CURRENT METHODS FOR OPTIMIZING RAIL MARSHALLING YARD OPERATIONS

by

DONALD H. TIMIAN

B.A. in Chemistry, University of Arizona, 1979

A REPORT

Submitted in partial fulfillment of the

Requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas
1994

Approved by:

R. Michael Harnett

Dr. R. Michael Harnett

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	13 Oct 94	Final 1981 - 1993

4. TITLE AND SUBTITLE
Current Methods for Optimizing Rail Marshalling Yard Operations

5. FUNDING NUMBERS

6. AUTHOR(S)
Donald H. Timian

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of Industrial Engineering
Kansas State University
Manhattan, KS

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited

12b. DISTRIBUTION CODE
—

13. ABSTRACT (Maximum 200 words)
See Attached

14. SUBJECT TERMS	15. NUMBER OF PAGES	16. PRICE CODE
Rail, Marshalling Yard, Optimization	48	

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

ABSTRACT

America's railroads handle 37.5% of all freight shipped within the continental United States. Because of the competition from other modes of transportation, the rail industry is eager to improve its operational effectiveness. Smooth and efficient rail yard operations can improve delivery date reliability, reduce the time it takes a rail car to travel from its origin to its destination, and decrease the amount of time it takes to sort incoming cars or assemble outbound trains. This paper will focus on three different models - Shi's Hump Sequencing System, Kraft's Mixed-Integer Optimization Model, and Ferguson's Switching Process Model - designed to help optimize freight rail marshalling or classification yards.

ABSTRACT

America's railroads handle 37.5% of all freight shipped within the continental United States. Because of the competition from other modes of transportation, the rail industry is eager to improve its operational effectiveness. Smooth and efficient rail yard operations can improve delivery date reliability, reduce the time it takes a rail car to travel from its origin to its destination, and decrease the amount of time it takes to sort incoming cars or assemble outbound trains. This paper will focus on three different models - Shi's Hump Sequencing System, Kraft's Mixed-Integer Optimization Model, and Ferguson's Switching Process Model - designed to help optimize freight rail marshalling or classification yards.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

pg no

1. FIGURES

2. TABLES

1. INTRODUCTION

A. Overview

America's railroads handle 37.5% of all freight shipped within the continental United States (3). Because of the competition from other modes of transportation, the rail industry is eager to improve its operational effectiveness. Smooth and efficient rail yard operations can improve delivery date reliability, reduce the time it takes a rail car to travel from its origin to its destination, and decrease the amount of time it takes to sort incoming cars or assemble outbound trains. This paper will focus on three different models - Shi's Hump Sequencing System (30), Kraft's Mixed-Integer Optimization Model (21), and Ferguson's Switching Process Model (12) - designed to help optimize freight rail marshalling or classification yards.

B. Rail Marshalling Yard Operations

As described by Assad (2), "one may regard the rail transportation system as a network." In the case of freight rail, the nodes of the network can be thought of as the marshalling yards and the network's links that connect specific rail yards can be thought of as lines of track. For example, in Figure 1, a rail car that is starting in Barstow which is loaded with cargo for a consignee in Richmond would first have to pass through the Kansas City intermediate marshalling yard where it would be inspected, sorted, and then assembled together with other cars to await the departure of an outbound train to Richmond. These four operations - inspection, verifying the make-up of the train, sorting, and assembly - are called

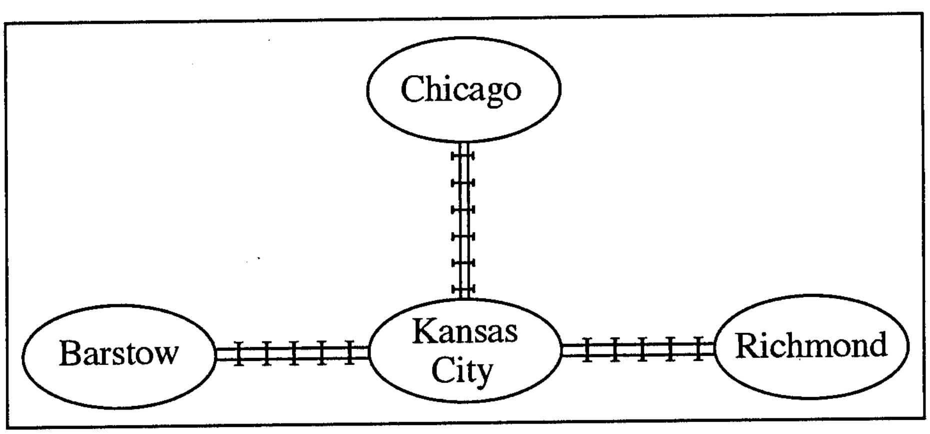


Figure 1 - Rail Network

yard activities; see Figure 2. The main activity being the sorting or classification of the incoming cars into groups or *blocks* according to their *forwarding point* or destination. It is this consolidation of cars into blocks that allows railroads to take advantage of the economies associated with full trainloads (2).

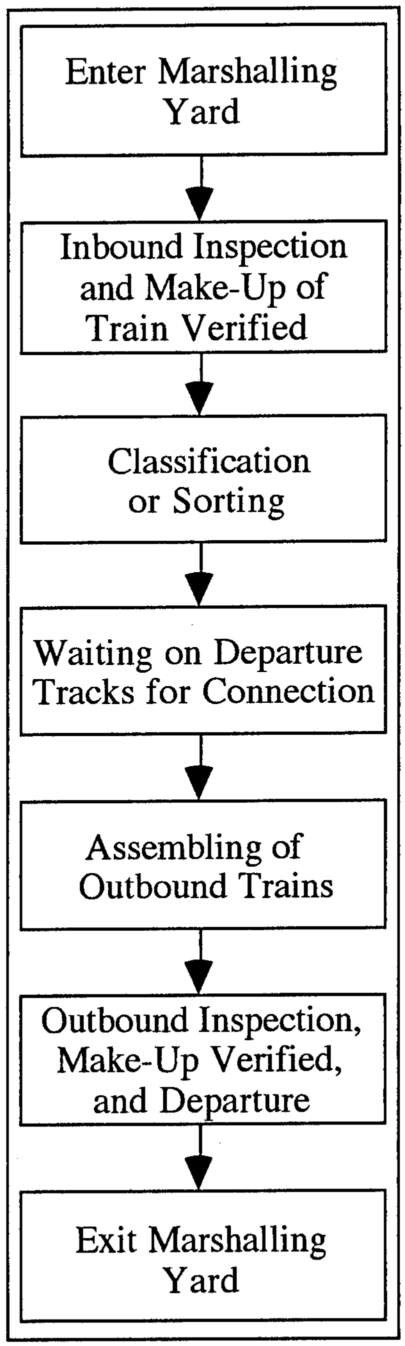


Figure 2 - Processing Steps for a Rail Car in a Marshalling Yard

Once sorted, blocked cars are then placed on classification tracks to await the departure of an outbound train. This process of sorting cars into blocks at intermediate marshalling yards is accomplished in one of two ways:

(i) In *hump yards* complete trains are pushed slowly up a raised portion of track - called a "hump" - where, at the crest, individual cars are uncoupled and allowed to coast to the desired classification tracks.

(ii) In *flat yards* switching engines move individual cars from receiving tracks onto classification tracks.

Because of their efficiency, most modern marshalling yards are hump yards (2); see Figure 3. Switching and braking of the cars - as they roll off of the hump - is controlled automatically. Any improperly sorted or *bad order* cars are collected by switching engines and resorted.

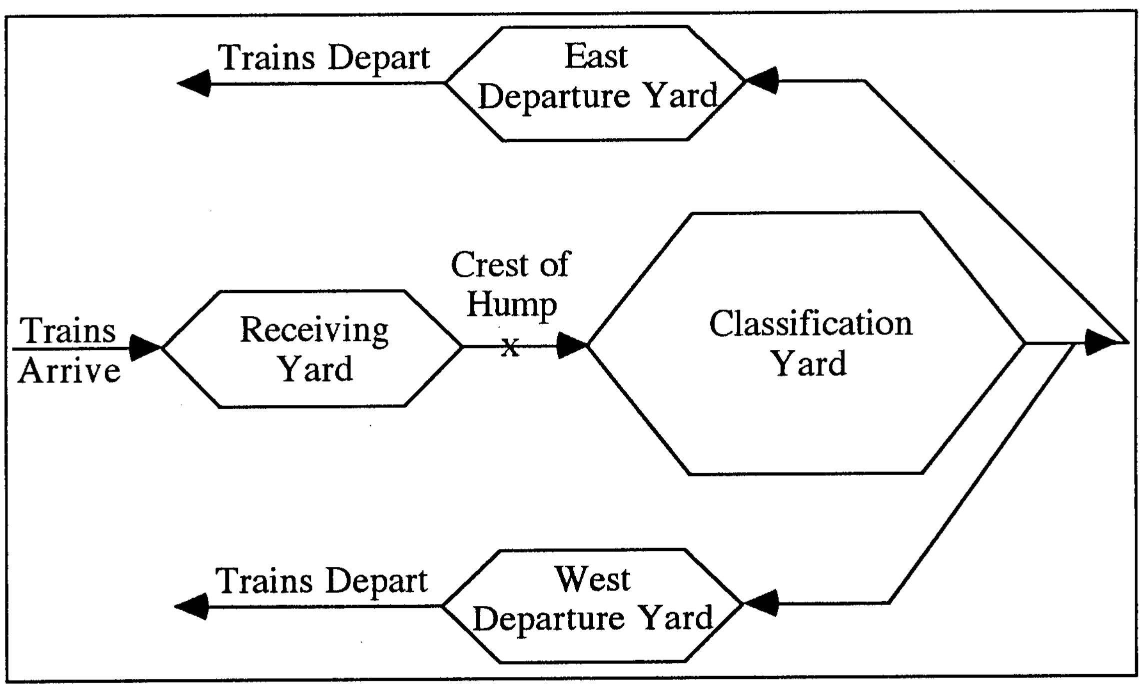


Figure 3 - Operations of a Hump Yard with Tree Stages

If sufficient tracks are not available to allow a yard master to sort inbound cars simply by destination or block, he may elect to use a multistage or dynamic sorting strategy. While there is no limit to the number of possible multistage sorting strategies, just four strategies are commonly used (11):

(i) Sort-by-train (iii) Triangular

(ii) Sort-by-block (iv) Geometric

The simplest multistage sorting strategy is sort-by-train (11). All inbound cars are first sorted onto classification tracks according to their outbound train. Next, cars "belonging" to a given outbound train are resorted according to their block or final destination, connected together, and assembled for departure. Outbound trains then depart sequentially or when desired.

Sort-by-block is the most prevalent sorting strategy in the United States (12). When sorting-by-block is used, inbound cars are first sorted according to their *block number*. All cars belonging to the first block - the block closest to the engine - will be sorted onto one track, the second block of any train onto another, and so on. In the second stage, cars are pulled onto the departure tracks and resorted according to train number. All cars on track one - belonging to block one - are pulled and resorted on the departure tracks first. Then all cars on track two - belonging to block two - are pulled and resorted. This process is continued until all of the classification tracks are cleared. Thus, all trains are assembled simultaneously.

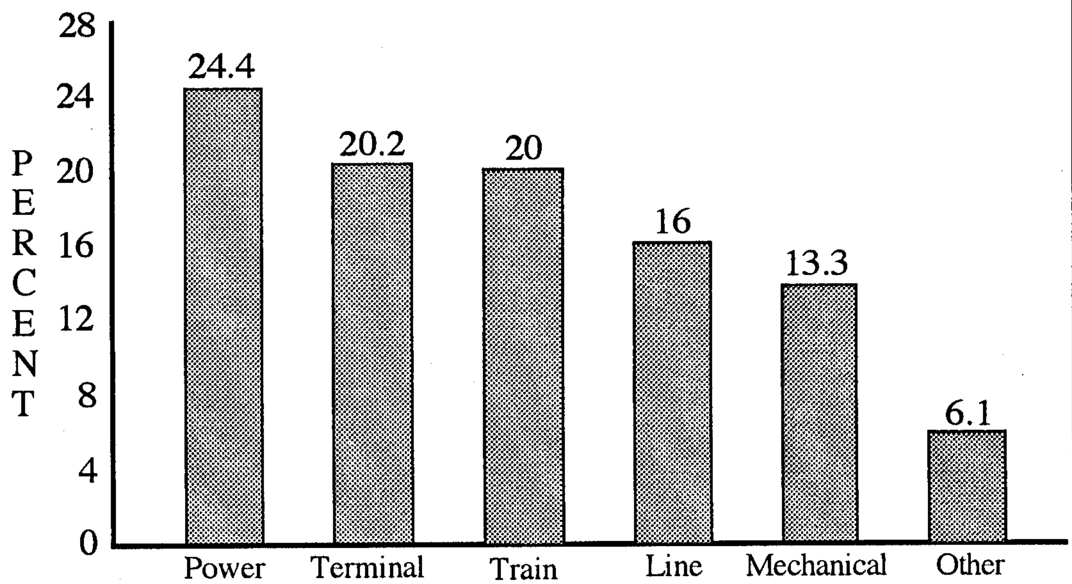
Triangular sorting is based on a numbering system which assigns an integer label to <u>each</u> block of <u>each</u> outbound train (11). Geometric or matrix sorting is the most powerful sorting technique, but due to its complexity it is seldom used (21). In geometric sorting, inbound cars are first sorted by *convoy*; a convoy - or a train convoy - being defined as those inbound cars that make-up a pair of consecutively outbound trains (11). Then, in the second stage, convoys are resorted by outbound train (11). In triangular sorting, a car may be resorted at most twice. In geometric sorting, a car may need to be resorted more than twice (11).

In general, rail cars can arrive at marshalling yards in one of two ways. They can arrive at a marshalling yard as part of inbound train or they can be delivered from a shipper's facility by a local freight train or an industry switch crew (2). These crews will often make regular rounds of the local industries that surround a given marshalling yard *spotting* empty cars and *pulling* loaded cars (2).

Over the past decade, most North American rail carriers have implemented - or are in the process of implementing - a rail car scheduling system (21). A *trip plan* is built for each car designating which trains should carry it for each leg of its journey. Once built, a trip plan becomes both the basis for customer committed delivery time and a way of measuring performance. Thus, a low priority car may not be immediately sorted in order to depart on the first available outbound train, while a later arriving, higher priority car may need to be sorted as soon as possible in order to depart on the next available outbound train.

C. Why Optimize Rail Marshalling Yards Operations?

In their 1993 study on the causes of unreliable rail service, Little and Martland (23) state, "A number of recent studies have highlighted the importance of service reliability to shippers in deciding which mode and carrier to use for the movement of freight." In 1990, the Association of American Railroads undertook a series of audits that looked at the present level of service. It was found that for boxcar traffic less than 80% of the cars arrived at their final destination within a two day window and that for one railroad terminal delays accounted for 20.2% of all the boxcars that failed to meet the agreed customer delivery date (23); see Figure 4.



Power - Power or Engine Shortage
Terminal - Terminal Delays, Yard Congestion, Switching Errors, Late Train Make-Up
Train - Maximum Tonnage or Length, Crew Shortage, Lack of Traffic
Line - Train Meet, Track Work
Mechanical - Bad Order, Engine Failure, Not Inspected
Other - Derailments, Unknown, Weather, Holidays

Figure 4 - Causes of Shipment Delay (23)

On average, a rail car will spend 12.4 days of it's 18.2 day car-cycle waiting in various *downstream* marshalling yards (26); see Figure 5. Some of this delay may be because of a failed inspection and the maintenance needed to correct the deficiency. Alternatively, a part of this delay may be because of the late arrival of one or more outbound trains. None-the-less, given a conservative car-day cost of $12.86 (26) and a fleet of 40,000 freight rail cars, a decrease in the car-cycle of only 6 hours can result in a potential annual saving of over two and one-half million dollars. Additionally, reducing the number of personnel and/or switching engines need to sort incoming trains can also result in significant savings.

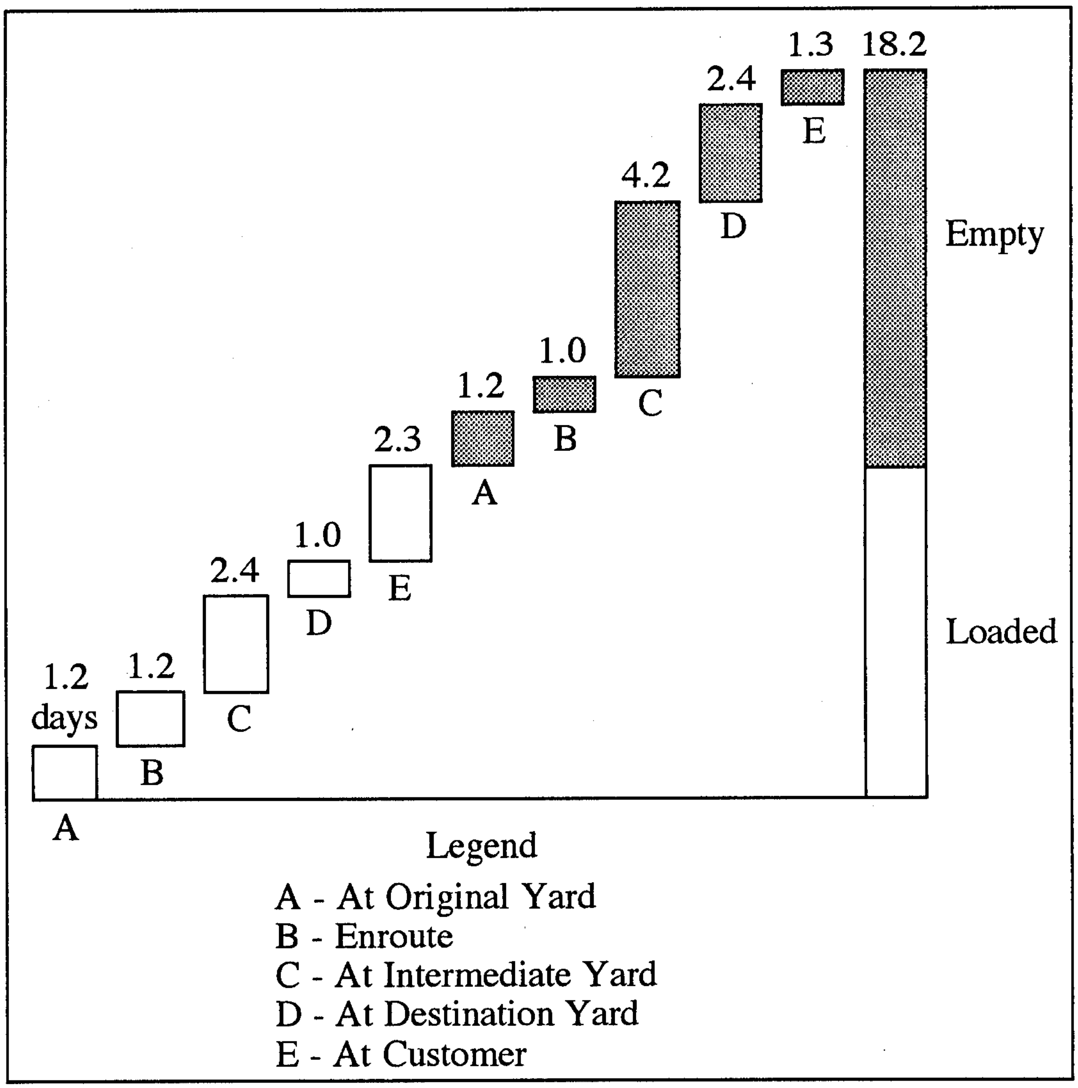


Figure 5 - Average Car-Cycle (26)

D. Railroads - The Choice for the Future

Given that America's economy is currently built on the assumption of "cheap oil", it is important to remember that railroads are both more fuel efficient and more environmentally friendly than trucks or automobiles (4, 32); see Figure 6. According to the Association of American Railroads (4), "On a single gallon of fuel, railroads can move a ton of freight three times as far as a truck can move that same ton." Thus if America were again to see an increase in the of price of oil similar to what it saw in the 1970s, for the "long haul" shipper, rail would quickly become the transportation mode of choice.

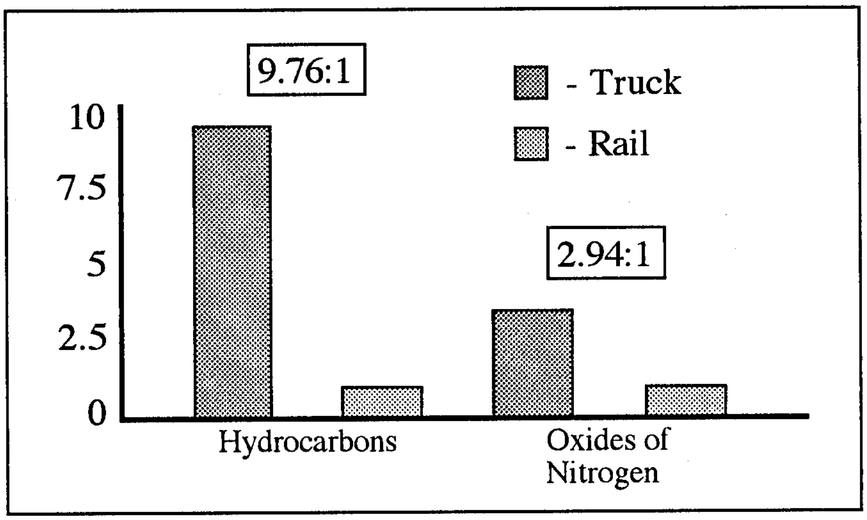


Figure 6 - Ratio of Rail to Truck Emissions (per billion ton miles) (4)

E. Strategies for Optimizing Rail Operations

Keaton (19) states, in his 1991 article on service-cost tradeoffs, "Strategies for improving rail service can be grouped into three categories." One strategy is technological improvements such as the pacing of more than one train over a given set of track in order to permit each train to travel at less than maximum speed in order to minimize fuel consumption and still meet planned arrival and departure times. Another strategy is to provide more direct and frequent "non-stop" train connections; thus, decreasing the amount of time it takes for a rail car to travel from its origin to its destination while also increasing the number of cars bypassing intermediate marshalling yards or terminals. A third strategy is to improve the decision support systems at rail marshalling yards. This third strategy - which is the underlying theme of this literature survey - can be further subdivided into techniques to optimize terminal operations and techniques to optimize blocking plans. In chapter three a model that attempts to optimize both terminal operations and blocking plans will be described. In chapters two and four models designed to optimize terminal operations - specifically hump sequencing - will be described.

2. THE HUMP SEQUENCING SYSTEM

A. Hump Sequencing Policies

In order to determine the optimum "humping" sequence for rail cars, two general policies or approaches have been suggested (38):

(i) Maximizing (minimizing) the number of likely connections (missed connections) between inbound and outbound rail cars.

(ii) Minimizing - from the time of its arrival to the time of its departure - the average length of time (or cost) that a rail car spends in a marshalling yard.

This first approach was used by both Deloitte, Haskins & Sells (21) in 1977 and by Allen and Rennicke (38) in 1978 to develop, respectively, the Terminal Hump Sequencing System and the Terminal Sequencing System. While open literature describing the Terminal Hump Sequencing System is not readily available, it appears that Deloitte, Haskins & Sells (21) did include in their system a set of constraints that take into account the capacities of the receiving, classification, and departure tracks as well as the amount of time it takes to block, assemble, and inspect an outbound train. However, neither model allows outbound trains to be built to tonnage - or to their *max-min* car requirements - rather than to time. (An outbound train's *max-min* car requirement is the maximum number of cars that it may depart with or the minimum number of cars that it must have in order to depart.) This assumption of a fixed departure time for each outbound train may introduce a "Catch 22" into the humping sequence. If for no other reason than simply the large number of "moving pieces" (i.e., track outages, the availability of *power* or road engines, the possible ripple effect - with regards to the departure of an outbound train or the sorting an inbound train - caused by a *down* rail car, etc.), freight trains very seldom run on time.

In 1981 Shi used this second strategy - minimizing the average length of time (or cost) that a car spends in rail yard - to develop the Hump Sequencing System (HSS) (30, 38). Shi's system takes into account the capacities of the receiving, classification, and departure tracks; the time it takes to block, assemble, and inspect an outbound train; and allows outbound trains to be built to max-min car requirements rather than time requirements (30, 38).

B. Shi's Model

As described by Shi (30), "HSS is a computer-aided simulation model which attempts to produce the most desirable [or 'best'] humping sequence in terms of average rail yard throughput costs". It does this by working backwards through the rail marshalling yard (starting with the departure and assembly of outbound trains then the humping, inspection, and arrival of inbound trains); see Figure 7 (30, 38). HSS does this by sequencing through three principle procedures or phases (38):

(i) *Evaluating and modifying yard operating status/requirements*

To date, most North American rail carriers have implemented a rail car scheduling or inventory system (19). In HSS, Shi has elected to use the Canadian National Railway's Yard Inventory System (YIS). Initially all of the factors that are used in HSS are stored in one of the YIS's two databases; permanent and temporary. The permanent database contains information relating to the design of the rail yard and its current status (i.e., the maximum and current capacity of each receiving track, the max-min capacity of each outbound train, the current departure schedule, etc.). The temporary database contains information that can be keyed-in or updated by the General Yard Master prior to each run in order to reflect the yard's current operating situation (i.e., actual arrival time of inbound trains, car priority values, current clock time, etc.). While each rail company has different car priority policies, normally a car's priority value is expressed in $/car-hour and is usually determined according to one or more of the following conditions:

- The type of merchandise loaded in the car.
- The car's consignor; some shippers receive preferential treatment with regards to rate and/or guaranteed arrival date.
- Whether or not the car has been designated "special" (e.g., a refrigerator car, a tank car, or a double-stack car).
- Whether or not the car is "foreign"; i.e., the rail car is owned by another railway company.

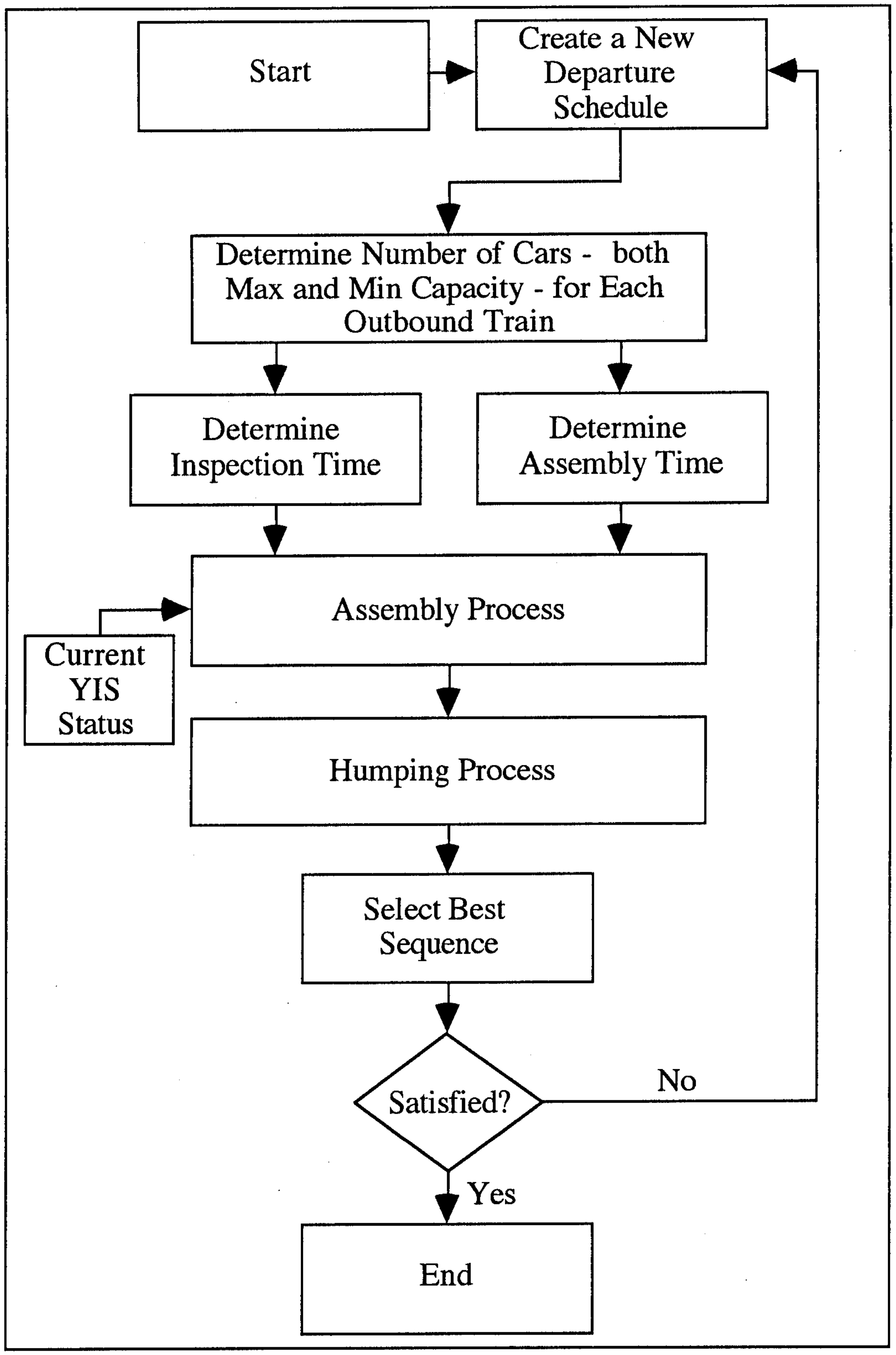


Figure 7 - Flow Chart of HSS Model (30, 38)

(ii) *Determining classification track pull schedule*

The next phase involves the estimation of the Latest Assembly Time (LAT) for each outbound train. LAT is the time beyond which the departure of outbound train k would have to be delayed, if its assembly has not yet been completed.

$$LAT_k = TD_k - TI_k - TA_k \qquad [1]$$

Where TD_k is the current scheduled departure time of outbound train k, TI_k is the inspection time of train k, and TA_k is the assembly time for train k. Current scheduled departure time for each train is expressed in terms of clock time. Inspection, as well as assembly time, is assumed to be proportional to the number of cars in outbound train k.

$$TI_k = \alpha_I N_k \qquad [2]$$

$$TA_k = \alpha_A N_k \qquad [3]$$

Where N_k is the number of cars on outbound train k, α_I refers to the average inspection time for each car on train k, and α_A refers to the average assembly time for each car on train k.

The difference between the current clock time and the LAT establishes the *block time* for each classification track. Within the block time, a classification track is able to receive rail cars from the humping process without causing delays to the outbound train. Block time sets the basis for the humping sequence.

(iii) *Determination of the "best" hump sequence*

In order to determine the "best" hump sequence, HSS must first review the YIS's permanent database. Only those inbound trains that have been both inspected and had their cars tagged with the appropriate destination code are considered for humping. As with the assembly and inspection times for outbound trains, in HSS the hump time of an inbound train is assumed to be proportional to the number of cars it contains; HSS makes no provisions for *no hump* cars.

A comparison is made between the required humping time of each inbound or candidate train and the block time of each classification track for which a candidate train has tagged cars. If some of the cars belonging to a candidate train cannot be humped within the block time or if the maximum capacity of a given classification track has been met, these cars are allocated to a *gauge* or side track where they are delayed until the next humping phase. If, on the other hand, the number of cars on a specific classification track is greater than an outbound train's maximum capacity, the surplus cars are delayed for the next scheduled departure of an outbound train headed for the same forwarding point or destination.

C. Screening of Candidate Trains

The key to Shi's system, is the use of dynamic programming (30, 38). While both Deloitte, Haskins & Sells (21) and Allen and Rennicke (38) elected to use linear programming to determine their optimum humping sequence, Shi elected to use dynamic programming to compare all of the possible inbound train humping sequences in order to find the "best" or optimum (21, 30, 38). Therefore, in order to calculate the optimum humping sequence for four candidate trains, Shi's system must calculate twenty-four (i.e., 4!) different time (cost) permutations; see Figure 8.

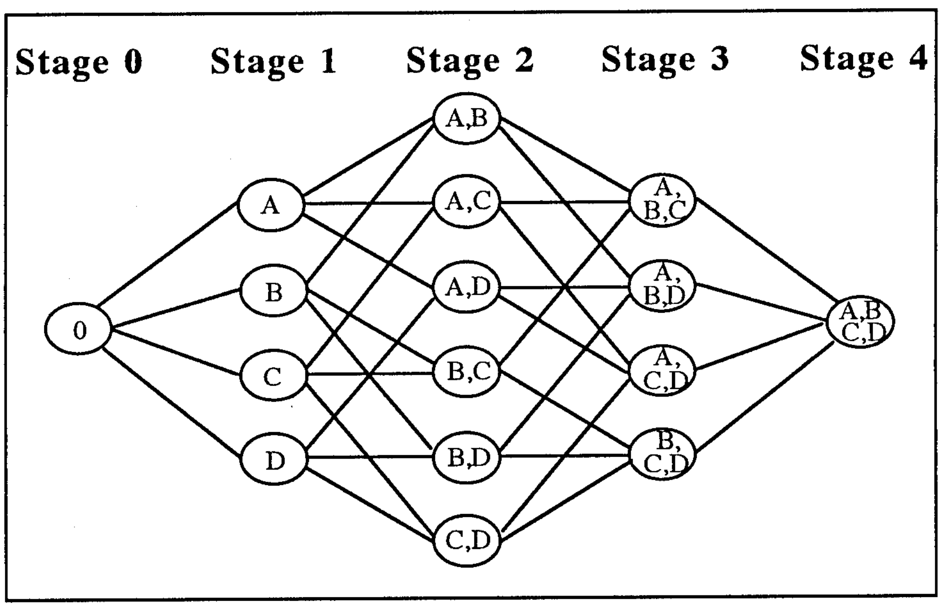


Figure 8 - Time (Cost) Permutations for Humping Sequence using Dynamic Programming (30, 38)

But, with each increase in the number of candidate trains available to hump a rapid increase in the amount of computer memory and computing time occurs; the problem or "curse" of dimemsionality. For example, twenty candidate trains would involve over 2.4 x 10^{18} comparisons. In 1983 Shi estimated at such an exercise would take almost an entire day (38). Even with the significant increase in computing speeds and storage over the last decade, such an exercise is simply not practical. It is for this reason that Shi developed a heuristic which is designed to reduce the number of candidate trains available for humping (30, 38). This screening procedure involves the calculation of load and priority factors together with decision values for each candidate train with respect to the given destinations or forwarding points of outbound trains (30, 38). Those candidate trains with the highest decision values are then selected to be input into HSS (30, 38). The following algorithm describes this procedure.

Step 1 - Calculate the load (M_k^j) and priority (P_k^j) factors for each candidate (or inbound) train with respect to each appropriate outbound train.

Step 2 - Calculate the decision value for each candidate train.

Step 3 - Rank the candidate trains - from high to low - in order of their decision values.

The load factor (M_k^j) reflects the proportion of cars in candidate train j that are likely to find connections on outbound train k (30, 38).

$$M_k^j = 10^{10} \left(\frac{\sum_i N_i^{jk}}{E^k} \right)^3 \cdot \frac{1}{N^k T_k^8} \qquad [4]$$

Where N_i^{jk} is the number of cars with tag number i in candidate train j bound for destination k; N^k is the train capacity of outbound train k; T_k is the time interval from current clock time to the scheduled departure time for outbound train k; and E^k is the critical number of cars that outbound train k must have in order to depart (30, 38). The critical number of cars (E^k) that outbound train k must have prior to being assembled is nothing more than the train capacity of outbound train k (N^k) minus the number of cars with tag i on the classification track(s) designated for train k ($\sum n_i^k$) (30, 38).

$$E^k = N^k - \sum_i n_i^k \qquad [5]$$

The priority factor (P_k^j) reflects the profitability of candidate train j with respect to outbound train k provided that all of train j's cars are humped prior to train k's latest assembly time (LAT) (30, 38). A high priority factor for candidate train j with respect to outbound train k suggests that train j is important or profitable relative to the makeup of train k.

$$P_k^j = \sum_p \left(\frac{\left[\sum_i N_{ip}^{jk} \right]}{E^k} \right) I_p \qquad [6]$$

N_{ip}^{jk} number of cars with tag number i and priority value p on inbound train j with respect to outbound train k

I_p intrinsic car priority value p ($/car-hour)

The decision value (V^j) of each candidate train j is the sum of the product of each candidate train's load and priority factors with respect to outbound trains k_1, k_2, k_3, . . . (30, 38).

$$V^j = \sum_k P_k^j M_k^j \qquad [7]$$

Example (38):

Consider two candidate trains, j_1 and j_2, consisting of 100 cars each. There are two outbound trains k_1 and k_2 with departure times of 5:00 and 6:00 respectively. The current clock time is 0:00.

Both outbound trains are going to the same destination and each train has a maximum capacity of 100 cars. Currently, train k_1 has 100 cars on its classification track(s) and train k_2 has zero.

The intrinsic car priority value for all of train j_1's cars, with respect to both outbound trains, is 0.91 $/car-hour. The intrinsic car priority value for all of train j_2's cars, with respect to both outbound trains, is 0.45 $/car-hour.

Step 1 - Calculate the load ($M_k{}^j$) and priority ($P_k{}^j$) factors for each candidate (or inbound) train with respect to each outbound train.

$M_{k1}{}^{j1} = 10^{10}(0/0)^3/100x5^8 = 0$ $\quad$ $P_{k1}{}^{j1} = (100/0)0.91 = 0$
$M_{k1}{}^{j2} = 10^{10}(0/0)^3/100x6^8 = 0$ $\quad$ $P_{k1}{}^{j2} = (100/0)0.45 = 0$
$M_{k2}{}^{j1} = 10^{10}(100/100)^3/100x5^8 = 256$ $\quad$ $P_{k2}{}^{j1} = (100/100)0.91 = 0.91$
$M_{k2}{}^{j2} = 10^{10}(100/100)^3/100x6^8 = 59.5$ $\quad$ $P_{k2}{}^{j2} = (100/100)0.45 = 0.45$

Step 2 - Calculate the decision value for each candidate train.

$V^{j1} = P_{k1}{}^{j1}M_{k1}{}^{j1} + P_{k2}{}^{j1}M_{k2}{}^{j1} = 0 + (0.91)(256) = 233$
$V^{j2} = P_{k1}{}^{j2}M_{k1}{}^{j2} + P_{k2}{}^{j2}M_{k2}{}^{j2} = 0 + (0.45)(59.5) = 26.8$

Step 3 - Rank the candidate trains - from high to low - in order of their decision values.

$V^{j1} = 233 > V^{j2} = 26.8$ $\therefore$ candidate train j_1 should receive priority for humping.

D. Cost

Figure 8 (introduced earlier) illustrates the time (cost) permutations for humping inbound trains A, B, C, and D using dynamic programming (30, 38). Within the figure, the four steps or *stages* represent the completion of humping one, two, three, or all four trains; the nodes represent the various hump sequencing *states* that are possible; and the arcs represent the available number of choices which can be made to *transition* from one stage to the next. Because for any transition between two stages a *stage return* must be defined, Shi in HSS defines his stage return "as the cost associated with each humping of an inbound train" (30, 38).

Therefore, the key component in Shi's HSS is the cost ($C_j{}^t$) of transitioning from one stage to the next or of humping train j at state t given that trains ($Q_i{}^{t-1}, Q_m{}^{t-2}, \ldots Q_p{}^1$) have been previously humped (30, 38).

$$C_j^t(Q_i^{t-1}, Q_m^{t-2}, \ldots Q_p^1) = C_1 + C_2 + C_3 \qquad [8]$$

Where C_1 is the rehumping and delay costs associated with cars that must be humped onto a gauge track because of a time overlap between the humping time of inbound train j and the latest assembly time (LAT) of outbound train k; C_2 is the rehumping and delay costs for cars on inbound train j that have been humped onto a *gauge* or side track because of a lack of room on outbound train k's classification track; and C_3 is the cost assessed on cars which are humped within a suitable sequencing period or are delayed pending certain conditions be met (30, 38). The expressions which represent the cost components C_1 and C_2 are of the form (30, 38):

$$C_1 = \alpha_1 \sum_p \sum_i N_{ip}^{j(t)} I_p x_i \qquad [9]$$

$$C_2 = \alpha_2 \sum_i I_{avg\,p}^{j(t)} \left\{ N_i^{j(t)} \left[Y_i + Z_i \right] - \left[C_i - n_i - \sum_{t=1}^{n} N_i^{j(t)} \right] Y_i \right\} \qquad [10]$$

α_1	average unit cost for rehumping and car delay
$N_{ip}^{j(t)}$	number of cars with tag number i and priority value p on inbound train j at state t
I_p	priority value p ($/car-hour)
X_i	binary variable; 1 if overlay exists, 0 otherwise
α_2	average unit cost for rehumping and car delay
$I_{avg\,p}^{j(t)}$	average priority value of the cars on inbound train j at state t
$N_i^{j(t)}$	number of cars with tag number i on inbound train j at state t
Y_i	binary variable; 1 if room exists, 0 otherwise

Z_i binary variable; 1 in no room exits or if no classification tracks have been assigned for cars with tag number i, 0 otherwise

C_i the capacity of the classification tracks for cars with tag number i

n_i the current number of cars, before humping of any inbound train, on classification tracks which contain cars with tag number i

With regards to the expression that represents C_3 (Equation 11), it is important to remember that C_3 in fact represents three different cases where processing costs may occur (30, 38):

• *Case 1* - Inbound train j at stage t contains at least the critical number (E^k) of cars for outbound train k and the humping plus assembly time is sufficient to allow outbound train k to depart its first departure time (t_1^k). Any surplus cars in inbound train j(t), tagged for outbound train k, are delayed until outbound train k's second departure time (t_2^k).

• *Case 2* - Inbound train j at stage t contains at least the critical number of cars for outbound train k, but humping and assembly time is not sufficient to allow outbound train k to depart at the first departure time. The critical number of inbound train j's cars for outbound train k are delayed to the second departure time. Surplus cars in inbound train j, in excess of the critical value for outbound train k, are delayed for 180 minutes, after the second departure time for outbound train k. (It appears that Shi's 180 minute value is based on the three hour time horizon that was used during the evaluation of HSS at the Taschereau Rail Yard.)

• *Case 3* - Outbound train k has zero number of critical cars; outbound train k has on its classification track the maximum number of cars it is allowed to depart with. The cars on the classification track allocated to outbound train k depart at the first departure time. Any surplus cars in inbound train j(t) destined for train k are delayed at least to the second departure time for outbound train k.

$$C_3 = \sum_k \sum_i \{[E^k (t_1^k - t_c) I_{high\,p}^{j(t)} + (N_i^{j(t)k} - E^k) \cdot$$

$$I_{low\,p}^{j(t)} (t_2^k - t_c + 180\ W_i)](Q_i + W_i)\} \qquad [11]$$

$$+ \sum_k \sum_i [N_i^{j(t)k} (t_2^k - t_c + 180\ V_i) I_{avg\,p}^{j(t)} (R_i + u_i + V_i)]$$

Where t_c is current clock time; $I_{high\ p}$, $I_{low\ p}$, $I_{avg\ p}$ is the high, low, and average priority value for inbound train j(t)'s cars; W_i is 1 if $E^k > 0$, $N_i^{j(t)k} > 0$, and train k will not depart at t_1^k and 0 otherwise; Q_i is 1 if $E^k > 0$, $N_i^{j(t)k} > 0$, and sufficient time exists for train k to depart at t_1^k and 0 otherwise; V_i is 1 if $E^k > 0$, $N_i^{j(t)k} < E^k$, and outbound train k will not depart at t_1^k and 0 otherwise; R_i is 1 if $E^k > 0$, $N_i^{j(t)k} > 0$, and sufficient time exists for train k to depart at t_1^k and 0 otherwise; and u_k is 1 if $E^k = 0$ and 0 otherwise.

E. Hump Sequencing through Dynamic Programming

As described by Shi (30), when "determining the optimum sequence, it is necessary to calculate, for each node, the least cost of arriving at the place represented by that node (i.e., the least cost of humping a set of trains represented by that node). For one of the N nodes of stage 1, the cost is just the cost of humping one [inbound] train represented by that node in the first stage of the sequence [(i.e., $M_{k,1}$)]." The minimum cost of arriving at any other node in stage 2 though N, can be determined by the following recursive formula (38):

$$M_{k,t} = \min_{j \in J} [M_{j,t-1} + C(j; k, t)] \text{ for } k = 1, 2, \ldots N! \quad t = 2, 3, \ldots N \qquad [12]$$

Here J is the set of nodes at stage t-1 that are connected to node k at stage t; $M_{k,t}$ is the minimal cost of arriving at the k^{th} node of stage t; C (j; k, t) is the cost of humping one train required to move from node j of stage t-1 to node k of stage t; and N is the original number of inbound or candidate trains to be humped.

With only one node associated with stage N, the minimum cost of humping all N trains is read by HSS directly. Then, by means of a standard backward pass, HSS determines the most desirable or "best" humping sequence in terms of average rail yard throughput costs.

F. Taschereau Marshalling Yard

Using input the yard status at Taschereau in early October of 1980, Shi measured the benefit of the HSS Model by simulating an actual day's humping and a further 3 days of humping (30, 38). In comparisons with the humping sequence specified by Taschereau's General Yard Master (GYM) and the First-In-First-Out Criterion (FIFO), HSS was found to significantly improve on the existing operation by reducing the cost of "tied-up cars" (30, 38).

In an example of one day's humping of eight inbound and six outbound trains, Shi reports the total time for the FIFO criteria as 16.72 hours/car (11,103.58 hours/664 cars), the total time based on the General Yard Master's experience as 14.68 hours/car (9,833.58 hours/664 cars), and the total time as determined by HSS as 13.6 hours/car (8,937.83 hours/664 cars) (38). Assuming Canadian National Rail's 1981 cost of $0.65 per car-hour, HSS's potential savings is $1,407.64 versus FIFO and $582.23 versus GYM (38). If we use McKinsey & Company's 1992 cost of $0.54 car-hour (26), HHS's potential savings is reduced, respectively, to $1,169.42 and $483.70. If one conservatively assumes that only 15% of the HSS vs. GYM potential savings is realized, a savings of $72.56 per day multiplied over 365 days at 5 different rail yards that handle a similar number of cars, quickly turns into a savings of one hundred twenty-five thousand dollars per year.

3. A MIXED-INTEGER OPTIMIZATION MODEL

A. Background

In June of 1993 Kraft (12) presented a paper in which he described his ongoing work in developing a Mixed-Integer Linear Programming Optimization Model designed to improve both the effectiveness and efficiency of freight rail marshalling yards As described by Kraft (21), "Effectiveness being defined as sorting all cars into required blocks and trains within required time frames" and "Efficiency is accomplishing this at minimum cost." The "key word" being effectiveness, because the measure of a rail company's performance - as seen by the *consignor* or customer - is its ability to provide reliable service.

> "We don't just run a railroad. We provide transportation and distribution services . . . . We're in business to meet customer needs . . . . The right measure [of our effectiveness] is whether the goods reach the customer by the time promised."
>
> - Union Pacific Railroad Chairman Michael Walsh (12)

Instead of looking at just the humping problem or at various classification track assignment strategies, Kraft's model integrates three lines of previous research into one model: the Terminal Hump Sequencing System developed by Deloitte, Haskins & Sells, the Dynamic Classification Track Assignment Model developed by SRI International, and freight car scheduling systems recently implemented by most North American rail companies (21).

B. The Terminal Hump Sequencing System

Developed in the late seventies by Deloitte, Haskins & Sells, the Terminal Hump Sequencing System attempts to maximize the number of scheduled connections between inbound and outbound rail cars (21). While open literature describing the Terminal Hump Sequencing System is not readily available, it appears that Deloitte, Haskins & Sells did include in their system a set of constraints that take into account the capacities of the receiving, classification, and departure tracks as well as the amount of time it takes to block, assemble, and inspect an outbound train (21). However, because Deloitte, Haskins & Sells' model assumes that the outbound train schedule <u>must</u> be maintained, a bias may be introduced into the humping sequence where outbound trains are build to time instead of max-min requirements (21, 30, 38).

C. The Dynamic Classification Track Assignment Model

Developed in the late seventies and field tested in the early eighties at Southern Pacific's West Colton Marshalling Yard, SRI's Dynamic Classification Track Assignment Model uses a rule-based heuristic to optimize the sort-by-block multistage sorting strategy (21). In short, SRI's model takes the hump sequence "as given" and attempts to minimize both the number of classification tracks used and the amount of work the *makeup-engine* (the engine that *builds* or assembles an outbound train) must do by assigning blocks scheduled to depart on the same outbound train either to the same or an adjacent classification track (21).

D. Kraft's Goal

Kraft's Mixed-Integer Optimization Model's goal "is to formulate a single model which can optimize hump sequence and track assignments simultaneously" (21). If inbound trains are running late, Kraft's yard planning model can obtain an estimated time of arrival from the rail carrier's train dispatching or tracking system (21). Thus the model can, if necessary, determine which inbound cars will depart on which outbound trains; <u>but</u>, Kraft's Mixed-Integer Optimization Model "is designed to classify cars <u>based directly on their trip plans</u>" (21).

Because it is based on Deloitte, Haskins & Sells' work, Kraft's model requires the capacities of the receiving, classification, and departure tracks as well as the amount of time it takes to block, assemble, and inspect an outbound train (21). Additionally, data as to how many classified and unclassified cars are currently on hand, together with which receiving or classification tracks these cars are sitting on, is input via a link with the yard's own inventory control system (21).

Designed primarily for larger rail yards, where there are separate receiving, classification, and departure tracks, Kraft's model has two "degrees of freedom" (21):

(i) <u>Hump Sequence</u> - The ideal hump sequence is determined by two primary factors: first, the schedule of a train's outbound connections, and second, the current block/track assignments in effect. In the model, protecting scheduled connections is given high priority. "Cutoff times" - or Latest Assembly Times - are established and act as constraints . If, for a pair of inbound trains, cutoff time is not a factor, then the choice of which train should be humped first is based on the composition of each inbound train and the degree to which it matches the current or ongoing classification block assignments. For example, suppose that one train has primarily southbound cars, and a second train has primarily northbound cars. If northbound blocks are currently being made, the model will elect to process the northbound train first.

(ii) Classification Track to Block Assignments - Throughout the time period covered, assignment as to which blocks are sorted onto which classification tracks can be fixed or variable. If allowed to vary, the model will use a sort-by-train strategy to optimize the number of classifications - and therefore the number of connections made - while holding operating costs under control.

E. Objective Function

Just as one may envision a rail transportation system to be nothing more than a network, so too can one view a rail marshalling yard. The nodes representing either inbound trains, classification tracks, or outbound trains and the arcs representing the humping, switching, or reswitching needed to process an inbound car. In his model Kraft uses just such a network to portray a three dimensional rail marshalling yard; the third dimension being time (21).

For example, in Figure 9, an inbound car is processed over a span of four time periods with each event/arc numbered as follows (21):

#1 - An inbound car arrives on inbound train node #2 during the first time period.

#2 - The car is not processed during the first time period and is moved forward into the second time period.

#3 - During the second time period the inbound car is humped onto classification #1 and blocked with a set of cars bound for the same destination.

#4 - The block on classification track #1 not pulled during the second time period onto a departure track, so it is moved forward into the third time period.

#5 - The block on classification track #1 is reswitched to classification track #2.

#6 - The block on classification track #2 is again not pulled onto a departure track and is moved forward into the fourth time period.

#7 - During the fourth time period, the block on classification track #2 is pulled to the node representing outbound train #2 on departure track #3, assembled with the other blocks schedule also to depart on outbound train #2, and departs.

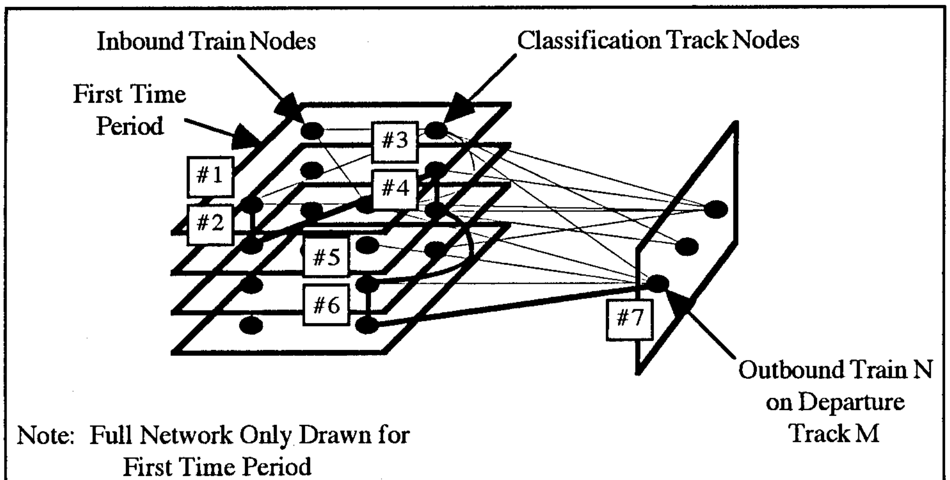


Figure 9 - Kraft's Network Model of a Marshalling Yard (21)

Kraft imbeds this marshalling yard network in his model and, via his objective function, strives to minimize the total number of times each car is rehandled as well as which departure tracks are used and the number of connections missed (21). The first two factors - number of cars reswitched and which *pullout leads* or departure tracks used - are measures of a yard's operational efficiency (21), while the last factor, number of missed connections, is a measure of a yard's effectiveness as perceived by the consignor (21). As stated by Kraft, "The relative weights given each factor depend on the level of congestion in the yard, the importance of making connections, and other management priorities" (21).

$$\text{Min } Z = \sum_{J,K,L,LP \neq L} \text{Reswit}\ (J, K, L, LP) * \text{Rswfact} \ +$$

$$\sum_{J,K,L} \text{Rswinv}\ (J, K, L) * \text{Rswfact} \ + \qquad [13]$$

$$\sum_{M,N} \text{Uselead}\ (M, N) * \text{Leadcost}\ (M, N) \ +$$

$$\sum_{K,L} \text{Leftover}\ (K, L) * \text{Infeas}$$

J Time Period
K Outbound Block
L Classification Track
M Pullout Lead
N Outbound Train
LP The Set of Classification Tracks

- The sums Reswit (the switching of a car from one classification track to another) and Rswinv (the reswitching of a car back to its "original" classification track during a reswitching event) are multiplied by the cost of handling a car (Rswfact).

- The binary variable Uselead indicates which leads or departure tracks are used and Leadcost keeps track of any preference that might exist to group blocks for a given outbound train on a specific departure track or in a general area of the departure yard.

- The total number of Leftover cars times the cost for each missed connection (Infeas).

F. Scale Up Difficulties

Kraft feels that in order for his Mixed-Integer Model to be of commercial use it must be able to handle 60 tracks, 90 blocks, and 25 trains simultaneously (21). As of the summer of

1993, this is considerably larger than any application he had yet attempted (21). Kraft states that as larger applications are created two problems will arise: excessive execution time (growing exponentially) and excessive memory/workspace requirements (21).

Results from sensitivity analysis indicates that this increase in both execution time and memory/workspace requirements may be due to the "searching" done by model to find the exact integer solution of the classification track to block assignment problem (21). Because Kraft's model is expected to be used in a commercial environment, a quick "B-" answer is better than an exact, "A+" provable optimum solution. Kraft suggests (21) "that a heuristic approach . . . or a non-exact technique . . . may prove more successful" in solving the track to block assignment problem.

4. THE SWITCHING PROCESS MODEL

A. Scheduling Theory and the Switching Problem

Because in a rail marshalling yard we are dealing with a specific lists of tasks that must be done on a group of inbound trains in order to meet a given departure schedule, one can characterize the humping or switching problem at a rail yard as a *general job-shop problem.* In general, a job-shop problem has been described as follows:

> "Given N jobs each having operations to be performed on each of M machines with the order of the machines for each job specified, determine the order of operations on each machine, or alternately, the schedule for each operation on each machine, so as to achieve some desired result." (12)

Where the *jobs* are the assembling of N outbound trains and the machines (M_1 and M_2) are respectively the hump locomotive and the make-up or *pull-back* engine (12).

In scheduling theory a job is defined as being "late" when its completion time (C) minus its due date (D) is a positive number. Thus lateness (L) is positive when the job is completed after its due date and negative when it is completed early.

Within the rail industry a job or switch is considered "late" if it is completed *after* its due date. Given that the rail industry penalizes for tardy occurrences, the goal should be minimizing the maximum lateness of all outbound trains. Because in a rail marshalling yard the order of the machines is set (i.e., hump inbound trains first, then assemble outbound trains), a better description of this scheduling or switching problem is a two machine flow shop; see Figure 10.

n/2/F/f(L)	
n - # of inbound trains	F - Flow Shop
2 - # of machines	f(L) - Minimize Maximum Lateness

Figure 10 - A Two Machine Flow Shop Problem

B. The Switch Processing Model

In his doctoral thesis, Ferguson (12) developed a mathematical model of the switching process at a large hump marshalling yard that he called the Switch Processing Model (12). Through the use of a series of six assumptions (listed below), Ferguson adapts a two machine flow shop problem to find a humping or switching which minimizes the maximum lateness of all outbound trains (12).

Assumption A1 - *All cars for a block will be switched before the block is considered completely switched.*

Assumption A2 - *An entire receiving track of cars will be switched before another track is started.*

Assumption A3 - *The model assumes sufficient track space to allow a "sort-by-block" switching scheme.*

Assumption A4 - *The model assumes one hump available for continuous operation.*

Assumption A5 - *The model assumes one pull-back engine in continuous service.*

Assumption A6 - *Values of the variables are assumed to be deterministic.*

Ferguson's first two assumptions deal with yard operations. Assumption A1 seeks to maximize the number of connections made between inbound cars and outbound trains. The second assumption simply describes how most rail yards operate. With regards to time, it is more efficient to complete the humping of an <u>uncompleted</u> string of cars than it is to withdraw and secure a different string. In short, the Assumption A2 does not allow job pre-emption. As noted by Ferguson (12), "Such pre-emption, if allowed, would exponentially increase the number of ways the hump sequence could be determined."

The next three assumptions - A3, A4, and A5 - address the resources available within the yard. And the last assumption - A6 - requires that the information about the block type and location of each car in every inbound train is available, as well as the blocking plan and the departure time for each outbound train.

The sixth assumption also requires that the yards average humping rate and the assembly time for each outbound train are known. As stated by Ferguson (12), this last part of "Assumption A6 is probably the least realistic. Preprocessing times are variable both in a stochastic sense and in a sequence-dependent sense, based on the track configurations and train locations in a yard." None the less, Assumptions A6 allows Ferguson to "get his arms around the problem" and identify areas for future work.

C. Ferguson's Heuristic

The following heuristic - in general - describes Ferguson's model:

Step 1 - Calculate the Processing Time (P_k) for each Inbound Train (I_k) by dividing the Yard's Average Humping Rate (cars/minute) into the total number of cars in I_k.

Step 2 - Determine the Hump Processing Time (H_k) for each I_k by adding the Yard's Average Hump Set-up Time to each P_k.

Step 3 - Determine the Slack Time (s_{ijk}) for each Block (B_{ij}) on I_k.

$$s_{ijk} = \frac{(\text{Position of Block } B_{ij}\text{'s Last Car } I_k - 1)}{\text{Average Humping Rate (cars/minute)}} \quad [15]$$

Step 4 - Given both the Blocking Plan, the Number of Blocks (n_j), and the Due to be Complete Time (C_j) for Outbound Train O_j, calculate the Due to be Complete Time (d_{ij}) for each B_{ij}. (Based on Ferguson's observations, at an unnamed rail yard, the processing or assembly time for <u>any</u> outbound train if 120 minutes.)

$$d_{ij} = C_j - (120 \text{ minutes}/n_j)(n_j - i) \quad [16]$$

Step 5 - If B_{ij} is spread over more than one I_k, set B_{ij}'s Slack Time (s_{ijk*}) equal to the $\max\{s_{ijk}\}$ where $k = 1, 2, 3, \ldots n$.

Step 6 - Calculate the Pull-Back Start Time (ps_{ij}) for each B_{ij}.

$$ps_{ij} = D_j - p_{ij} + s_{ijk} \qquad [17]$$

Step 7 - Sort all ps_{ij}'s from low to high. This is the "Suggested" Block Hump Sequence.

Step 8 - Calculate the Hump Completion Time (h_{ij}) for all B_{ij}.

$$h_{ij} = H_k - s_{ijk} \qquad [18]$$
$$+ \text{ Average Hump Set-up Time}$$

Step 9 - For each B_{ij}, calculate its Block Completion Time c_{ij}.

$$c_{ij} = \max\{h_{ij}, c_{ij} \text{ of previous } B_{ij}\} + p_{ij} \qquad [19]$$

Step 10 - For each B_{ij}, calculate its Lateness (l_{ij}).

$$l_{ij} = c_{ij} - d_{ij} \qquad [20]$$

Step 11 - For each O_j, calculate its Lateness (L_j).

$$L_j = \max\{l_{ij}\} \qquad [21]$$

Example (12):

There are three inbound trains to be humped which affect the departure of two outbound trains; see Tables 1 and 2:

Inbound Train 1 - 82 Cars	
Block	Last Car Position
31	16
11	38
Inbound Train 2 - 63 Cars	
Block	Last Car Position
12	41
Inbound Train 3 - 40 Cars	
Block	Last Car Position
12	19

Table 1 - Inbound Train Data

Outbound Train 1 Due to be Complete - 180 minutes	
Block	
11	
21	
31	
Outbound Train 2 Due to be Complete - 310 minutes	
Block	
12	
22	

Table 2 - Outbound Train Schedule

For this rail yard, the average hump set-up time is 20 minutes, the average humping rate is 1.5 cars/minute, and the hump start time is zero. (Note that blocks B_{21} and B_{22}, have already been humped and are waiting on Outbound Trains 1 and 2 to be assembled.)

Step 1 - Calculate the Processing Time (P_k) for each Inbound Train (I_k) by dividing the Yard's Average Humping Rate (cars/minute) into the total number of cars in I_k.

P_1 = (85 cars)/(1.5 cars/minute) = 55 minutes
P_2 = (63 cars)/(1.5 cars/minute) = 42 minutes
P_3 = (40 cars)/(1.5 cars/minute) = 27 minutes

Step 2 - Determine the Hump Processing Time (H_k) for each I_k by adding the Yard's Average Hump Set-up Time to each P_k.

H_1 = 55 minutes + 20 minutes = 75 minutes
H_2 = 42 minutes + 20 minutes = 62 minutes
H_3 = 27 minutes + 20 minutes = 47 minutes

Step 3 - Determine the Slack Time (s_{ijk}) for each Block (B_{ij}) on I_k.

s_{311} = (16 cars - 1 car)/(1.5 cars/minute) = 10 minutes
s_{111} = (38 cars - 1 car)/(1.5 cars/minute) = 25 minutes
s_{122} = (41 cars - 1 car)/(1.5 cars/minute) = 27 minutes
s_{123} = (19 cars - 1 car)/(1.5 cars/minute) = 12 minutes

Step 4 - Given both the Blocking Plan, the Number of Blocks (n_j), and the Due to be Complete Time (C_j) for Outbound Train O_j, calculate the Due to be Complete Time (d_{ij}) for each B_{ij}.

d_{11} = 180 minutes - (120 minutes/3) (3-1) = 100 minutes
d_{21} = 180 minutes - (120 minutes/3) (3-2) = 140 minutes
d_{31} = 180 minutes - (120 minutes/3) (3-3) = 180 minutes
d_{12} = 310 minutes - (120 minutes/2) (2-1) = 250 minutes
d_{22} = 310 minutes - (120 minutes/3) (2-2) = 310 minutes

Step 5 - If B_{ij} is spread over more than one I_k, set B_{ij}'s Slack Time (s_{ijk*}) equal to the max{s_{ijk}} where k = 1, 2, 3, . . . n.

s_{12k*} = max {27 minutes, 12 minutes} = 27 minutes

Step 6 - Calculate the Pull-Back Start Time (ps_{ij}) for each B_{ij}.

ps_{11} = 100 minutes - 40 minutes + 25 minutes = 85
ps_{21} = 140 minutes - 40 minutes + 0 minutes = 100
ps_{31} = 180 minutes - 40 minutes + 10 minutes = 150
ps_{12} = 250 minutes - 60 minutes + 27 minutes = 217

ps_{22} = 310 minutes - 60 minutes + 0 minutes = 250

Step 7 - Sort all ps_{ij}'s from low to high. This is the "Suggested" Block Hump Sequence.

"Suggested" Block Hump Sequence: (B_{11}, B_{21}, B_{31}, B_{12}, B_{22})
∴ the Inbound Train Hump Sequence is (I_1, I_3, I_2). (Inbound Train I_3 is humped before I_2 because I_2 has the greater Hump Processing Time; $H_2 > H_3$.)

Step 8 - Calculate the Hump Completion Time (h_{ij}) for all B_{ij}.

h_{11} = 55 minutes - 25 minutes + 20 minutes = 50 minutes
h_{21} = 0 minutes; B_{21} has already been humped
h_{31} = 55 minutes - 10 minutes + 20 minutes = 65 minutes
h_{12} = 184 minutes - 12 minutes + 20 minutes = 172 minutes
h_{22} = 0 minutes; B_{22} has already been humped

Step 9 - For each B_{ij}, calculate its Block Completion Time c_{ij}.

c_{11} = max {50 minutes, 0 minutes} + 40 minutes = 90 minutes
c_{21} = max {0 minutes, 90 minutes} + 40 minutes = 130 minutes
c_{31} = max {65 minutes, 130 minutes} + 40 minutes = 170 minutes
c_{12} = max {172 minutes, 170 minutes} + 60 minutes = 232 minutes
c_{22} = max {0 minutes, 232 minutes} + 60 minutes = 292 minutes

Step 10 - For each B_{ij}, calculate its Lateness (l_{ij}).

l_{11} = 90 minutes - 100 minutes = -10 minutes
l_{21} = 130 minutes - 140 minutes = -10 minutes
l_{31} = 170 minutes - 180 minutes = -10 minutes
l_{12} = 232 minutes - 250 minutes = -18 minutes
l_{22} = 292 minutes - 310 minutes = -18 minutes

Step 11 - For each O_j, calculate its Lateness (L_j).

$$L_1 = \max \{-10 \text{ minutes}, -10 \text{ minutes}, -10 \text{ minutes}\} = -10 \text{ minutes}$$
$$L_2 = \max \{-18 \text{ minutes}, -18 \text{ minutes}\} = -18 \text{ minutes}$$

The hump sequence in Step 7 is only "suggested" because of Ferguson's second assumption (A2); "An entire receiving track of cars [or inbound train] will be switched before another track [or inbound train] is started" (12). If an inbound train, I_A, contained not only the first "suggested" block, B_A, but also several other complete blocks that appear later in the "suggested" hump sequence (e.g. blocks B_D, B_E, and B_F), then the hump sequence "as executed" would be (B_A, B_D, B_E, B_F, B_B, B_C) versus the "suggested" sequence of (B_A, B_B, B_C, B_D, B_E, B_F).

5. COMPARISON OF THE THREE MODELS

A. General

Admittedly, trying to compare Shi's, Kraft's, and Ferguson's models is a lot like trying to compare "apples and oranges". The objective of both Shi's Hump Sequencing System and Ferguson's Switch Processing Model is to optimize a yard's hump sequence while the goal of Kraft's Mixed-Integer Optimization Model is to formulate simultaneously optimum hump sequence and classification track assignments. None the less, it is possible to compare the underling assumptions of all three models and their objectives or goals.

B. Goals

In the first chapter, Keaton's (19) three strategies for improving rail service were introduced:

(i) technological improvements such as pacing more than one train over a given set of track,

(ii) providing more direct and frequent "non-stop" trains, and

(iii) improving the decision support systems at rail marshalling yards by optimizing both terminal operations and blocking plans.

The underling theme of this paper is, of course, Keaton's third strategy. As noted above, the goal of both the Hump Sequencing System and the Switch Processing Model is to optimize a rail yard's terminal operations; specifically optimize a rail yard's hump sequence. Of the three models addressed, only Kraft's Mixed-Integer Optimization Model takes a more "holistic approach" in an attempting to optimize simultaneously a yard's hump sequence and its classification track assignments.

As discovered by Kraft (21), the addition of the classification track assignment problem to the hump sequencing problem significantly increases the difficulty of the "overall" problem. Thus, Kraft's goal can be said to be the most challenging.

C. Assumptions

In the process of developing their models, Shi (Hump Sequencing Model -HSS), Kraft (Mixed-Integer Optimization Model - MIOM), and Ferguson (Switch Processing Model - SPM) each had to make a series of relatively minor assumptions; see Table 3.

	HSS	MIOM	SPM
Assumption A1	Only those inbound trains or blocks that have been inspected will be considered.	The "trip plan" for all cars, both incoming and on hand, is known.	All cars for a block will be switched before the block is considered completely switched.
Assumption A2	Advance information about "new" inbound trains will not impact on a humping sequence.	All cars in an inbound train will be switched before the inbound train is considered completely switched.	An entire receiving track of cars will be humped before another track is started.
Assumption A3	The failure of an outbound train to depart at the scheduled time will not be due to lack of power, a switch engine, or an inspection crew.	Sufficient personnel and switch engines are available to operate all three yards - receiving, classification, and departure - simultaneously.	The model assumes sufficient track space to allow a "sort-by-block" switching scheme.
Assumption A4	No car on an inbound train will be considered humped before the inbound train is considered completely humped.	The failure of an outbound train to depart on schedule will not be due to a lack of power.	The model assumes one hump available for continuous operation.
Assumption A5	Each inbound train is assumed to be humped entirely or not at all.		The model assumes one pull-back engine in continuous service.
Assumption A6	All cars on one or more classification track(s) that have been tagged for the same outbound train will have the same "block-out" time.		Values of the variables are assumed to be deterministic.

Table 3 - Assumptions (12, 21, 30)

Given that all three models address operations in a rail marshalling yard, it is not surprising that there are a number of similar assumptions; see Table 4.

HSS	MIOM	SPM
A3	A4	
A4		A1
A5	A2	A2
	A3	A5

Table 4 - Matrix of Similar Assumptions

Because the Hump Sequencing System and the Switch Processing Model seek to find an optimum hump sequence for only those rail cars that have both been inspected and tagged (HSS Assumptions A1 and A2), knowledge of the trip plans for incoming cars (MIOM Assumption A1) is not relevant. With regards to HSS's Assumption A6 and SPM's Assumption A3, Kraft's Mixed-Integer Optimization Model makes do without these assumptions because its goal is to simultaneously find both an optimum hump sequence and an optimum set of classification track assignments. Of special note though, is the one assumption that missing in all three models; *No hump cars will be switched out of inbound trains while they are being inspected and tagged.*

Just one no hump car "buried" in an inbound train can significantly perturb a hump sequence. If the car is not switched out prior to the train being humped, when the car comes to the top of the queue, the inbound train will need to be backed away from the hump in order to allow the car to be switched. While Shi does state in his thesis that HSS makes no provisions for no hump cars, neither Kraft nor Ferguson mention this potential problem.

6. SUMMARY AND CONCLUSION

A. Summary

In this research paper, three different models designed to help optimize rail marshalling yard operations have been discussed; Shi's Hump Sequencing System, Kraft's Mixed-Integer Optimization Model, and Ferguson's Switch Processing Model. Using, respectively, dynamic programming and scheduling theory, both Shi's and Ferguson's models seek the optimum solution to a yard's hump sequence. Kraft's model - using integer programming - seeks simultaneously a rail yard's optimum hump sequence and optimum classification track assignments.

B. Conclusion

Based on the description of the three models addressed in this survey, and the open literature listed in the bibliography, it is the author's opinion that work similar to Kraft's Mixed-Integer Optimization Model - trying to solve simultaneously two or more different marshalling yard problems - is the direction in which future research should proceed. Though perhaps the most "industry ready" of the three, Shi's Hump Sequencing System is somewhat dated and can probably be better modeled using integer programming. Ferguson's work on the application of scheduling theory to a rail yard may be able to help solve Kraft's excessive memory/execution time problem. Yet still, as described by Ferguson, the Switch Processing Model can be best thought of as a "proof of a scheduling theorem for minimizing maximum lateness in a disassembly/assembly process" (12).

However, all this having been said, it is important not to forget the rail yard's "human dimension". As noted by Ferguson (12), "The potential for application of scheduling theory [or of any other operations research discipline] in railroad operations exists only with the caveat that these are not true machines performing the switching work. There is still a great deal of human involvement, with all its vagaries, in the switching process . . . ."

As rail companies try and implement dynamic sorting strategies, it will become more and more difficult for switching crews to understand which blocks are where, which blocks belong on which outbound trains, and where the cars in "Outbound Train X, Block B" will be reswitched to while cars in "Outbound Train X, Block A" are assembled. Perhaps what is needed is some type of a flat screen color monitor that can be placed in the switch engine that shows the switch engineer an "overhead shot" of the rail yard and indicates where he need to go for the next block of cars, what departure track to place them on, and which *side track* or classification track to use if reswitching is necessary.

BIBLIOGRAPHY

1. Armstrong, J. H. (19--), The Railroad -- What it is, What it Does (Third Edition). Simmons-Boardman Books, Inc.

2. Assad, A. A. (1979), Models for Rail Transportation. *Transportation Resources*, **14A**, 205-220.

3. Association of American Railroads (1992), Railroad Facts - 1992 Edition. Economic and Finance Department, Washington, D.C.

4. Association of American Railroads (1993), Background Paper; Railroads - The Clean Choice for the Future. Communications Department, Washington, D.C.

5. Bick, D. E. (1984), A History of the Dowty Marshalling Yard Wagon Control System by D. L. Turner. *Proceedings of the Institution of Mechanical Engineers*, **198B**, 19-26.

6. Bodin, L. D., Golden, B. L., Schuster, A. D., and Romig W. (1980), A Model for the Blocking of Trains. *Transportation Resources*, **14B**, 115-120.

7. Crainic, T., Ferland, J. A., and Rousseau, J. (1984), A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, **18**, 165-184.

8. Daganzo, C. F. (1986), Static Blocking at Railyards: Sorting Implications and Track Requirements. *Transportation Science*, **20**, 189-199.

9. Daganzo, C. F. (1987), Dynamic Blocking for Railyards: Part I. Homogeneous Traffic. *Transportation Resources*, **21B**, 1-27.

10. Daganzo, C. F. (1987), Dynamic Blocking for Railyards: Part II. Heterogeneous Traffic. *Transportation Resources*, **21B**, 29-40.

11. Daganzo, C. F., Dowling, R. G., and Hall, R. W. (1983), Railroad Classification Yard Throughput: The Case of Multistage Triangular Sorting. *Transportation Resources*, **17A**, 95-106.

12. Ferguson, R. T. (1993), Mathematical Modeling of the Railroad Switching Process. Doctor of Philosophy Thesis (Systems Engineering), University of Virginia.

13. General Accounting Office (1992), Operation Desert Shield -- Problems in Deploying by Rail Need Attention. Report to the Chairman, Subcommittee on Readiness, Committee on Armed Services, House of Representatives.

14. Gentzel, R. A. (1981), A Comparison of Rail Yard Queuing Models: Analytic Queuing Models Versus Computer Simulation. Master of Science in Transportation Thesis, Northwestern University.

15. Hammesfahr, R. D. (1981), A Simulation Model for the Analysis of Railway Intermodal Terminal Operations. Doctor of Philosophy Dissertation, Virginia Polytechnic Institute and State University.

16. Jiahao, W. (19--), Optimized Design of Marshalling Yard for Heavy Haul Railway. *Transportation and Economics Research Institute - China Academy of Railway Sciences*, 67-71.

17. Keaton, M. H. (1985), Optimizing Railroad Operations. Doctor of Philosophy Thesis, University of Wisconsin-Madison.

18. Keaton, M. H. (1989), Designing Optimal Railroad Operating Plans: Lagrangian Relaxation and Heuristic Approaches. *Transportation Resources*, **23B**, 415-431.

19. Keaton, M. H.(1991), Service-Cost Tradeoffs for Carload Freight Traffic in the U. S. Rail Industry. *Transportation Resources*, **25A**, 363-374.

20. Kraay, D., Harker, P. T., and Chen B., Optimal Pacing for Trains in Freight Railroads: Model Formulation and Solution. *Operations Research*, January-February 1991, 82-99.

21. Kraft, C. (1993), A Mixed-Integer Optimization Model to Improve Freight Car Classification In Railroad Yards. TIMS/ORSA 1993 Joint National Meeting Phoenix, Arizona.

22. Lawrence, M. B. and Sughart, L. A. (199-). *Transportation Research Forum*, ---, 54-69.

23. Little, P. and Martland, C. D. (1993), Causes of Unreliable Service in North American Railroads. Unpublished Study for the Center for Transportation Studies, Massachusetts Institute of Technology.

24. Martland, C. D. (1982), PMAKE Analysis: Predicting Rail Yard Time Distributions Using Probabilistic Train Connection Standards. *Transportation Science*, **16**, 476-506.

25. Martland, C. D. (1992), Managerial and Economic Approaches to Measuring Freight Railway Performance. The 6th World Conference on Transport Research Lyon, France.

26. McKinsey & Company (1992), Improving CBU Performance through Integrated Fleet Management: Phase 1 - Progress Review. Unpublished Study for the Atchison, Topeka, and Santa Fe Railway.

27. Melhuish, A. W. (19--), Economic Design Methods for Automated Miniyards. *Transportation Research Record*, **1030**, 45-48.

28. Missouri Pacific Railroad (1981), Missouri Pacific's Computerized Freight Car Scheduling System. U. S. Department of Transportation.

29. Petersen, E. R.(1977), Railyard Modeling: Part II. The Effect of Yard Facilities on Congestion. *Transportation Science*, **11**, 50-59.

30. Shi, Q. (1981), On the Railway Hump Yard Sequencing Problem. Unpublished Masters of Arts and Science Thesis, Department of Civil Engineering, University of Waterloo.

31. Shinohara, T. (1963), Theory of Freight Yard Planning. *Monthly Bulletin of the International Railway Congress*, **40**, 161-172.

32. Sidey, H. (1993), Hugh Sidey's America - Back at Full Throttle. *Time*, **142**, August 23, 1993, 52-55.

33. Tadi, R. R. and Palaniswamy S. P. (1984), Comparison of Various Freight Train Dispatching Policies for Indian Railways. *Journal of Transportation Engineering*, **110**, 357-365.

34. Turnquist, M. A. and Daskin, M. S. (1982), Queuing Models of Classification and Connection Delay in Railyards. *Transportation Science*, **16**, 207-230.

35. Van Dyke, C. D. (1986), The Automated Blocking Model: A practical Approach to Freight Railroad Blocking Plan Development. *Transportation Research Forum*, **27**, 116-121.

36. Wetzel, J. A. (1983), Transportation Research Record 927 Innovation in Classification Yard Technology: Proceedings of the Third Railroad Classification Yard Workshop. Transportation Research Board.

37. Marshalling Yard. *Rail International,* **3**, 336-348.

38. Yagar, S., Saccomanno, F. F., and Shi, Q. (1983), An Efficient Sequencing Model for Humping in a Rail Yard. *Transportation Resources*, **17A**, 251-262.

39. Yun, H. (1989), Fuzzy Model of Hump-Yard Locomotive Speed Control System. *Modeling, Simulation & Control,* **26**, 53-64.