESEARCH FORUM

Volume 34 Number 2 **1995**

TRANSPORTATION RESEARCH FORUM

THE LINK BETWEEN DEMAND VARIABILITY AND RAILROAD SERVICE RELIABILITY[1]

by *Edwin R. Kraft**

ABSTRACT

Railroads must improve transit time reliability, particularly on single railcar shipments, if they hope to remain competitive with trucks in the future. Service failures can result if there is a mismatch between demand and the amount of capacity provided. A simulation model is used to establish the link between demand variability and transit time reliability under several scenarios; tightly scheduled, loosely scheduled, tonnage-activated, and a scheduled, but power constrained railroad. Priorities are demonstrated effective in improving performance of the high priority traffic class in all except the tonnage-activated scenario.

The model results establish a clear relationship between the level of variability in demand and the transit time reliability produced by a rail transportation system. They establish that "precision execution" is a necessary, but not sufficient condition for achieving service reliability. It is also necessary to schedule sufficient capacity to handle demand peaking, or else to implement a traffic priority system. Since providing slack capacity to handle traffic peaks is expensive, carriers have a financial incentive to figure out how to make a priority system work. The payoff will be the ability to run <u>full trains with the right cars, on time, every day</u>. By breaking the paradigm that reliable service always costs more, a traffic priority system may be the key to solving the railroad industry's long-standing service reliability problem, while retaining the inherent cost and operating efficiencies of the rail mode.

INTRODUCTION

Railroads must improve transit time reliability, particularly on single railcar shipments, if they hope to remain competitive with trucks in the future. Customers require ever increasing quality and reliability in transportation service. In a survey of emerging logistics trends, OECD[2] found that:

> Customers do not ask for high speed in transport, nor for continuous information on the goods, nor for containers and quick load systems. They ask for reliable delivery of goods at the right time and place, without damage and at a fair price. Customers need to have the feeling that their goods transported are in capable hands.

In a study of the trucking industry, Allen and Liu[3] found that "the reason the existing motor carrier cost studies did not find that large carriers have significant cost advantages over small ones is not because the trucking industry has a constant returns to scale cost structure, but because large carriers have used up their cost advantages to provide higher quality services than small carriers. Since shippers are usually more sensitive to service quality than to costs (see Winston[4] and McGinnis[5]), high quality services and competitive costs provide shippers with strong incentives to switch to larger, higher quality carriers. This is characteristic of what has happened in the LTL industry: continuing bankruptcies of low quality service carriers and increasing domination by a few large and high quality ones." This finding strongly suggests that a low-cost, low-service strategy may not be the right formula for railroad success in merchandise freight markets.

A review of the literature found much research which focuses on network management with uncertain demand and on production of reliable service. This includes the stochastic modeling work of Powell[6] and Jordan and Turnquist[7]; the priority traffic classes proposed by Nozick and Morlok[8]; the SCAN train dispatching program of Jovanovic and Harker[9]; and the Service Planning Model by McCarren and Martland.[10] Unfortunately, much other research misses the target by discussing neither demand variability nor service reliability. Very little research was found (see Kwon[11]) which addresses the fundamental tradeoffs between demand variability, the amount of slack capacity scheduled and resultant railroad transit time variability.

Service failures can result if there is a mismatch between demand and the amount of capacity provided. If demand is higher than the capacity of the train, excess cars spill over to the next day's train, unless an extra train is operated, which may not always be physically possible or economic. If demand is much lower than planned, some trains may be annulled, consolidated, or held for tonnage, leading once again to unreliable transit times. There is a direct link between the variability of customer demands and the reliability of transit time produced by a railroad freight transportation system.

Management has some latitude to allocate capacity among different customers and traffic lanes, but in the short term only within fixed limits determined by the number of locomotives in the system, train crew availability, and requirements to reposition both crews and locomotives to handle future demands. Passing siding lengths and train handling considerations determine the maximum train size which can be operated over any route. But, there may not be sufficient locomotives to power all the trains at this maximum size.

In this paper, a simulation model is used to establish the link between demand variability and transit time reliability under several scenarios: tightly scheduled, loosely scheduled, tonnage-activated, and a scheduled, but power constrained railroad. Priorities are demonstrated effective in improving performance of the high priority traffic class in all except the tonnage-activated scenario.

A NETWORK SIMULATION MODEL

A network simulation model was built to test the effect of various operating strategies on transit time reliability of a single origin-destination pair. Large-scale network simulations can be so complicated that it is hard to see or understand what is really going on. This *simple* simulation is designed to be easy to understand and to effectively demonstrate how demand variability can affect service reliability. Having said this, even for a single origin-destination pair, there are a practically infinite number of combinations of operating strategies, demand uncertainty, slack capacity measured by the Capacity/Volume or C/V ratio, traffic mixes and priority systems which could be simulated. It is still necessary to be selective in the generation of alternatives and in presentation of results.

The purpose of this study is to explore in a "controlled" manner how demand uncertainty can cause service unreliability. The adverse reliability impact of late inbound train arrivals or irregular departures from classification yards is already well understood. So this factor has been excluded from the present analysis by assuming that all the trains always run on time. The purpose of this study is to demonstrate additional, less well-understood reliability effects due to cars overflowing scheduled capacity, in an environment where demand is not the same every day.

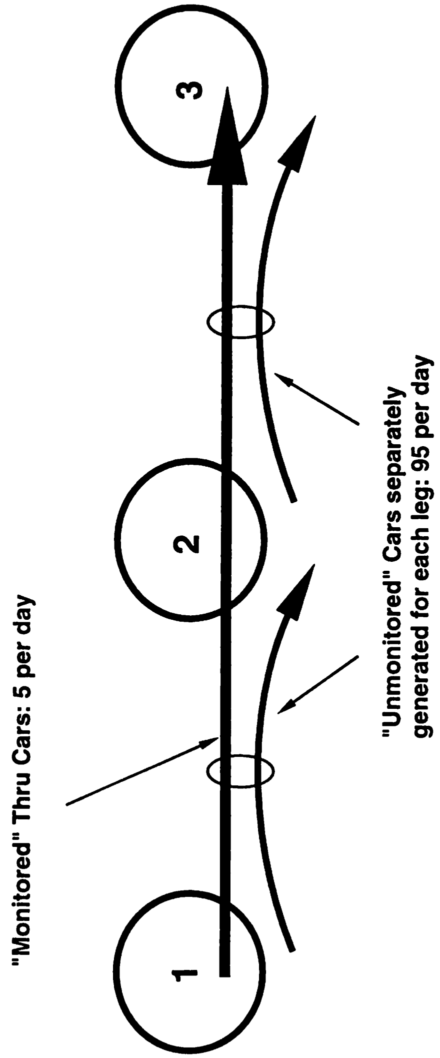
As shown in Figure 1, the simulation includes two endpoint terminals, one intermediate terminal and two road trains which must make a connection at the intermediate terminal. Only one group of "monitored" cars goes all the way through from terminal 1 to terminal 3; the performance of these cars is measured and reported. Other cars riding on legs 1-2 and 2-3 are randomly generated as needed; their sole purpose is to occupy space on trains, and these non-monitored cars are discarded as soon as they arrive at their destination terminal.

Major parameters affecting system performance include the level of demand uncertainty, the amount of slack capacity scheduled and if a priority system exists, the mix of high versus nonpriority traffic. The level of demand uncertainty was parametrically varied between zero (deterministic demand) and +/- 50% of the average cars per day. Variability in demand can be buffered by scheduling spare capacity, so different levels of spare capacity, as measured by C/V ratio, were simulated. Finally, some experimentation was performed with the implementation of a priority traffic system.

A Description of the Simulation Model

The model performs an event-oriented Monte Carlo simulation. Daily input traffic is assumed normally distributed having user specified mean and standard deviation. Train

Figure 1: The Simulated Network with Three Yards and Two Trains



departures occur on a scheduled basis, and each train has a maximum capacity. Cars fill train capacity on a first-in-first-out basis within each priority class. Excess cars beyond the capacity of the train must wait for the next train departure.

In the real world, congestion is likely to occur during traffic peaks, which would further degrade service reliability. The simulation model assumes that the yard has sufficient capacity to hold overflow cars without extending the required processing time. Thus, only the direct effect of exceeding scheduled train capacity is included in the model. Congestion effects are not included.

To simplify presentation of model results, the same level of uncertainty is applied to all demands in a single model run. This level of uncertainty is specified by one parameter, the coefficient of variability, defined as:

Coefficient of
Variability = Standard Deviation/
Average Daily Volume

For each model run, this coefficient was varied from a value of zero to .50 in steps of .05. Thus, 11 separate simulations were performed for each scenario to see how each strategy would perform under a wide range of traffic demand variability assumptions.

Each time an event occurs, the yard status is updated by making a call to a random number generator. This determines the number of cars which arrived since the previous event. As simulation proceeds, the cars arriving before, on, and after each train arrival are generated. Table 1 shows an example of how the number of arriving cars is calculated.

Beginning and end effects are handled by simply running the model for long enough to render them insignificant. A typical simulation generates in excess of 25,000 carloads over four and a half months of operations. Of these, usually about 625 are monitored cars. Each simulation requires about a minute and a half to run on a Macintosh SE/30. After the required number of cars have been generated, the simulation shuts down abruptly in a fully loaded condition, so that there are no "end" effects. Results are reported based on the number of cars which have successfully reached their destinations.

Simulation Model Results

The first scenario represents a limiting case where all connections are scheduled very tightly. Only three hours are allowed for each connection from inbound train arrival until outbound train cutoff time. The running time between yards (measured from scheduled cutoff time until arrival at the next yard) is standardized in all model runs at 10 hours. Obviously such tight scheduling would likely produce many missed connections due to irregularities in line operations causing late inbound train arrivals. However, these results do not take this factor into account.

Simulation results for the tightly scheduled scenario are contained in Tables 2 through 4. As expected, the deterministic demand case always produces 100% reliable service; but Table 2 shows that each 5% increase in demand uncertainty leads directly to one additional day in the "spread" in the O-D transit time distribution for this case. Tables 3 and 4 explore the effect of scheduling 10% or 20% spare capacity into the system; this improves the picture, but only up to a point. Scheduling 10% slack capacity "insures" against 10% variability in demand; 20% slack "insures" against 20% variability in demand, etc. Once the level of demand uncertainty exceeds the C/V ratio, the trip time distributions begin to spread out again.

This finding suggests that one way for the rail industry to ensure reliable service is to schedule slack capacity in the system; however, the provision of this extra capacity might be quite expensive. A traffic priority system can protect reliability for the most service-sensitive cars without needing so much slack capacity.

The second scenario represents a limiting case where all connections are scheduled very loosely. Twenty hours are allowed for each connection from inbound train arrival until outbound train cutoff time. Loose scheduling minimizes the probability of missed connections due to late inbound train arrivals. However, it has an additional benefit, since cars arrive shortly after the queue has been cleared out, they are "first out" for departure on the next train. Thus, the probability of overflowing the next train is quite small. Overflow would only occur if the previous train had also overflowed and left behind nearly a whole trainload of cars.

TABLE 1: CALCULATING RANDOM NUMBER GENERATION PARAMETERS

Input Data:

- Local Train scheduled arrival 7 AM with an average of 5 monitored cars
- Road Train scheduled cutoff time 12 Noon, with an average daily volume of 100 cars: 5 monitored, 95 unmonitored
- Coefficient of Variability = .30

Calculations:

- Daily Standard Deviation = .30 * 100 cars = 30 cars
  Variance = 30 * 30 cars = 900 cars
- Number of Unmonitored Cars = 100 cars - 5 cars = 95 cars
- Number of Hours "Prior to" monitored cars' arrival = Previous 12 Noon until 7 AM = 19 hours
- Number of Hours "After" monitored cars' arrival = 7 AM until next 12 Noon = 5 hours
- Pro-rate expected number of car arrivals:

  "Prior to" = 19 hrs/24 Hrs * 95 cars = 75.2 cars
  "On" = Specified as 5 cars = 5.0 cars
  "After" = 5 Hrs/24 Hrs * 95 cars = 19.8 cars
  TOTAL 100.0 cars

- Pro-rate Variance:

  "Prior to" = 75.2 cars/100 cars * 900 cars = 75.2 cars
  "On" = 5 cars/100 cars * 900 cars = 5.0 cars
  "After" = 19.8 cars/100 cars * 900 cars = 19.8 cars
  TOTAL 900.0 cars

- Take square root of variance to develop standard deviation:

		Mean Cars	Std Dev Cars
"Prior to"	=	75.2 cars	26.0 cars
"On"	=	5.0 cars	6.7 cars
"After"	=	19.8 cars	13.3 cars

TABLE 2: TIGHT CONNECTION SCENARIO

Transit Days Showing Number of Cars

Scheduled Capacity=100 Average Demand=100 Capacity/Volume Ratio=1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	645	0	0	0	0	0	0	0	0	0	645
0.05	83	314	237	0	0	0	0	0	0	0	634
0.10	37	181	370	29	0	0	0	0	0	0	317
0.15	17	102	302	172	23	0	0	0	0	0	616
0.20	8	71	115	286	120	10	0	0	0	0	610
0.25	5	56	95	157	212	83	1	0	0	0	609
0.30	5	54	71	77	174	175	47	1	0	0	604
0.35	6	50	50	53	126	228	42	35	8	0	598
0.40	6	48	37	47	90	181	102	55	29	1	596
0.45	5	48	37	31	95	116	135	77	32	10	586
0.50	4	44	39	27	59	102	122	98	56	31	582

TABLE 3: TIGHT CONNECTION SCENARIO

Transit Days Showing Number of Cars

Scheduled Capacity=110 Average Demand=100 Capacity/Volume Ratio=1.10

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	645	0	0	0	0	0	0	0	0	0	645
0.05	639	0	0	0	0	0	0	0	0	0	639
0.10	622	5	0	0	0	0	0	0	0	0	627
0.15	497	127	7	0	0	0	0	0	0	0	631
0.20	296	281	46	0	0	0	0	0	0	0	623
0.25	221	322	79	0	0	0	0	0	0	0	622
0.30	173	321	123	0	0	0	0	0	0	0	617
0.35	134	304	167	5	0	0	0	0	0	0	610
0.40	116	257	185	50	0	0	0	0	0	0	608
0.45	92	260	170	65	10	0	0	0	0	0	597
0.50	82	223	164	78	40	1	0	0	0	0	588

TABLE 4: TIGHT CONNECTION SCENARIO

Transit Days Showing Number of Cars

Scheduled Capacity=120 Average Demand=100 Capacity/Volume Ratio=1.20

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	645	0	0	0	0	0	0	0	0	0	645
0.05	639	0	0	0	0	0	0	0	0	0	639
0.10	627	0	0	0	0	0	0	0	0	0	627
0.15	630	1	0	0	0	0	0	0	0	0	631
0.20	570	53	0	0	0	0	0	0	0	0	623
0.25	454	161	7	0	0	0	0	0	0	0	622
0.30	339	241	37	0	0	0	0	0	0	0	617
0.35	270	279	61	0	0	0	0	0	0	0	610
0.40	222	308	78	0	0	0	0	0	0	0	608
0.45	206	303	88	0	0	0	0	0	0	0	597
0.50	190	297	106	0	0	0	0	0	0	0	593

Tables 5 and 6 show that loosely scheduled traffic can be expected to perform much more reliably than its tightly scheduled counterpart: the loose schedule acts as a form of "insurance," up to a point, against demand variability. In Table 5, the transit time distribution does not begin to spread significantly until variability reaches 15%. Provision of just a little (10%) slack capacity in Table 6 pushes this spread point almost off the chart making these connections very reliable. This finding suggests that <u>another way for the industry to improve service reliability is to intentionally schedule loose, rather than tight connections for priority traffic</u>.

Intentionally extending priority car schedules may not be acceptable from a transit time or equipment utilization point of view. However, some non-priority cars will inevitably be scheduled with long connect times. Such non-priority cars moving over a stabilized train service network should experience definite improvements in reliability. This finding supports the view that improvements to the performance of a priority subset need not always be accomplished at the expense of non-priority cars. To the contrary, by stabilizing core train network operations, service to a substantial portion of non-priority "free rider" cars can also be improved.

The third scenario represents a "demand actuated" or "tonnage" railroad operation. This was simulated as follows: First, train departures were scheduled every four hours throughout the day. Second, if a particular train departure would not be full, that departure was canceled and the cars allowed to accumulate another four hours. The most distinguishing feature of the tonnage railroad operation is that the train schedules take a "random walk." That is, the departure time of the next train equals the departure time of the previous train plus a random amount of time which depends on the arrival rate of inbound traffic.

Additionally, schedule "drift" will occur even with deterministic demand. The schedule will drift anytime daily demand does not precisely match scheduled capacity. Take, for example, the case where train capacity is 100 cars, but demand is only 80 cars. This leads to a fixed headway of 100/80 days or 30 hours. This train schedule will drift by an average of six hours per day, repeating itself every four departures.

The unstable nature of tonnage train schedules makes this operation difficult to plan and control. It often leads to emergency moves of crews and power, leading some to question whether it is even an economical operating strategy. Simulation of this is beyond the scope of the model employed in this study. What is known for sure, however, is that the tonnage railroad operation leads to a wide spread in the transit time distribution. Table 7 shows that the tonnage railroad leads to an almost constant spread in transit time regardless of demand variability. With volume in Table 7 sufficient to support daily train departures, train schedules drift to produce transit times distributed between the best and worst cases simulated in scenarios A and B. Lower traffic volume in Table 8 results in lower train frequency, increasing average transit time and significantly widening the transit time distribution.

One benefit of the tonnage railroad scenario, however, is the notable absence of extremely long transit times present in both the tightly and loosely scheduled scenarios. The high end of the transit time distribution has been clipped off. Demand activated operation appears to place an upper bound on the worst case scenario. Providing extra departures when needed eliminates the growth of persistent queues in terminals, which developed in the two earlier fixed capacity scenarios.

Promptly moving overflow cars out of terminals benefits more than just the cars moved: later arriving cars are now first out in the queue and can depart on their originally planned trains. Railroads can improve service reliability by <u>operating additional advance or second sections when demand warrants</u>. Thus a tonnage-activated or demand-responsive operation is not "all bad": if implemented on top of a stable core operation of scheduled trains, the demand activated operation can clip the longest trips out of the transit time distribution for nonpriority cars, while priority cars ride the scheduled network and maintain a tight distribution.

The fourth scenario represents a power constrained, but still scheduled railroad. This strategy arises out of consideration of just what is meant by "spreading the peak load" in a railroad context. Rail operations typically run in weekly cycles. On Mondays through Fridays, a

TABLE 5: LOOSE CONNECTION SCENARIO

Transit Days Showing Number of Cars

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	0	0	0	635	0	0	0	0	0	0	635
0.05	0	0	0	619	0	0	0	0	0	0	619
0.10	0	0	0	607	54	0	0	0	0	0	607
0.15	0	0	0	604	310	0	0	0	0	0	598
0.20	0	0	0	607	299	72	17	0	0	0	595
0.25	0	0	0	606	221	148	61	0	0	0	592
0.30	0	0	0	574	118	262	56	35	0	0	593
0.35	0	0	0	562	85	261	88	31	10	0	584
0.40	0	0	0	525	55	226	119	53	25	1	583
0.45	0	0	0	494	44	196	140	70	25	0	579
0.50	0	0	0	431	33	143	171	71	52	8	583

TABLE 6: LOOSE CONNECTION SCENARIO

Transit Days Showing Number of Cars

Scheduled Capacity = 110 Average Demand = 100 Capacity/Volume Ratio = 1.10

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	0	0	0	635	0	0	0	0	0	0	635
0.05	0	0	0	619	0	0	0	0	0	0	619
0.10	0	0	0	607	0	0	0	0	0	0	607
0.15	0	0	0	604	0	0	0	0	0	0	604
0.20	0	0	0	607	0	0	0	0	0	0	607
0.25	0	0	0	606	0	0	0	0	0	0	606
0.30	0	0	0	574	33	0	0	0	0	0	607
0.35	0	0	0	562	44	0	0	0	0	0	606
0.40	0	0	0	525	74	7	0	0	0	0	606
0.45	0	0	0	494	93	17	0	0	0	0	604
0.50	0	0	0	431	157	20	0	0	0	0	608

TABLE 7: TONNAGE RAILROAD SCENARIO

Transit Days Showing Number of Cars

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	0	275	355	0	0	0	0	0	0	0	630
0.05	8	302	305	0	0	0	0	0	0	0	615
0.10	0	203	437	0	0	0	0	0	0	0	640
0.15	6	251	337	5	0	0	0	0	0	0	599
0.20	3	278	333	24	0	0	0	0	0	0	638
0.25	0	299	333	17	0	0	0	0	0	0	649
0.30	1	302	292	35	0	0	0	0	0	0	630
0.35	6	205	375	43	0	0	0	0	0	0	629
0.40	0	220	362	37	0	0	0	0	0	0	619
0.45	7	235	360	45	0	0	0	0	0	0	647
0.50	0	244	321	10	0	0	0	0	0	0	575

TABLE 8: TONNAGE RAILROAD SCENARIO
Transit Days Showing Number of Cars
Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 2.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	0	0	331	650	249	0	0	0	0	0	1230
0.05	0	153	363	498	301	0	0	0	0	0	1315
0.10	5	164	561	518	95	0	0	0	0	0	1343
0.15	6	163	500	543	158	0	0	0	0	0	1370
0.20	7	168	523	545	148	23	0	0	0	0	1414
0.25	2	203	531	480	187	17	0	0	0	0	1420
0.30	1	168	544	538	216	14	0	0	0	0	1481
0.35	19	145	416	500	297	4	0	0	0	0	1381
0.40	3	188	527	515	160	10	4	0	0	0	1407
0.45	3	187	586	470	172	10	8	0	0	0	1436
0.50	7	154	530	501	226	24	0	0	0	0	1442

substantial portion of the locomotive fleet is tied up in local train service; meanwhile, these same locals are bringing cars into origin terminals to be moved. Few locals operate on Saturdays and Sundays, releasing locomotive power back into over-the-road service, while simultaneously reducing the number of new shipments brought into the system. The availability of power, the absence of local train interference on mainlines and suspension of maintenance gang activities makes weekends an excellent time to run trains.

Since locomotives are expensive assets, maintaining a high utilization rate is very important. Locomotive requirements can be reduced by intentionally allowing car inventories to build up during the week and then "cleaning out" the terminals with intensive road train operations over the weekends. This leads to severe power shortages during the week as insufficient locomotives are on the property to move all demand on a current basis. As shown in Table 9, weekly demand spreading causes a progressively increasing number of cars to miss their connections as the week goes on.

Thus the power-constrained operation is even "worse" than the tonnage railroad scenario: the Tonnage Railroad assumes a train will be dispatched promptly, as soon as enough cars are available to operate the train. But the "Power Constrained" Railroad doesn't even assume that, because not enough locomotives are available to cover peak demand within the week. Trains, even though ready for departure will be systematically held for power.

There are two possible management responses in the face of locomotive power shortages. One is to attempt to maintain maximum train length, producing irregular train departures; the other is to reduce the train length and power allocation per train as needed, but maintain regular schedules. By maintaining regular schedules, power and crews can be "cycled" on an efficient basis through preplanning and schedule coordination; emergency moves and "deadheading" can be kept to a minimum. All power assignments would still be fully utilized, just as they would be in the tonnage railroad scenario. Running less than maximum train lengths may increase crew costs above the tonnage railroad case. (However, it is a matter of current debate whether this increase would be greater than the cost of unplanned crew deadheading and emergency power moves which would be associated with a reversion to the tonnage strategy.)

As will be shown later, reliable service can still be maintained for a priority class of traffic if regular schedules are operated. Reverting to an irregular operation, however, would defeat any priority system and widen the transit time spread for all traffic classes. The strategy of maintaining regular schedules with reduced power allocation was simulated. Train departure times and total capacity provided were the same as in the "tightly scheduled" scenario; however, instead of scheduling, for example, 100 cars capacity on each outbound train, a pattern was used assuming 80 cars capacity on weekdays and 150 cars on weekends.

TABLE 9: Weekly Demand Spreading Causes an Increasing Number of Holdover Cars

(Starting with an empty yard on Sunday night:)

Day	# Arrivals	# Departures	# Holdovers
Monday	100	80	20
Tuesday	100	80	40
Wednesday	100	80	60

and so on.

To simulate slack capacity, the weekly pattern was scaled up by a constant percentage. Thus a C/V ratio of 1.10 would result in weekday capacity of 88 cars and weekend capacity of 165 cars.

Table 10 shows that the policy of weekend peak spreading results in considerable transit time spread even for the deterministic case, and transit time reliability deteriorates markedly as demand uncertainty increases. Provision of slack capacity in Tables 11 and 12 decreases transit time sensitivity to demand uncertainty, but any scenario which uses weekend peak spreading results in significant service unreliability.

Intermodal systems experience similar demand peaking. Many customers cannot accept deliveries over the weekends; consequently the intermodal peak generally occurs on Monday mornings rather than Friday evenings. Yet the fundamental problem is the same: weekend line haul capacity outpaces weekday delivery capacity. This leads either to origin yard queues building during the week, destination yard queues dissipating throughout the week, or both. Both kinds of queues adversely affect service reliability.

Priority Scenario Model Results

Recognizing that not all customers require the same service, this section explores the effectiveness of establishing a priority traffic class. Priority shipments would always have first rights to available train capacity; other shipments would move on a space available basis. A two-tier priority traffic system has been implemented by the French National Railways.[12]

At first glance, intermodal operations seem more amenable to the introduction of priorities than do railcar operations. Although sometimes inconvenient, there are no physical barriers to "cherry picking" priority shipments off intermodal trains in a mechanized terminal. There is evidence that priority systems already exist in intermodal operations and there is customer pressure to improve the responsiveness of these systems. As Jay Hirst (of Alliance Shippers) put it in Railway Age,[13] "Terminals have a bad habit of not being able to prioritize. Terminal operations just seem to allow for one set format to process trailers and containers, rather than being able to do what the customer requires."

Efforts to establish priorities in railcar operations have been generally implemented at aggregated levels, such as the block, train, or even service network levels. For example, several carriers have established dedicated train networks for automotive traffic. There are concerns that establishing individual railcar priorities would increase the amount of switching work in yards and even prove operationally infeasible. It appears that the key to implementing a priority system for individual railcars could be based on an adequate car scheduling system. Such a system could optimize switching instructions **before** the cars reach the hump or are classified, rather than attempting to "cherry pick" priority cars out of the middle of an already-classified cut of cars.

Two questions can be asked when evaluating the effectiveness of any priority system: first, is the system effective in improving service reliability and/or transit time for the priority traffic class? Second, what is the effect on the "other" traffic? Are the gains experienced by the priority class worth the costs imposed on the non-priority class?

Tables 13 and 14 report the impact of introducing traffic priorities into the tightly

TABLE 10: POWER CONSTRAINED RAILROAD

Transit Days Showing Number of Cars

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	95	90	465	0	0	0	0	0	0	0	650
0.05	16	61	415	150	1	0	0	0	0	0	643
0.10	16	50	257	246	61	0	0	0	0	0	630
0.15	16	48	161	246	153	7	0	0	0	0	631
0.20	16	29	86	165	222	96	9	0	0	0	623
0.25	17	28	70	139	179	151	28	10	0	0	622
0.30	18	29	57	66	104	200	108	29	6	0	617
0.35	16	30	48	63	77	158	155	46	17	0	610
0.40	10	33	38	53	64	131	136	111	26	6	608
0.45	5	30	42	57	45	109	137	111	51	10	597
0.50	4	28	34	53	49	56	138	112	66	53	593

TABLE 11: POWER CONSTRAINED RAILROAD

Transit Days Showing Number of Cars

Scheduled Capacity = 110 Average Demand = 100 Capacity/Volume Ratio = 1.10

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	280	90	280	0	0	0	0	0	0	0	650
0.05	217	142	284	0	0	0	0	0	0	0	643
0.10	175	192	263	0	0	0	0	0	0	0	630
0.15	162	219	247	3	0	0	0	0	0	0	631
0.20	142	212	259	10	0	0	0	0	0	0	623
0.25	134	212	266	10	0	0	0	0	0	0	622
0.30	115	213	277	12	0	0	0	0	0	0	617
0.35	107	204	270	29	0	0	0	0	0	0	610
0.40	104	179	239	85	1	0	0	0	0	0	608
0.45	83	168	222	114	10	0	0	0	0	0	597
0.50	69	156	200	133	29	6	0	0	0	0	593

TABLE 12: POWER CONSTRAINED RAILROAD

Transit Days Showing Number of Cars

Scheduled Capacity = 120 Average Demand = 100 Capacity/Volume Ratio = 1.20

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	489	107	54	0	0	0	0	0	0	0	650
0.05	369	221	53	0	0	0	0	0	0	0	643
0.10	307	234	89	0	0	0	0	0	0	0	630
0.15	283	238	110	0	0	0	0	0	0	0	631
0.20	257	241	125	0	0	0	0	0	0	0	623
0.25	245	245	132	0	0	0	0	0	0	0	622
0.30	222	258	135	2	0	0	0	0	0	0	617
0.35	215	254	136	5	0	0	0	0	0	0	610
0.40	200	261	141	6	0	0	0	0	0	0	608
0.45	178	256	156	7	0	0	0	0	0	0	597
0.50	166	260	158	9	0	0	0	0	0	0	593

TABLE 13: SCHEDULE PRIORITIES IN TIGHT SCENARIO
Transit Days Showing Number of Cars
for High Priority Traffic: 50/50 Mix

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	645	0	0	0	0	0	0	0	0	0	645
0.05	634	0	0	0	0	0	0	0	0	0	634
0.10	615	0	0	0	0	0	0	0	0	0	615
0.15	608	0	0	0	0	0	0	0	0	0	608
0.20	599	0	0	0	0	0	0	0	0	0	599
0.25	594	0	0	0	0	0	0	0	0	0	594
0.30	583	0	0	0	0	0	0	0	0	0	583
0.35	580	0	0	0	0	0	0	0	0	0	580
0.40	578	0	0	0	0	0	0	0	0	0	578
0.45	574	0	0	0	0	0	0	0	0	0	574
0.50	571	0	0	0	0	0	0	0	0	0	571

TABLE 14: SCHEDULE PRIORITIES IN TIGHT SCENARIO
Transit Days Showing Number of Cars
for High Priority Traffic: 80/20 Mix

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	640	0	0	0	0	0	0	0	0	0	640
0.05	634	0	0	0	0	0	0	0	0	0	634
0.10	620	0	0	0	0	0	0	0	0	0	620
0.15	595	18	0	0	0	0	0	0	0	0	613
0.20	555	49	0	0	0	0	0	0	0	0	604
0.25	470	111	18	0	0	0	0	0	0	0	599
0.30	403	157	29	0	0	0	0	0	0	0	589
0.35	363	185	38	0	0	0	0	0	0	0	586
0.40	308	234	42	0	0	0	0	0	0	0	584
0.45	284	233	61	2	0	0	0	0	0	0	580
0.50	261	243	65	8	0	0	0	0	0	0	577

scheduled scenario. Table 13 shows that even with a 50% high/50% low traffic mix, a priority system would result in absolutely reliable service for the high priority class. Increasing the high priority traffic to 80% of the total, Table 14 shows that the priority system is still able to buffer against up to 20% uncertainty in demand. Table 15 shows that, in the tonnage-actuated scenario, the priority system is not effective at any level of uncertainty in improving reliability. This is because most variability is introduced by the shifting train schedules and connection opportunities, not by overflowing scheduled train capacity. Table 16 shows that the priority system is very effective in the power constrained scenario, so long as scheduled train departure times are maintained. Priority traffic is moved on a current basis everyday with only minor spreading, even at high uncertainty levels.

Tables 17, 18, and 19 show that in fixed train capacity scenarios, priority classification adversely affects the movement of nonpriority traffic, as expected. The degree of impact depends directly on the percentage mix of priority versus nonpriority traffic. Tables 17 and 18 report transit time distributions for nonpriority traffic classes under the tightly scheduled scenario. Table 17 reports results for a 50/50 traffic mix; Table 18 for an 80% High/ 20% Low mix. It can be readily seen that the

TABLE 15: SCHEDULE PRIORITIES IN TONNAGE RR SCENARIO

Transit Days Showing Number of Cars
for High Priority Traffic: 20/80 Mix
Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	0	480	150	0	0	0	0	0	0	0	630
0.05	5	631	8	0	0	0	0	0	0	0	644
0.10	50	333	264	0	0	0	0	0	0	0	647
0.15	84	453	124	0	0	0	0	0	0	0	661
0.20	62	520	92	0	0	0	0	0	0	0	674
0.25	161	426	102	0	0	0	0	0	0	0	689
0.30	128	485	74	0	0	0	0	0	0	0	687
0.35	81	552	69	2	0	0	0	0	0	0	704
0.40	107	477	109	0	0	0	0	0	0	0	693
0.45	142	429	152	11	0	0	0	0	0	0	734
0.50	165	475	101	0	0	0	0	0	0	0	741

TABLE 16: SCHEDULE PRIORITIES IN POWER CONSTRAINED SCENARIO

Transit Days Showing Number of Cars
for High Priority Traffic: 50/50 Mix
Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	650	0	0	0	0	0	0	0	0	0	650
0.05	639	0	0	0	0	0	0	0	0	0	639
0.10	620	0	0	0	0	0	0	0	0	0	620
0.15	613	0	0	0	0	0	0	0	0	0	613
0.20	604	0	0	0	0	0	0	0	0	0	604
0.25	589	10	0	0	0	0	0	0	0	0	599
0.30	564	25	0	0	0	0	0	0	0	0	589
0.35	532	54	0	0	0	0	0	0	0	0	586
0.40	501	83	0	0	0	0	0	0	0	0	584
0.45	473	104	3	0	0	0	0	0	0	0	580
0.50	455	114	8	0	0	0	0	0	0	0	577

TABLE 17: SCHEDULE PRIORITIES IN TIGHT SCENARIO

Transit Days Showing Number of Cars
for Nonpriority Traffic: 50/50 Mix

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	645	0	0	0	0	0	0	0	0	0	645
0.05	48	219	356	0	0	0	0	0	0	0	623
0.10	16	121	348	123	0	0	0	0	0	0	608
0.15	16	84	215	203	79	0	0	0	0	0	597
0.20	14	44	81	183	165	91	5	0	0	0	583
0.25	13	36	49	104	146	155	64	12	0	0	579
0.30	13	25	44	104	68	136	106	57	9	0	562
0.35	14	11	36	94	49	74	125	99	50	7	559
0.40	14	8	25	54	59	42	99	79	108	61	549
0.45	12	11	23	51	36	47	46	88	66	159	539
0.50	10	12	23	48	10	62	43	67	63	194	532

TABLE 18: SCHEDULE PRIORITIES IN TIGHT SCENARIO

Transit Days Showing Number of Cars
for Nonpriority Traffic: 80/20 Mix

Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	645	0	0	0	0	0	0	0	0	0	645
0.05	18	245	297	63	0	0	0	0	0	0	623
0.10	6	58	114	170	135	135	32	1	0	0	597
0.15	7	7	60	90	76	76	89	42	41	33	585
0.20	9	5	48	65	43	43	106	55	57	110	567
0.25	7	3	20	45	37	37	54	72	52	224	549
0.30	1	4	18	39	30	30	31	73	49	262	536
0.35	0	5	11	40	21	21	23	28	63	314	533
0.40	0	5	10	34	19	19	23	17	53	333	518
0.45	0	5	10	4	28	28	27	10	12	387	500
0.50	0	5	10	2	11	11	17	0	12	407	488

TABLE 19: SCHEDULE PRIORITIES IN POWER CONSTRAINED SCENARIO
Transit Days Showing Number of Cars
for Nonpriority Traffic: 50/50 Mix
Scheduled Capacity = 100 Average Demand = 100 Capacity/Volume Ratio = 1.00

	DAYS										
VAR	1	2	3	4	5	6	7	8	9	10+	TOT
0.00	95	90	185	208	72	0	0	0	0	0	650
0.05	19	79	175	207	157	2	0	0	0	0	639
0.10	10	54	125	197	182	52	0	0	0	0	620
0.15	4	44	93	190	156	113	13	0	0	0	613
0.20	4	35	63	147	117	137	76	25	0	0	604
0.25	3	21	61	80	126	116	97	87	8	0	599
0.30	1	13	77	51	96	99	100	88	63	1	589
0.35	2	8	85	40	41	120	40	85	109	50	580
0.40	2	7	74	40	13	102	55	82	70	133	578
0.45	1	5	67	42	14	51	52	49	33	248	562
0.50	0	5	64	39	3	43	75	34	32	263	558

effect on nonpriority traffic is more pronounced the higher the proportion of priority traffic. These tables can be compared with Table 1 to see how the traffic would have moved had there been no priorities in effect. Table 19 reports how the nonpriority traffic class would move under the power constrained scenario with a 50% high/50% low traffic mix. The primary effect seems to be to increase the average transit times, since the transit time spread was already quite wide in the unprioritized case. A practical way to limit the effect of the priority system on nonpriority traffic might be to arbitrarily limit the size of the priority class to say, no more than 15% of the total traffic.

OPERATING STRATEGY FOR SERVICE RELIABILITY

These model results clearly show that even with precisely executed train-operations, reliable origin to destination service is not guaranteed. "Precision Execution" of train operations is a necessary, but not a sufficient condition for producing reliable service. Additionally, either sufficient slack capacity must be scheduled to "buffer" uncertainty in demand, or a priority traffic system must be implemented to ensure that service-critical shipments always have first priority on available train capacity. Since providing slack capacity is expensive, the railroad industry has a financial incentive to explore ways to implement a priority traffic system.

The flexibility should be retained to operate extra trains, second or advance sections as an "overlay" on top of a basic scheduled network. This is needed to clear out yard queues in response to traffic surges: it both improves service reliability and helps keep terminals "fluid." Railroads should retain the flexibility to operate an extra train if there are sufficient nonpriority cars to justify it.

Really, the objective for nonpriority traffic is to eliminate the excessively long trips from the transit time distribution. From the tonnage railroad simulation, we learned that a demand responsive system is capable of accomplishing this. The longest trips result from the formation of persistent queues in yards; running an occasional extra train or second section would clear out these queues. Alternatively, nonpriority cars could be dynamically rerouted into different blocks or trains if their primary choice is filled. This provides a means to keep these cars moving towards their destination by filling space on existing trains, reducing the need for extras. Finally, nonpriority cars could be "aged" and their priority increased as their target delivery time approaches. Through some combination of these three strategies railroads can not only control service reliability for a priority class of traffic, but in fact ultimately for *all* traffic.

SUMMARY AND CONCLUSIONS

In spite of inherent environmental and cost advantages, the rail freight mode has steadily lost market share in recent years. In part, this loss of share is due to unreliability of rail transit times, making it difficult for rail freight to compete profitably with truck for high revenue, high value traffic. Until rail can better compete with trucks on a service basis, its cost and environmental advantages will remain irrelevant.

This simulation modeling effort establishes a clear relationship between the level of variability in demand and the transit time reliability produced by a rail transportation system. It establishes that "precision execution" is a necessary, but not sufficient condition for achieving service reliability. It is also necessary to schedule sufficient capacity to handle demand peaking, or else to implement a traffic priority system. Since providing slack capacity to handle traffic peaks is expensive, carriers have a financial incentive to figure out how to make a priority system work. The payoff will be the ability to run full trains with the right cars, on time, every day. By breaking the paradigm that reliable service always costs more, a traffic priority system may be the key to solving the railroad industry's long-standing service reliability problem, while retaining the inherent cost and operating efficiencies of the rail mode.

ENDNOTES

* University of Pennsylvania

1. This paper is a condensed version of Kraft, E. R., *Demand Variability, Service Reliability and Railway Freight Research Needs*, Working Paper 94-04-03, Department of Operations and Information Management, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6366.

2. *Advanced Logistics and Road Freight Transport*, Organisation for Economic Cooperation and Development (OECD), 1991, p. 100.

3. Allen, W. Bruce and Dong Liu, "Service Quality and Motor Carrier Costs: An Empirical Analysis," *Proceedings, Thirty-fifth Annual Meeting, New York City, Transportation Research Forum*, 1993, p. 25.

4. Winston, Clifford, "A Disaggregate Model of the Demand for Intercity Freight Transportation," *Econometrica*, 49/4, 1981, pp. 981-1006.

5. McGinnis, Michael, "The Relative Importance of Cost and Service in Freight Transportation Choice: Before and After Deregulation," *Transportation Journal*, 30/1, 1990, pp. 12-19.

6. Powell, W. B., "An Operational Planning Model for the Dynamic Vehicle Allocation Problem with Uncertain Demands," *Transportation Research*, Vol. 21B, No. 3, 1987, pp. 217-232.

7. Jordan, William C. and Mark A. Turnquist, "A Stochastic, Dynamic Network Model for Railroad Car Distribution," *Transportation Science*, Vol. 17, No. 2, May 1983, pp. 123-145.

8. Nozick, Linda K., *A Model of Intermodal Rail-Truck Service for Operations Management, Investment Planning and Costing*, Ph.D. Dissertation, the Department of Systems, University of Pennsylvania, 1992.

9. Jovanovic, Dejon and Patrick J. Harker, "A Decision Support System for Train Dispatching: An Optimization-Based Methodology," *Journal of the Transportation Research Forum*, Vol. 30, No. 1, 1990.

10. McCarren, J. Reilly and Carl D. Martland, *The MIT Service Planning Model*, Studies in Railroad Operations and Economics, Vol. 31, MIT, 1980.

11. Kwon, O. K., "Car Flow Management for Service Differentiation in Rail Freight Transportation Systems," Presentation at the ORSA/TIMS Joint National Meeting, Phoenix, AZ, November, 1993.

12. Dejax, Pierre and James H. Bookbinder, "Goods Transportation by the French National Railway (SNCF): The Measurement and Marketing of Reliability," *Transportation Research*, Vol. 25A, No. 4, 1991, pp. 219-225.

13. "What's Needed to Keep Intermodal Growing," *Railway Age*, Vol. 194, No. 10, October 1993, p. 60.