Terminal Operating Plan and Achievability: an Approach to Improve Rail Terminal Performance

by

Yan Dong

B.S., Southwest Jiaotong University, 1984
M.S., Northern Jiaotong University, 1987

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Master of Science
in Transportation

at the

Massachusetts Institute of Technology

August 1994

Signature of Author ....
Department of Civil and Environmental Engineering
August 11, 1994

Certified by ....
Patrick Little
Research Associate, Center for Transportation Studies
Thesis Supervisor

Accepted by ....
Joseph M. Sussman
Departmental Committee on Graduate Studies

Terminal Operating Plan and Achievability: an Approach to Improve Terminal Performance

by

Yan Dong

Submitted to the Department of Civil and Environmental Engineering
on August 12, 1994 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Transportation

ABSTRACT

Recent studies have shown that rail terminals are a critical element in railroad operations, particularly for handling manifest and intermodal traffic. In spite of their importance to the reliable functioning of rail networks, terminal managers are often left to their own devices in developing and evaluating alternative methods of operation. Most terminal managers, whether at the level of the Terminal Superintendent or the Hump Yardmaster, conduct their operations on a *de facto* operating plan. In many cases, this terminal operating plan calls for following a simple decision rule, such as First In/First Out (FIFO), with *ad hoc* exceptions for specific trains or unusual circumstances. There are virtually no tools available to assist terminal managers in evaluating the quality of their current or alternative terminal operating plans, either in terms of the effect on accomplishing system objectives or the likelihood of carrying out specific tasks.

The research in this thesis focuses on developing models to assist terminal managers in formulating and evaluating *terminal operating plans (TOP)*, which is different from the *de facto* plan. In order to better understand terminal operations, terminal operations' *behavior models* are developed, which are based on a detailed activity data base. To assist terminal studies, the *data requirements* are discussed in the thesis. The results from the behavior models can be used in real-time PMAKE analysis and an assignment model. *Real-time PMAKE* analysis is an aggregate model to predict terminal train connection performance and to provide useful information to the terminal managers in the development of TOP. The *assignment model* presented in the thesis can be used to generate a detailed TOP, which uses train connection performance as the objective function, constrained by terminal managers' expectation about the time to perform each task in the terminal.

When selecting TOPs, not only should train connection performance be considered but also the likelihood of accomplishing the plans. To assist terminal managers in evaluating TOPs, the notion of *achievability* is developed. This is a measure of the probability that a set of tasks that are assigned by the terminal managers will be accomplished within the time allotted for completing the tasks. It can be used at the system level to assess the overall likelihood that the terminal will complete its set of tasks, at the terminal level to determine which processes are most in need of careful supervision, and at the task level to ensure that work assignments are reasonable. Methods have been developed to measure achievability at each level of the organization. A case study is presented to demonstrate that the achievability measure is potentially a very useful tool to support terminal operations management.

Thesis supervisor: Dr. Patrick Little
Title: Research Associate, Center for Transportation Studies

Acknowledgments

I would like to express my great appreciation for my supervisor, Dr. Patrick Little. His guidance, encouragement, and involvement during the entire process of this research have been invaluable and I learned a lot from him. Dr. Little is not only my supervisor but also a good friend of mine. We spent wonderful time together in rail terminals, seminars, and discussions.

I deeply thank Carl Martland who gave me a lot of guidance in my research, read the thesis and provided insightful advice, comments, and suggestions to revise the thesis.

I am grateful to Professor Joseph Sussman who cares about my research and provided many good suggestions throughout the thesis development.

I thank my friends and colleagues at MIT for their help and support. They are too many to mention. It is these friends that make me feel warmth in the study at MIT.

The research is supported by grants from the Association of American Railroads and Draper Lab to the Center for Transportation Studies of MIT. I appreciate the financial support from these organizations. Also, I thank CSXT for providing the data for the case study and the managers of CSXT for their insightful help about the terminal operations.

I would like to acknowledge my parents for their encouragement and affection. Finally I would like to express my great thanks to my wife, Zhili Li, for her love, care, patience and support.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

This thesis is focused on improving the operations of rail freight terminals. An approach for measuring and improving yard[1] service performance is presented. In addition to the methodology, a case study is conducted using the data from a major terminal in a Class I railroad[2]. Before presenting the approach, the research motivation and an outline of the thesis is presented in this chapter.

1.1 Research Motivation

Rail freight transportation is an important transportation mode for some customers, particularly for high volume, long haul shippers. The railroads are trying hard to attract more shippers, especially those looking to other modes. Various surveys and previous studies show that the most important factor preventing the railroads from getting more market share is service performance (for example, [Vieira 1991]). The railroad service performance is far behind compared with that of motor carriers for most markets ([Kulman 1974], [Temple et al. 1989, 1990] and [Vieira 1991]).

In a rail transportation system, there are many components affecting service performance. One of the most important of these is the terminal (for example, [Lang and Martland 1972] and [Sussman et al. 1972]). Many studies show that terminals are still a major problem in the improvement of rail service performance. For example, terminal operations are much more critical than train operations to the reliable movement of cars

---

[1] In this thesis, "terminal" and "yard" are used interchangeably.

[2] The US. railroads are organized into "classes" which were defined by the Interstate Commerce Commission based on their annual operating revenues. Class I railroads are the largest railroads. In 1986, for example, the basis for the Class I railroads was $88.6 million in annual operating revenues [Association of American Railroads 1987]

[Martland et al. 1982]. Terminal delays and related causes of service failure are one of the most important failures in a Class I railroad [Little et al. 1993].

A great portion of car cycle time is spent at yards. Statistics show that average car cycle time is about 26 days. Among them, about 15.8 days are spent at various yards from cars' origins to their destinations, which accounts for about 62% of the total car cycle time [Trope, 1975].

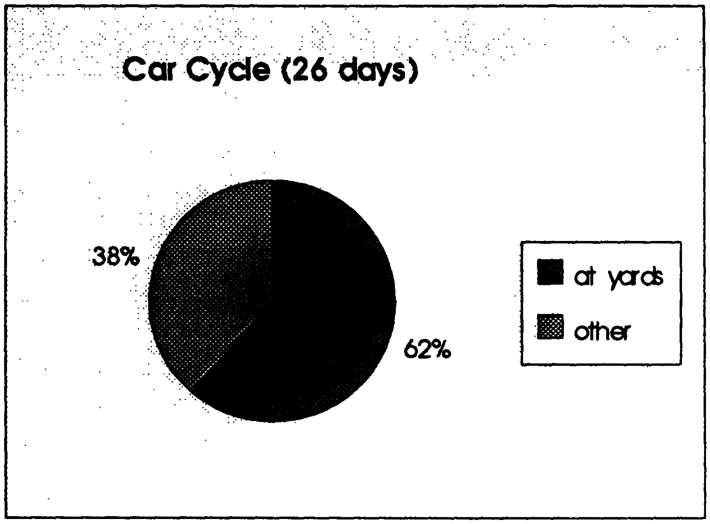

Figure 1. Average Car Cycle Time

For loaded cars, the average trip time from origin to destination is about 8.8 days, of which about 6.8 days are spent at yards, accounting about 77% of the total trip time [Reebie, 1972].

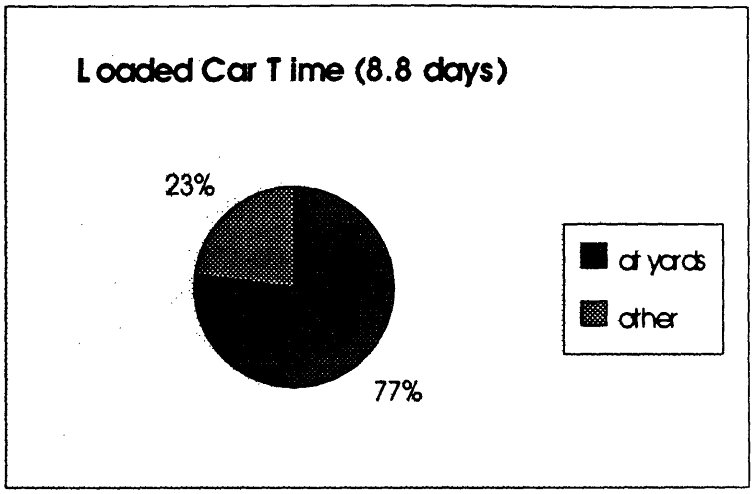

Figure 2. Average Loaded Car Time

In a recent study [Little et al. 1991], the 1970's data are compared with 1989 data. Cycles are not substantially changed. There is little reason to believe the percent of time in terminals have changed either. For a specific yard, there may exist large variation in terms of average yard time at different period of times. For example, in a case study conducted in 1993 in a major terminal of a Class I railroad, the scheduled average yard time is about 28 hours. For some periods of time, the observed average yard time was about 22 to 24 hours. For other periods of time, the observed average yard time was about 35 to 40 hours [Duffy, 1994].

The above figures (great portion and large variation) indicate that improvement in yard service performance may result in significant improvement in car cycle time, car trip time and average yard time which are all important performance measures. In other words, the improvement in yard operations has a big positive effect on the system level service improvement.

The car trip time may be divided into three parts: the time spent at yards, the time spent at line haul and the time spent at shipper and consignee. At present, railroads have very accurate line haul models to predict the running time over line segments based on the curves, the grades of the segments, the weights and lengths of trains and the available locomotives dragging the trains (for example, TEM model and TOES model). Also, the railroads have advanced algorithms to arrange the meets and passes of trains during their trips from their origins to their destinations [Morlok and Peterson 1970] and [Tsisiklis 1992]. The result is that line haul performance (or train-level performance) is much better than yard performance (car-level performance). The time cars spend at shippers and consignees, due in part to contracts between railroad and customers, is also more reliable than yard reliability[3]. In any case, customers are concerned with "dock to dock" time, so that delays at shipper and consignee are generally significant only in terms of asset utilization, not service reliability.

Terminal operations, on the other hand, do not have satisfactory performance. Many rail system service problems happen at terminals. For example, railroads found that cars are delayed at terminals, which affects rail service performance [Reid et al. 1972]. One reason for the delay is that cars are missing connections. The railroads found that missed connections are the number one cause of rail system service failure. This indicates that more effort should be focused on terminal service improvement.

Unfortunately, yard operations are still poorly understood, especially by system level managers. This leads system level managers to treat terminals as "black boxes", where cars enter and depart, but with processes that are not directly controlled or monitored by system level planners and managers. The relationship between input and output of the terminals is still not clearly understood. The yard managers, on the other hand, are expected to achieve better yard service performance under demanding time and resource constraints, often with poor or non-existent tools.

Contemporary production systems require more of transportation systems including integrated logistics systems and just-in-time systems. Higher levels of service are required by the customers. Various surveys show that customers treat service reliability as the most

---

[3] Shippers and consignees have an inherent incentive to control the time they hold cars for loading and unloading. In general, this is the result of car rental charges (demurrage) and in other cases represents the use of the car as a de facto warehouse.

important attribute of transportation modes [Vieira 1991]. To compete with other modes, railroads must improve their service performance. To cope with the requirements of the market, based on their experience, some railroads are attempting to run railroads strictly according to the operating plan.

The plan the railroads are referring to is the system level or operating plan. It includes train schedules, car scheduling, power plan, crew schedules and so on. The yard operations and yard plan are not explicitly addressed except in terms of connections between inbound trains and outbound trains at yards. As discussed earlier, a large portion of car time is spent at yards and there is substantial variation in car time. If yards are treated as black boxes, it may be very difficult to develop practical system level plans.

This thesis focuses on the issue of how to improve terminal service performance. A yard plan, which is called **Terminal Operating Plan (TOP)**, is proposed as a tool for yard managers to improve yard service performance. The definition of TOP is developed in chapter three. When developing the terminal operating plan, the **achievability** of the terminal operating plan, which measures the feasibility or robustness of the plan, is explicitly considered.

1.2 Outline of the Thesis

In this chapter, the importance of the terminal in rail transportation system has been briefly introduced. In chapter two, terminal operations and terminal performance measurement are briefly introduced, the current practice is presented, and previous studies are reviewed.

In chapter three, a strategy for improving yard service performance is presented. The major point of the strategy is to use the terminal operating plan as a tool to manage and control terminal operations. The chapter begins with a framework addressing this strategy.

Then the definitions of the terminal operating plan are given. The individual issues in the framework such as data requirements, process behavior models, real-time PMAKE analysis, and assignment model are presented in this order. In chapter four, the concept of achievability of the terminal operating plan is presented and methods to measure the achievability of terminal operating plan are developed and applied. A case study using data from a major terminal of a Class I railroad in US is presented in chapter five. The results of the case study indicate the usefulness of the approach presented in this thesis. In chapter six, the major conclusions of the thesis and possible future studies are given.

Chapter 2. Review of Terminal Operations and Studies

2.1 Terminal Processes

Rail traffic at terminals is basically handled in one of two ways. Some traffic bypasses the yard, either because it does not need classification or because it is handled in specific, high priority movements. This traffic can include trains which stop for other services or groups of cars which are set off by one train for picking by another ("block swapping"). For bypassing traffic, terminal operations are simple. They may include changing crew or road engines, adding fuel, water or sand to the engines and some paper work. The bypassing traffic is not likely to have service performance problems. In this thesis, the focus is placed on non-bypassing traffic. This is the traffic which comes to or from customers or local trains, or which is set off by trains for further classification en-route. The major work of the terminal is to assemble outbound trains from inbound traffic. In order to do so, the inbound trains must be first classified. So the terminal has two major functions: classifying inbound trains, which is also called the hump operation for hump yard, and assembling outbound trains[4].

A rail terminal can be thought of as an assembly plant for trains. Like other assembly plants, the raw materials, which are the inbound trains, and final products, which are outbound trains, may be inspected. So there are two inspection operations: inbound inspection and outbound inspection. For rail terminals, there are also another two operations: inbound trains' arrival which includes deciding where to put the train, and removal of power and crews and outbound trains' departure including attachment of power

---

[4] In practice, the terminal often has a number of other functions which are associated with it such as servicing power, serving as a reporting point for crews, repairing defective cars, etc. Since this thesis is primarily concerned with the processing of cars through the terminal, these other auxiliary functions can be treated as separate.

and crews to the train, and transferring authority to the line dispatcher. To sum up, there are six major processes in terminals. They are:

- inbound (IB) arrival
- inbound (IB) inspection
- classification (hump for hump yard)
- assembly
- outbound (OB) inspection
- outbound (OB) departure

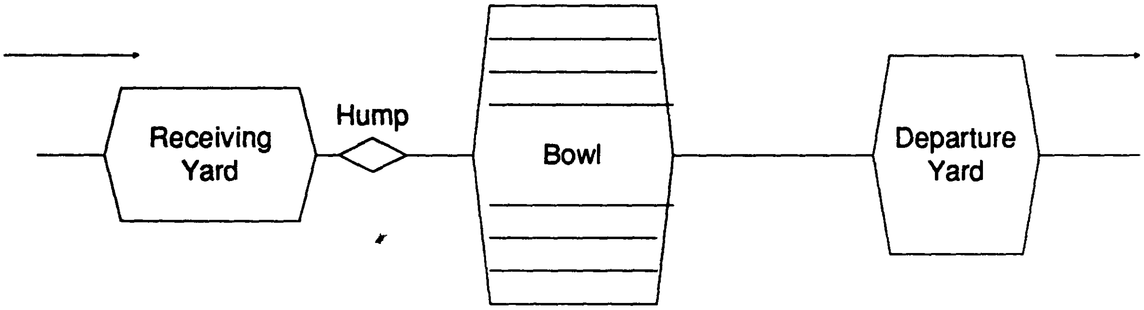

Figure 3. A Terminal Layout

Each of the operations or processes is described below:

Inbound Arrival: inbound train arrives at the yard. This process includes yarding the inbound train to the receiving track, disconnecting road engines, removing the end-of-train (EOT) device which is placed on the last car of the train to indicate the end of the train, and some paper work. After the inbound train arrives at a receiving track, it is waiting for inbound inspection.

Inbound Inspection: inspectors examine the cars in the inbound train to check the physical condition of the cars. If a car has some mechanical defect and may not continue its trip without repairs, this car is called a bad order car. For light bad order cars, if time is available, the inspectors may repair the car at the receiving track. Heavy bad order cars need go to the car shop for repair. After the inbound train is inspected, the cars in the inbound train are waiting for the classification operation. The inbound inspection is not required by federal law. Railroads conduct the inbound inspection to improve the mechanical reliability and safety of the fleet, and to reduce the likelihood of defective cars being detected on outbound trains.

Classification: This process consists of disassembling the inbound train and reorganizing the cars into outbound groups or blocks of cars with common intermediate or final destinations. The blocks are generally specified by the system level operating plan.

At a hump yard, a switching engine pushes the cars from the receiving track across a raised section of the track known as the hump. At the crest of the hump, a worker disconnects the cars. A string of cars having a common destination, which is called a cut, then go down to the bowl of the yard. During the hump process, some cars may not move completely to their assigned tracks. In this situation, the hump engine may need to stop the hump process and go down to the bowl tracks to push the cars to their tracks. In addition, cars carrying certain commodities such as high value items or hazardous materials are restricted from humping. These cars must be set aside and handled separately from the other cars. This process is called trim work[5]. After the cars at the receiving track are humped, they are waiting for assembly.

There are a number of ways to classify an inbound train in a flat yard. One way is as follows: the classification engine pushes the cars from the receiving track to a track in the bowl. After the cars arrive at the track of the bowl, a switch crew disconnects the cars and the classification engine drags the remaining cars up a "ladder track" for the second switch. This process continues until all the cars are classified into the tracks of the bowl. Another way is that during the push process, the switch engine first speeds up, then the switch crew

---

[5] There is a trade off between having sufficient speed to avoid trim work and avoiding loss and damage to loading.

disconnects the cars. The switch engine then brakes, making the disconnected cars go to a track of the bowl. This method, known as "kicking cars", may need less time compared with the first method.

Assembly: after enough traffic volume is accumulated or when a train is scheduled for departure, the assembly operation is started. Assembly is the gathering together of blocks into a group which constitutes an outbound train. In some cases, the cars within a block are not properly connected on a single track. The assembly engine needs to push the cars to be connected on the track. In other cases, cars from more than one block are on a track and must be separated out. Depending on the number of blocks, the degree to which the blocks are joined, the location of blocks in the yard (i.e., nearness of tracks) and amount of additional switching, assembly can vary from a very simple process to a very complex one. In many yards, more than one engines is involved in assembling trains, creating the possibility of conflict between them. Usually the assembly engine pulls the connected cars for the outbound blocks to the departure track. During the assembly process, there may be additional sorting work to do according to the consist and sequence requirements of the outbound train. After the outbound train is assembled, it is waiting for the outbound inspection operation.

Outbound Inspection: outbound inspectors examine the outbound train to check the physical condition of the running gear and brakes of the cars in the outbound train. If some bad order cars are found, these cars may not be departed from the yard before necessary repairs. They may be sent to car shop or be fixed at the departure track depending on the bad order situation. After all the cars are connected and the road power is attached, the air brake test is conducted. (In some yards, there is in-ground piping of compressed air, known as "yard air", which is used for the brake test prior to attachment of road power). This is a mandated inspection under the Power Brake Law. After the outbound train is inspected, the train is waiting for departure.

Outbound Departure: After the road engines are attached to the cars at the departure track and the outbound inspection process is finished, the outbound train is ready for departure. After the departure signal is given, the outbound train departs from the yard.

In addition to these processes, these are a number of other processes and tasks to be performed in a terminal, including switching local industries, spotting cars on repair tracks, and bringing trains which have exceeded the Federal Hours of Service Law[6] into the terminal. Local pick-up and delivery operations in the vicinity of the yard can be treated as arrival or departure processes. Some high priority non-bypassing traffic may not go through all the six processes if the available time to make a connection is limited. For example, they may omit inbound inspection or be classified by the assembly engine.

Some studies have focused on the middle four processes, omitting the inbound arrival and outbound departure processes. Since this thesis focuses on yard service improvement, all the processes in the yard are explicitly considered. Also, there are responsibility changes from arrival to inbound inspection and from outbound inspection to departure. From arrival to inbound inspection operation, the responsibility is transferred from transportation department to the mechanical department in the terminal. From the outbound inspection to the departure operation, the responsibility is transferred from the mechanical department to the transportation department in the yard. Also, there is a transfer of control in these operations from system level manager (dispatcher) to yard or vice versa. In order to achieve better yard service performance, all the time the cars spend in the yard should be explicitly considered.

---

[6] The rule that a train crew may not work more than 12 hours before being given an eight hour rest. Upon reaching the 12 hour limit, the train must be stopped and a replacement crew used to complete the train's run.

2.2 Terminal Performance Measures

There are different dimensions of terminal performance and many terminal performance measures have been offered. Rather to list all of these, some of the most often used measurements in the terminal operations are given. They are:

service performance: service performance is the extent to which processes in the terminal corresponds to that called for in the operating plan or in commitments to customers. What the railroad provides to the customers is the transportation service. This measurement is also important for the service planners or the market planning department of the railroad. Some widely used service performance measures in terms of yard operation are as follows:

- connection performance (e.g., percent of cars making their most appropriate (first) connections, PMAX, T50, T90 [Martland 1982])
- average yard time

resource utilization: resource utilization measures the efficiency of the terminal resources being utilized. Both the yard manager and system managers are concerned with the yard resources level and utilization. Since the yard resource utilization is related with operating cost, the operating departments may also be concerned with this performance measurement. Some of the yard resources utilization measures are:

- crew working time in a shift
- engine working time in a shift
- crew and engine idle time in a shift
- percentage of time crew and engine working in a shift

processing rates: processing rates measure the speed that terminal tasks can be performed. Yard managers are concerned with the processing rates. When the yard manager plans his work, the processing rates are some of the factors being considered. They are also concerned by the system level manager in system level planning . Some of the most often used yard processing rate measures are:

- number of cars inspected per inspector per shift
- number of cars humped per engine per shift
- number of cars assembled per engine per shift
- number of cars handled per clerk hour

operating costs: operating costs measure the cost of terminal operations. Cost control department and the yard manager are concerned with this performance measure. System level managers may also be concerned about this measure. Some of the most often used measures of operating cost are:

- total costs
- costs relative to budget
- cost per car handled

Of the above measures, terminal service performance measures are of the highest importance to the system level managers, particularly those concerned with meeting customer commitments. As discussed earlier, missed connections are the number one cause of rail service failure. Connection performance is an important element of reliability. In this thesis, improving terminal service performance is the primary focus. Specifically, improving connection performance such as the percentage of cars making their most appropriate connection is considered by better assignment of tasks to available resources in the terminal.

In this thesis, whenever the service performance is referred to, connection performance is meant. The objective is to maximize the number of cars making their first connections by better scheduling tasks in the shift. The byproducts may be that the processing rates can be increased, resources are better utilized and hence the operating costs may be reduced. That is, the strategy used in this thesis to improve yard service performance may have a positive effect on other yard performance improvement as well.

While system managers are concerned with service performance, and planners may be concerned with yard processes, yard managers must direct attention to specific tasks. These are activities which are assigned to particular crews or workers in the yard at a particular time. For example, "to classify inbound train #103" is a task. The purpose of the task is to classify the inbound train to make the cars of this inbound train available for assembly operation of outbound trains. There are many tasks performed in a terminal in a shift. The terminal may not be able to perform all the assigned or necessary tasks in the shift. The tasks to be performed in the shift constitutes the heart of the terminal operating plan (TOP). Since the purpose of TOP is to achieve better yard service performance, the tasks in TOP are important for the yard in terms of yard service performance.

2.3 Current State of the Practice

2.3.1 Decision Makers at Terminal: Organization Issues

Generally speaking, there are three layers of yard managers in yard operations management. The titles of the managers depend on the individual railroad company. The first layer manager or highest level manager is the person who is responsible for entire terminal (around the clock) usually called *terminal superintendent*, or *assistant superintendent*. The second layer manager is the person who is responsible for the all operations on a shift, usually called the *trainmaster*. The third layer managers are the persons who are responsible for specific functions within yard. *Yardmasters* are the assistants of the trainmaster for specific car movement operations. For example, there may be a hump yardmaster, who assists the trainmaster to plan, manage and control inbound traffic operations including arrival, inbound inspection and hump operations. There may also be a bowl yardmaster, who assists the trainmaster to plan, manage and control outbound operations including assembly of outbound trains, outbound inspection and departure operations.

There is also a *mechanical department supervisor* who is responsible for car inspection and locomotive shop operations. The *car inspection supervisor* is responsible for inbound and outbound inspection operations on a shift. The *locomotive shop manager* is responsible for the provision of road engines for the outbound trains on a shift.

In addition to the vertical layers of authority, there are interdepartmental limits on authority. The higher layer manager of one department may not have authority over the lower layer workers of another department. For example, the hump yardmaster can give directions or task requirements to the inbound inspectors, but the hump yardmaster does not have full authority over the inbound inspectors. The car inspection supervisor has authority over the inbound inspectors but he does not give detailed direction or task requirements to his subordinates except when there is conflict between the hump yardmaster and the inspectors. Under this situation, the car inspector supervisor will coordinate the two sides' work and make a decision, if necessary, such as whether the inspectors will do the task required by the hump yardmaster. One reason for the complexity and difficulty of terminal operations is due to the separation of authority and direction. For some subordinates, there are more than one source of instructions. This organization will result in some management problems. But on the other hand, it may be difficult to organize the terminal in such a way that every worker in the terminal has only one supervisor. The reason is that the terminal operations are accomplished by different functional sub-organizations in the terminal, such as the transportation department, and the mechanical department (car inspection, and power shop), and all these functional sub-organizations must be coordinated for the transportation service.

There are a number of workers in the yard, usually grouped along labor groupings, or "craft" lines. They include switching crews, who operate a switching engine or trim engine and do the tasks the yard managers such as trainmaster and yardmasters give them and inspectors, who inspect inbound trains or outbound trains. The organization of the yard can be represented using Figure 4.

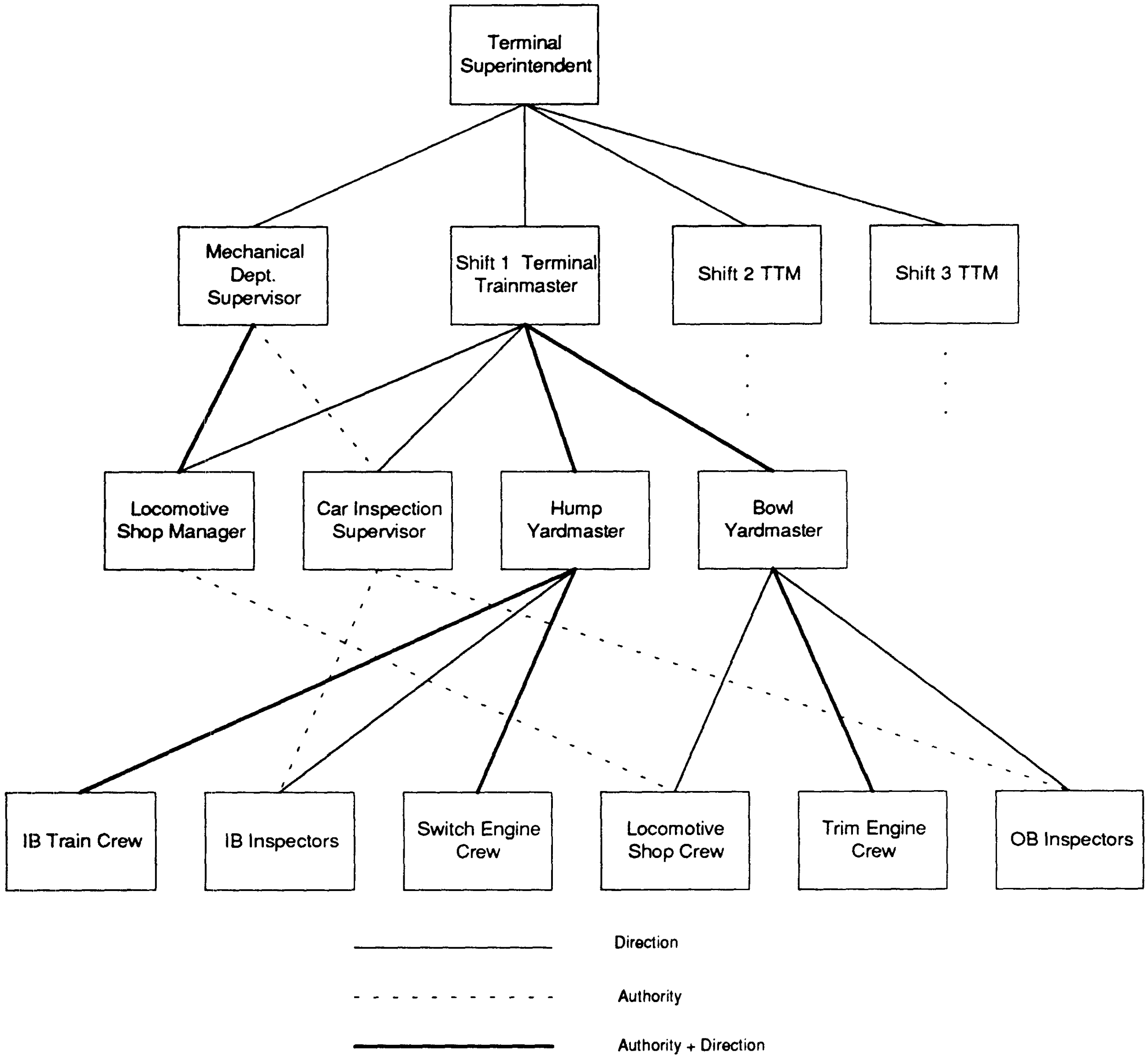

Figure 4. Yard Organization/Structure

2.3.2 Task Management in Terminal Operations

The major work of the yard managers is task management. The purpose of task management is to attempt to achieve better yard service performance such as connection performance. In this section, the decision processes used by trainmasters are used to represent the overall task management process. The trainmaster's work in a shift from planning the shift work to the implementation of his plan can be represented by figure 5.

Using the available information as inputs, the trainmaster develops a plan for the shift, using a behavioral or anecdotal model. Here, the model is not a computer or analytic model. There may not be a clear relationship between inputs and outputs in the model. It is the trainmaster's simple rules such as fixed cut off time for connections, first in-first out for the operations, and so on, plus his experience. The trainmaster's tool is very simple, heuristic, and based on his experience. Different trainmasters may have different models and may generate different plans, which may have different yard service performances.

At the beginning of a shift, the trainmaster will collect the following information as his input in his "model".

Current traffic situation at the yard: how many cars in the yard, where these cars are located (e.g., receiving tracks, bowl or departure tracks), what are their destinations and priority, if the yard is too congested and so on. This kind of information provides traffic basis for the whole shift.

ETAs and ETDs information and other important issues are available from system level managers and the various information systems. This information includes the number of inbound trains expected during the shift, arrival time of each inbound train, consist of each inbound train, traffic priority of inbound trains; number of outbound trains, predicted departure time of each outbound train, outbound traffic requirements; and other important issues from the system level manager such as special traffic requirements. To sum up, this information provides the trainmaster the incoming traffic in the shift, departure requirements, and other special requirements from the system level managers. The trainmaster's work is to make connections of cars from current traffic at the yard, the incoming traffic in the shift to the outbound trains.

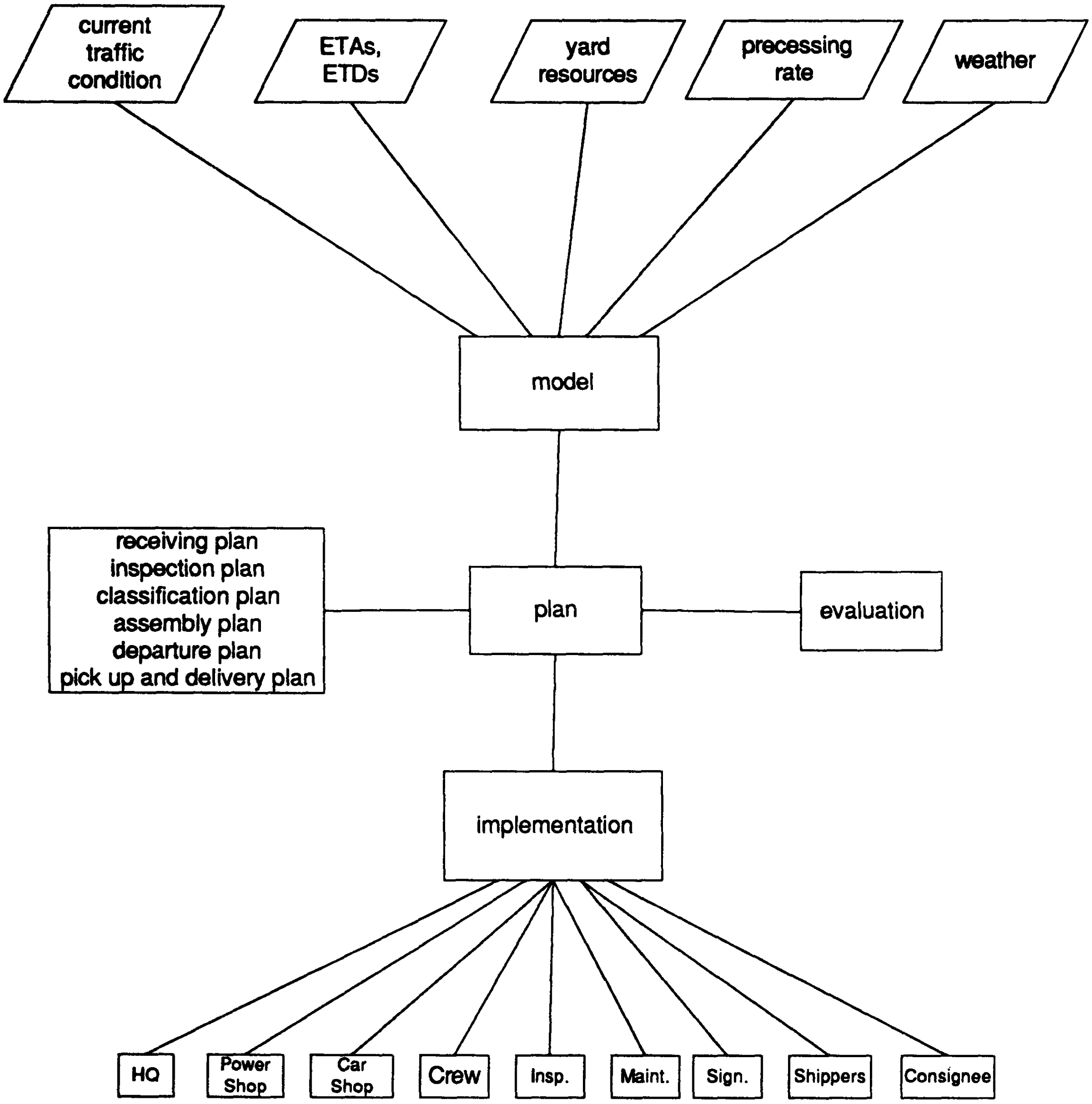

Figure 5. Trainmaster's Work in a Shift

Yard resource information: this information includes number of inspectors, switching crews in the shift; yard layout information such as number of receiving tracks, length of each receiving track, number of tracks in the bowl, length of each track at the bowl, number of tracks at the departure yard, length of each track at the departure yard, number of leads, and the interrelationship of receiving yard, hump, bowl, and departure yard; number of switching engines in the shift; number of other workers in the shift; the car shop facility and power shop facility and so on. This information provides resource constraints in the yard operations in the shift.

Processing rate information: includes what are the typical times to arrive, inspect, and hump a typical inbound train, to assemble, inspect, and depart a typical outbound train. The trainmaster may use his experience to decide these times in his work planning.

Other information such as weather condition in the shift: if the weather is not very good e.g. snowing, raining and so on, it may have adverse effect on the yard work in the shift. Other information includes any unusual track maintenance activities.

Using the trainmaster's "model", the output is the plan of how to do the work in the shift. The plan may not be a detailed plan. It may not include the complete list of tasks to be performed in the shift. At the beginning of the shift, the trainmaster may not decide all tasks in the shift. Also, for a specific task the trainmaster or yardmaster plans to do, there is usually no time interval for the task. The trainmaster or yardmaster directs the switching crew or inspection team to do the task, monitoring their performance, and after the task is completed, gives them another task. In other words, the instructions from terminal managers to the yard crews or workers are "one step ahead".

Since the trainmaster does not have a detailed plan, it is not possible to predict the connection performance from the plan. There is also no measure of the plan's feasibility. The plan is based on trainmaster's experience; he may not have alternative plans nor a method of evaluating the plan. As can be seen, the effectiveness of the plan depends greatly on the experience, wisdom, and judgment of the trainmaster and yardmasters. In cases where there is a great deal of knowledge or where the circumstances are "normal", this may be sufficient to result in acceptable performance of the yard. Where experience is lacking or conditions are unusual, good performance is likely to deteriorate rapidly. There may also be no detailed specific plan for each process such as a receiving plan, inspection plan (inbound and outbound), hump or classification plan, assembly plan, departure plan, and pick up and delivery plan.

After having some idea of how to do the work in the shift, the trainmaster then moves to the implementation of his plan. In this stage, the trainmaster may contact different departments and levels of the organization such as headquarters (system level managers), locomotive shop, yard car shop, yard switching crew, yard inspection team, track department, yard signal department, shippers, consignees, and so on to attempt to achieve better yard performance[7]. For example, the trainmaster may contact headquarters to get instructions about any special circumstances in terms of traffic operations. Also, if the trainmaster believes that certain tasks may not be accomplished that the system level manager expects the yard to perform, he may contact the system level manager to explore alternatives. The yardmaster may contact the yard power shop to get the information regarding when a specific road engine is going to be available for the next line-haul task. The trainmaster may have a clerk contact the car shop to confirm the time the bad order cars can be repaired and available for the connections to the outbound trains. The trainmaster may contact the track department to arrange a time window for maintenance activities if it is required.

During the whole shift, yardmasters act as the assistants of the trainmaster and are responsible for the planning and implementation for specific yard operations. For example, the hump yardmaster may be the assistant of the trainmaster in the planning and implementation of receiving traffic operations including arrival of inbound trains, inbound inspection, and hump. The bowl yardmaster assists in the planning and implementation of

---

[7] In many cases, this contact will be through subordinates such as the yardmasters or clerks.

departure traffic operations including assembly, outbound inspection, and departure operations.

The yard managers first give a few tasks to the switching crew and inspection teams at the beginning of the shift. Then the yard managers make many decisions based on the available information and the current performance of the tasks given to the crew and inspection teams. The major purpose of the yard operations is to make connections of cars from inbound trains to the outbound trains. So the major concern of trainmaster should be the cars' making connections from inbound trains to the outbound trains. Although the normal concern of the trainmaster is with connection performance, he may focus on other related issues such as the timeliness of train departures or the management of yard congestion. The trainmaster may decide which traffic will make the connections in the shift. And then he may decide the processing sequence at each operation. The yardmaster is concerned with the detailed assignments of tasks to available resources such as inbound trains to receiving tracks, occupied tracks to inspection teams, inspected tracks to hump, available blocks or cars at the bowl tracks to outbound trains, assembled outbound trains to the departure tracks, assembled outbound trains to outbound inspection teams, and departure sequence if there are more than one outbound trains waiting for departure operation.

The detailed decisions made by the yard managers in each operation are given using Table 1.

The trainmaster may also need to make decisions to coordinate the hump and assembly work because there is some degree of conflicts between hump work and assembly work in the yard. The degree of conflicts will be determined by the detailed configuration of the yard.

To summarize the state of the practice, there are three characteristics. The first one is *very simple decision rules* such as First In - First Out (FIFO) for yard operations and fixed cut off time for making connections of cars from inbound trains to outbound trains.

Table 1: Decisions Made in Terminal Operations

Process	Decisions Made	Decision Maker	Performer
IB Arrival	which track to yard the inbound train	hump yardmaster	road train crew
	is the length of the track long enough to hold all the cars in the inbound train? need double over?	hump yardmaster	road train crew
	arrangements of receiving tracks for the successive inbound trains	hump yardmaster	--------
IB Inspection	whether or not to inspect an IB train	trainmaster	--------
	assign occupied receiving tracks to the inspection teams including sequence, time, lunch hour, work load, and so on	hump yardmaster	IB inspectors
Classification	hump sequence, time, work load, lunch hour	hump yardmaster	hump crew
Assembly	assembly sequence; leads, throats and tracks used; number of cars and blocks in each outbound train; work load	bowl yardmaster	assembly crew
OB Inspection	assign occupied departure tracks to the OB inspection teams including sequence, time, lunch hour, work load, and so on	bowl yardmaster	OB inspectors
	send out power from engine house and attach the power to the OB train	trainmaster	hostlers
OB Departure	the sequence of departures	system manager	---------
	call train crew for departure	system manager	crew callers
	signal times	line dispatcher	dispatchers
	departure	line dispatcher	road crew

The second one is *only sequence, no time requirements* for each task. The terminal managers usually can only give the sequence of operations in each process and they usually do not specify the time required for each task in each process. The major reason for no time requirements may be that there exists big variations in the times required for same or similar tasks and the time required for each task is not well understood. This results in the following characteristic.

The third one is "*one step ahead*", which means that each time the yard managers give only very few tasks to the switching crew and inspection teams (usually one to three tasks at a time). The reasons for this may be due to the following: first, the time required for each task is not well known; second, if the time were known, the benefit of giving all the tasks at the beginning of the shift is not well known. That is, there is no a model to justify "all steps ahead"; third, the division of authority and direction makes it difficult to give time requirements for all the tasks. When the crew and inspection team are doing the current task, they only know one or two tasks ahead. This situation may not give the workers the complete list of tasks for the shift.

These three characteristics may be part of the difficulties in the improvement of yard service performance. Later in this chapter, a methodology to address improvement of terminal service performance is presented, which may overcome these disadvantages and lead to better yard service performance.

2.4 Literature Review of Terminal Studies

Generally speaking, there are three levels of terminal and related studies: terminal specific operation studies, terminal performance studies, and terminal related system level studies. The terminal specific operation studies focus on one or two specific process in the terminal such as hump operation or assembly operation. Terminal performance studies focus on the terminal aggregate level performance. They do not consider how the individual processes are put together to achieve better performance. The terminal related system level studies treat the terminal as a "node" in the overall railroad network. They do not consider the yard operations to get reasonable values for the yard processing times. The previous specific operation studies, terminal performance studies, and the terminal related system level studies are reviewed in this order.

2.4.1 Terminal Specific Operation Studies

Most previous specific operation studies focus on hump and assembly processes. An early study was conducted by Siddiqee [1972]. In his paper, Siddiqee analyzes the required assembly work based on four sorting strategies named initial grouping according to subscript, initial grouping according to outbound trains, triangular scheme, and geometrical scheme. He analyzes the hump work, the assembly work, and the number of tracks and lengths of the tracks required for the four strategies. The suitability and applicability of the four strategies are also discussed in the paper.

Daganzo et al. [1982] further analyze the first three strategies in Siddiqee [1972]. They name the strategies as sorting-by-block, sorting-by-train, and triangular sorting. These are actually multistage sorting strategies. That is, in the first stage, more than one block will be assigned to some bowl tracks. And these tracks must be resorted by either the classification engine (re-humping) or during the outbound train assembly stage so that cars and blocks are sequenced correctly at the time of departure. Using a probabilistic model, the authors derive processing time and expected number of switches per group, at first and second stage formulas for the three multistage strategies. Only approximation formulas are given for the triangular sorting strategy due to the complicated nature of the strategy. The authors find that triangular sorting, which allows many more classifications on a fixed number of tracks than either sorting-by-block or sorting-by-train strategy, does not require significantly greater number of expected switches in flat yards, even for fairly large numbers of blocks per outbound train.

There are a few other studies about the hump operation. One of these studies is that of Yagar et al. [1982]. In this paper, a model named HSS (Hump Sequencing System) about hump sequencing is presented. The authors use a time or cost minimization procedure in the HSS model because the authors think that the connection criteria tends to overestimate actual linkages (connections made). The processing times such as inspection time, classification time, assembly time, and departure time are assumed only related with the number of cars being handled. Thus given the yard layout, arrival and departure trains' information (ETAs and ETDs), blocking policy, and processing times, the hump sequence is determined using a dynamic programming approach. This approach is modified using a heuristic called "screening candidate trains" because of the high computational cost. A rank criteria which is the product of load factor and priority factor is used to select the inbound trains in the dynamic programming approach. The load factor is the proportion of cars in the inbound trains that are likely to make connections to the outbound trains and the priority factor measures the importance of cars in each inbound train. The authors conclude based on the results of several examples that the HSS model can get better result compared with first in first out (FIFO) algorithm and the actual performance based on the experience of yard managers.

The advantages of the specific operation studies are that these studies analyze the individual processes in great depth and give either formulas about the operation or the optimized results based on some strategies or assumptions. The disadvantage in terms of improving yard service performance is that these studies do not focus on the whole set of yard operations. One specific operation analysis and optimization are not generally enough to improve yard service performance. All the yard operations affect the yard performance. All the operations in the yard should be analyzed together to improve the yard performance. The specific operation studies, however, did provide insights to the operations addressed.

The insights gained from these specific operation studies can be incorporated in the modeling for the improvement of the yard service performance in the following chapter.

2.4.2 Aggregate Level Terminal Performance Studies

There are many empirical aggregate level terminal performance studies especially those conducted by Center for Transportation Studies of MIT. There are not many methodological studies. The reason may be due to the complexity of the yard operations and higher requirements for the understanding of the yard operations and processes in great depth and breadth. The two most important aggregate level performance studies are Petersen [1977, 1 and 2] and Martland [1982].

Petersen [1977, 1] divides the yard operations into five processes: receiving and inbound inspection, classification or sorting, waiting for connection, trains marshaling or assembly, and outbound inspection and departure. He also classifies yards into five types: simple yard, single-ended flat yard, double-ended flat yard, directional flat yard, and hump yard. He does not explicitly model the receiving and departure operations because his data showed that these two operations are not bottleneck operations. He focuses on the modeling of classification, assembly, and waiting for connection operations using queuing theory. Specifically, if the classification and assembly are independent, then the two operations can be modeled separately using multiserver queuing models. If these two operations are not independent such as in the single-ended flat yard, these two operations are modeled as an nonpremptive priority queue with unequal service rates. The assembly of outbound trains is assumed to have priority over the classification of inbound trains. The queue models used to model these two operations are M/G/s, if the operations are independent, and M(i)/G(i)/s or NPPR if the common yard resources are used, where M denotes a Poisson input, G denote a general service time distribution, and s is the number of service channels.

For the cars' waiting time for connection or connection delay time in the paper, the queue model used is a $M/E_k/1$ bulk queuing model.

Using these models and deterministic service time per train, the predicted distribution of put-through times or yard times by destination is obtained. The comparison with the observed data showed that the predicted distribution is very accurate.

The queue theory and models require some distributions of arrival of customers and service time. For example, in this study, the arrival and departure processes are assumed statistically independent and are Poisson processes. Also, the arrival from classification to the bowl is assumed Poisson process. The service time distribution could be exponential (M), Erlang of order k (E_k), or deterministic (D). The time between services is assumed being independent and distributed as an Erlang distribution of order k. For a rail yard operations, these assumptions may be strong. More effort is needed to investigate the distributions of arrival and service time before applying the queue theory in the yard operations.

The study assumes a fixed service time. In Peterson [1977, 2], the expected number of switches per cut and the classification and assembly time which are based on the expected number of switches and other parameters are derived. Using a probabilistic model, Peterson derives the formulas of expected number of switches first for insufficient classification tracks and multiple classification engines situations assuming uniform block size. The formulas of expected number of switches are also derived for unequal block size and any allocation of these blocks to the available bowl tracks.

The classification and assembly time are then derived based on the expected number of switches per cut, the average train length in cars, average number of cars per cut, and the standard times for a classification and assembly switch. Using the number of receiving tracks, bowl tracks, and departure tracks, the classification and assembly rates are modified based on the assumptions discussed in Peterson [1977, 1].

A key concern with aggregate queuing approaches such as that of Peterson is that they simply do not give the terminal manager any insight into how to improve the operation. At best, the model can give an accurate estimate of the yard's performance, but it is nearly impossible to translate this into an understanding of the consequences of other activities or task assignment.

The other key paper in aggregate performance studies is Martland [1982]. In this paper, Martland presents PMAKE analysis and models addressing the connection performance issue. A PMAKE function relates the probability of making a particular train connection to the time available to make that connection and a number of other independent variables such as the priority of the traffic, the traffic volume through the yard, the pattern of train volume through the yard, the pattern of train arrivals through the day, the reliability of train arrivals, and the availability of power. The PMAKE analysis extends the fixed cut off time to determine the car connections to a probabilistic connection standard. Also, the PMAKE function can incorporate, at least in theory, all the factors affecting the yard performance in the functional form. These factors include facility, resources, traffic volume and distribution, processing rates, and so on. Using the PMAKE function, it is easy to get the predicted yard connection performance and the change in connection performance if the situations are changed. The paper presents in detail the PMAKE functional forms, the calibration of PMAKE functions, and the standard set of PMAKE functions used by some railroads. Possible methods to enrich PMAKE functions are also addressed such as using econometrics method to incorporate more independent variables and looking into processing levels to make it possible to use the PMAKE function as a control tool in the daily yard operations.

The paper also presents a summary of both the practice of railroads and the academic studies in the yard performance area. The paper analyzes other approaches used in the yard performance studies. These approaches addressed in the paper include simulation, queuing theory, and capacity scheduling. The major weakness of these approaches are analyzed in depth. For example, the simulation model, though capable of simulating the detailed activities in the yard operations, has had little success in terms of adoption by railroads for ongoing use. The steady state and random arrival of inbound trains are two very strong assumptions which may not be satisfied for the rail terminals. The capacity scheduling models ignore the unreliability of service times and the resulting problems in predicting delays. The strength of PMAKE is presented compared with the weakness of these approaches. Using PMAKE functions, the predicted connection performance is close to the observed indicating the validity of the PMAKE approach.

The earlier development of PMAKE analysis is presented by Kerr, Martland, Sussman, and Philip [1976]. The process PMAKE concept and modeling approach is presented in Tykulsker [1981].

There are other terminal empirical studies. Rothberg et al. [1980] propose a terminal control system which uses flexible connection standards for reliability and average yard time, unit costs linking car time to car cost, and a volume variable car cost budget that is integrated with the operating conditions. Applying the methodology developed in Rothberg et al. [1980], Ferguson [1979] conducts a case study to demonstrate the usefulness of the system. Martland et al. [1983] further develop the system by incorporating budgeting techniques, probabilistic train connection standards, and microcomputer applications together in the system. A case study is also conducted in the research.

In a recent study by Duffy [1994], the statistical process control approach is used to analyze terminal performance and to identify causes of poor performance within the terminal. The results from a case study in this research show that tight connections and out of control processing times within these tight connections are the two major causes of missed connections in the terminal studied.

Many studies discussed above use probabilistic modeling approach. The major concern of these studies, especially Peterson [1977 1 and 2] and Martland [1982] is to predict connection performance, specifically, to predict the distribution of through-put time or yard time. In this thesis, the achievability concept is used to measure the terminal operating plan's achievable degree, which has probabilistic nature.

2.4.3 Terminal Related System Level Studies

Folk [1972, 1] analyzed data from several railroads on yard, link, and total origin-destination (O-D) performance. The data analyzed include time spent in the receiving yard, total yard times, train arrivals at yards, train departures from yards, train line haul times between yards, and trip times.

Reid et al. [1972] analyzed the relationship of car movement performance through terminals to transit time reliability. Using data from one hump yard and two flat yards, the study showed the causes of car delays in these terminals and the causal relationship between car movement performance and yard time parameters. Specifically, the findings of the study are: a substantial number of cars missed normal connections at the yards studied; many of these missed connections resulted from the cancellation of outbound trains; if outbound cancellations are discounted, the predominant car delays were due to late inbound arrival; finally there is a causal relationship between time available to make a connection and the probability of making that connection successfully.

Folk [1972, 2] used two simulation models, one a network simulation model and the other a single car (probabilistic) movement model, to study the effects of railroad operating policies and practices on trip time reliability. A major result of the study is that operating policies such as train dispatching criteria, the number of yards in a car's schedule routing, and the train frequency between yards have significant effects on a car's trip time reliability. Another major finding of the study is that the yards, rather than the links between the yards, are the major centers of unreliability. In Folk [1972, 3], some simulation and optimization models are reviewed.

Lang and Martland [1972] analyzed O-D trip time reliability by investigating the reliability issues of the components of the system: line haul reliability, classification yard reliability and network reliability. Martland [1974] and Sussman and Martland [1974] conduct a case study to investigate the relationship of O-D trip time reliability and the component reliability. The results of the case study verify the conclusions of the previous MIT research discussed in this section. The major findings of the case study are that reliability can be improved in the short run without major capital expenditures, and policies for improving reliability can reduce mean trip times and operating costs as well.

Using PMAKE analysis, McCarren et al. [1979], Martland et al. [1979], and Martland et al. [1983] developed a service planning model and conducted several case studies to demonstrate the applications of the model.

The terminal related system level studies provide insights to address terminal performance issues from the system level point of view. They do not, however, provide much insight that is useful to yard managers in planning daily operations in terms of task management.

Chapter 3. Terminal Operating Plan (TOP)

3.1 Framework

The major focus of this research is to improve terminal service performance by using the terminal operating plan (TOP) as an operational tool to manage and control terminal operations in one shift. The general framework of TOP is presented in Figure 6.

First, a detailed data base in terms of terminal processes is established. This data base includes detailed activities performed by each car or train in the terminal for a long enough time period. For example, each inbound train's arrival time, inspection time, and classification time and each outbound train's assembly time, inspection time, and departure time should be included in the data base. The number of cars, blocks, cuts in each inbound or outbound train, the number of switching engines, and the number of inspectors working at the shift should also be included. That is, the data base contains the major activity information in the terminal for a representative period of time. This activity information does not need to be collected after the study period, although regular updates are needed to capture changes in practices and procedures.

Another part of the data in the framework contains the current traffic and resource information in the terminal. For many terminals, this data is collected in various formal or informal systems such as yardmaster logs, shift working sheet and so on. From the data base, the insights of the processes could be obtained. In next section, the detailed data requirements are given.

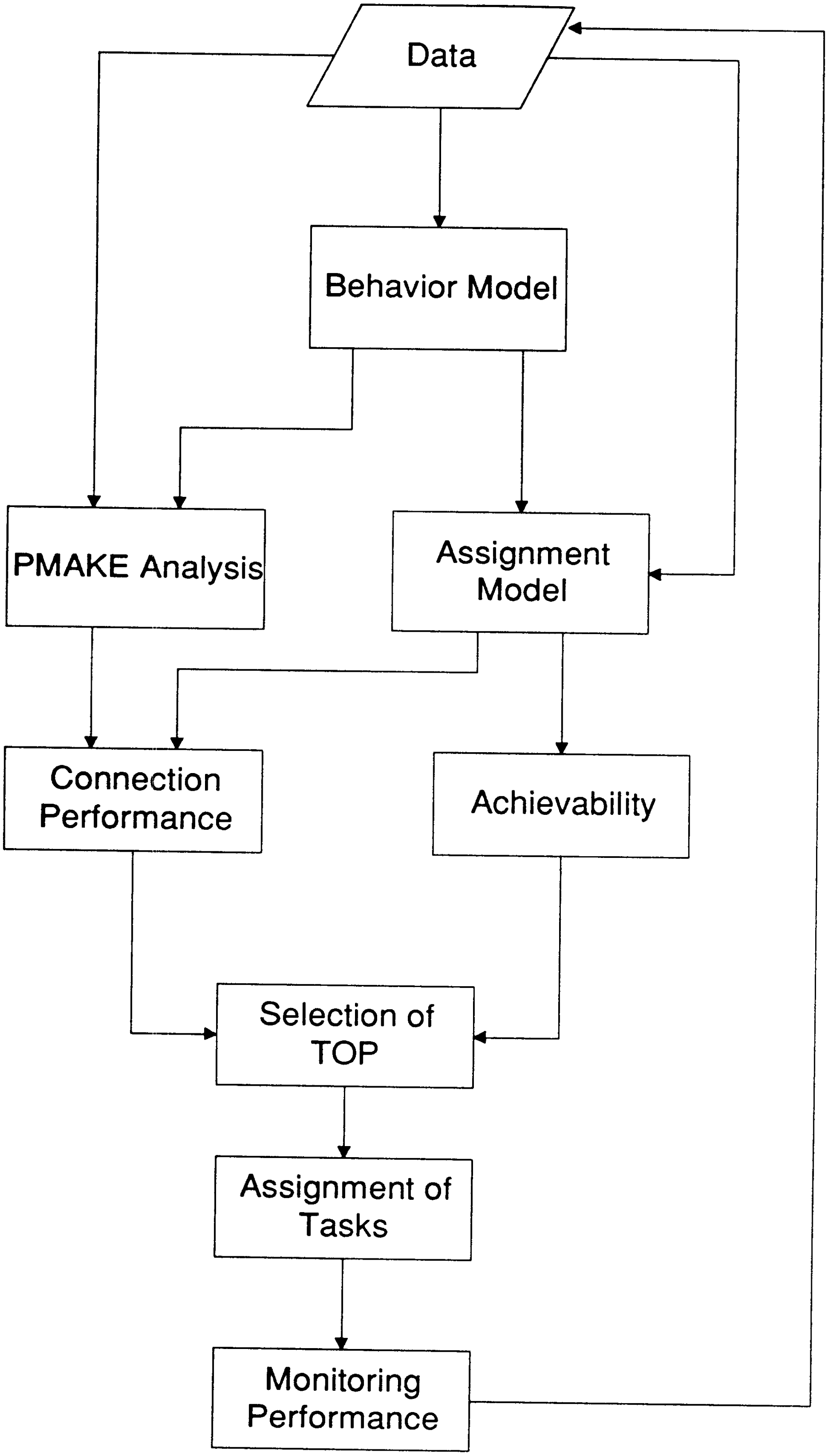

Figure 6. Framework

After a detailed data base discussed above is available, the next step is to analyze the data base and get some insights about the processes in the terminal. The process behavior models can be developed using factor analysis techniques. The purpose of this step is to attempt to understand the terminal operations in great depth. Here, the effort is made to reveal the "black box" of terminal operations. This is the *basis* for other analysis and modeling. The results from this step can also be used as benchmarking with other terminals' results. The analysis from this step may also reveal the major problems of the terminal operations and can be used in the improvement of the terminal operations.

The pure processing times are the major concerns of the process behavior models. The reason is that a basic requirement for a reliable terminal is the reliable pure processing times. The factors which are believed to affect the pure processing times can be searched out. Econometrics methods can be applied to find the effects of these factors on the pure processing times. The process behavior models can be used to get reasonable processing times for different conditions. The model results can also be used in better scheduling the tasks to achieve better service performance and as inputs of other models in the framework.

After terminal process behavior models are developed, there are two alternatives of addressing connection performance issue. One is an aggregate approach called real-time PMAKE analysis. The other one is to develop a detailed terminal operating plan. The real-time PMAKE analysis model predicts the best connection performance using real-time information. This model is based on the process PMAKE function developed by Tukulsker [1981]. From real-time PMAKE analysis, the projected best connection performance and the cars to make the connections to achieve the best performance can be obtained. The detailed task assignment is not given from the analysis. This approach can be used to predict aggregate connection performance and provide useful information to terminal managers to decide which cars make their most appropriate connections. Since many railroads have already applied PMAKE analysis technique in their daily terminal operations, this aggregate approach is easy to implement.

The other approach is to develop a detailed terminal operating plan. The current *de facto* plan is developed either using simple rules such as first in-first out for operations and fixed time for connection or using simple rules plus intelligent judgment based on training, experience and other "wisdom". In this thesis, an assignment model is proposed to create an "optimal" TOP which meets a specific objective function such as maximizing the number of cars making their first connections. The purpose of the assignment model is not to replace excellent terminal managers' experience and wisdom, but to make some of it available to less experienced terminal managers by a computer based tool. Using this detailed planning technique, the projected connection performance is available from the TOP. Also, the achievability of TOP, which is presented in detail in chapter four, can be computed. The TOP provides a detailed and complete picture of what will happen if the TOP is accomplished. As a planning result, the TOP gives detailed task management requirements to achieve better connection performance. The achievability measure gives the terminal managers, terminal crews and inspectors, and the system level managers the probabilities of accomplishing individual tasks, a set of tasks in each process, and the overall plan. As discussed earlier, the achievability measure is potentially very useful for these managers and workers in the terminal operations management and the improvement of terminal service performance.

From different achievability and connection performance values, the terminal manager may choose one TOP which satisfies his expectation in both connection performance and achievability measures. This selection process may be done by both system level manager, who may be more interested in higher connection performance, and terminal manager, who may be more interested in higher achievability measure.

After a TOP is selected, the requirements of the TOP for each crew (including inspectors) can be communicated to the terminal crews. According to the TOP, the terminal managers can assign the tasks to the crews in a manner that it is easy for the crews to accept these tasks and make an effort to accomplish these tasks. From the theory of Total

Quality Management (TQM) [Deming 1986], it is better to make the crews know all the tasks in the shift. But the reality may be different. The way the terminal managers communicate the terminal crews and assign the tasks to crews can be decided on a case by case basis.

During the implementation process, the terminal managers need to monitor actual performance. Based on the actual performance, the terminal managers need to make necessary decisions in the task management process. Also the actual performance should be recorded to the detailed data base for future analysis.

3.2 Definition of Terminal Operating Plan

In this section, the terminal operating plan is defined. The possibility of developing a detailed terminal operating plan from the available information at the beginning of each shift is then addressed.

3.2.1 Definition of Terminal Operating Plan (TOP)

As discussed in chapter two, the current terminal plan is heuristic and based on the terminal managers' experience. There is no complete list of tasks to be performed in the shift in the plan. Also, there are no time requirements for each task in the plan. From the current plan, the connection performance can not be predicted if the plan is accomplished. Compared with the current terminal plan, the terminal operating plan in this thesis is defined as follows:

Terminal Operating Plan (TOP) is a set of tasks in a shift with an explicitly stated start time and end time for each task to be performed in the shift to achieve better yard service performance. The key problem in the terminal operations is the task management. The system operating plan's performance can only be realized by assigning the tasks to available resources in the terminal. A realization of assigning the tasks to available resources in the terminal in planning period (such as a shift) is a TOP.

Compared with current yard plan, the TOP defined above has three distinct characteristics. The first one is that for each task, a start time and end time is explicitly stated. This is required in order to manage yard service performance and to control yard resources. The current plan does not impose time requirements for each task, which may indicate that not enough attention is paid to increase reliability of pure processing time. Here the *pure processing time* is defined as the time interval from the beginning of the task to the end of the task. For the same task, there will be some variation in the time needed to finish the task. But a reliable yard requires reliable operations, especially reliable pure processing times. That is, for a reliable yard, the pure processing time should have small variation and the expected time to perform tasks should be accordingly predictable. So we may specify a time interval for a task in the TOP in such a way that it is likely to complete the task in the time interval and the time interval is not too long.

The second distinction is that when developing the TOP, the yard service performance, i.e. connection performance, is explicitly considered. That is, the connection performance is used as an objective of the TOP. After the time interval for each task is specified, the TOP can be developed in such a way that the number of cars making their most appropriate connections is maximized.

The third characteristic is that the TOP developed at the beginning of the shift can give the yard manager a complete picture of what is expected to happen in the yard during the shift. The TOP provides a list of tasks that are going to be done in the shift. Each task in the TOP has a time window to perform the task. From the TOP, all the tasks can be divided into each of the six major processes. The crew who is responsible for a specific process such as classification will know his work in the shift clearly. At the beginning of the shift, the crew will know not only the tasks they will do in the shift but also when each task must be done in order to achieve high connection performance.

Since the TOP as defined here provides a clear picture of the yard operations in a shift, we can calculate, from the TOP, the predicted connection performance, the projected detailed processes for each car in the yard, and the probability of successfully performing the TOP.

3.2.2 Possibility of Developing TOP

From the current available information at the beginning of each shift, it is possible to develop TOP for the yard operations. Figure 7 shows the interaction between system level managers and yard managers and the relationship between the system level operating plan and the terminal operating plan, which result in much of the information available to the yard.

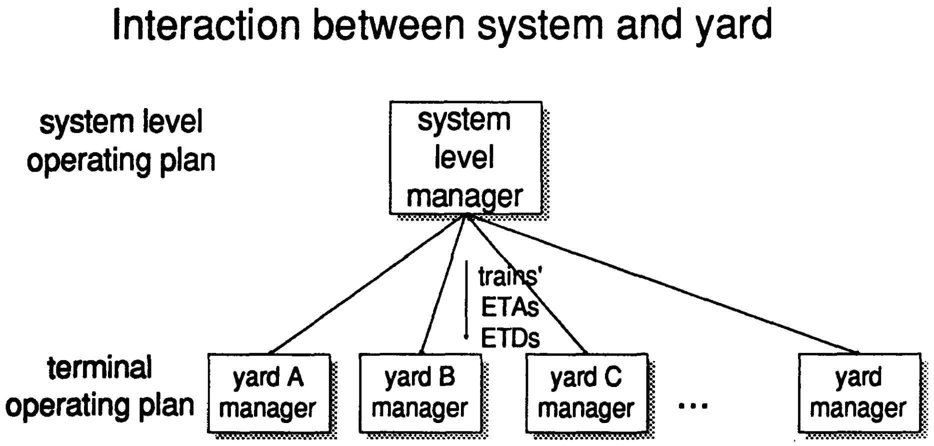

Figure 7. Terminal Available Information

Before the beginning of each shift, the system level manager will give each yard incoming traffic information (estimated time of arrival (ETA)) and outbound departure requirements (estimated time of departure (ETD)). The ETA information includes the number of inbound trains, arrival time for each inbound train, the consist of each inbound train including sequence of the cars in the train and the priority of each car. The ETD information includes number of outbound trains, departure time requirement, traffic volume requirement for each outbound train and so on. The system level manager's decisions are based on the system level operating plan such as train schedule, car scheduling, power and crew plans and so on, adjustment to the plan from the current performance and situation of the system such as line haul performance, terminal performance and even weather condition and so on.

The yard resources information is also available to the yard manager before the shift. For example, the yard manager will know how many inspectors, switching crew and engines, clerks, and so on in the shift.

Before each shift, the inbound traffic information, outbound departure requirements, and yard resources information is available to the yard manger. It is possible to develop TOP which assigns tasks in the shift to the available resources in the yard to satisfy departure requirements as much as possible. TOP is actually a realization of assigning the tasks to the available resources in the yard to maximize connection performance constrained by the departure requirements such as departure time and volume requirements.

Table 2 presents a simple terminal operating plan.

In this example, only four processes, IB inspection, classification, assembly, and OB inspection, are explicitly considered. There are four blocks in the terminal named block a, b, c, and d. The IB arrival and OB departure processes are considered as part of IB inspection and OB inspection respectively. This table shows that at the beginning of the planning period (0800-1600), there is an inspected train (inventory) waiting for classification operation. This train contains 25 cars of block a, 10 cars of block b, 35 cars of block c, and 20 cars of block d. The planned classification time for this train is from 0820 to 0900 (40 minutes). At the end of the shift, the planned inventory in the terminal is 15 cars of block a, 20 cars of block b, 28 cars of block c, and 7 cars of block d.

Before presenting the framework of TOP, a clarification of the difference and relationship between system level operating plan and terminal operating plan may be useful. Most railroads are concerned with the system level operating plan while leaving the terminal operating plan to the yard manager's discretion. Generally speaking, the system level operating plan is a method to move traffic through the whole system while the terminal operating plan is about how to do the yard work. System level managers may not really know or care how resources are used in yards as long as system level goals are met. Terminal managers are very concerned with how resources are assigned or used because this is the only way they can realize their target goals.

Table 2: A Terminal Operating Plan:

Train ID	Arrival Time	Block(a, b, c, d)	IB Inspection Time	Classification Time	Assembly Time	OB Inspection Time
Inventory		25, 10, 35, 20	(inspected)	0820-0900		
IB #1	0830	30, 15, 33, 12	0830-0900	0905-0945		
IB #3	0900	10, 10, 40, 30	0900-0935	0945-1030		
IB #2	1000	15, 20, 20, 45	1005-1040	1215-1255		
OB #1		65, 35, 0, 0			1030-1120	1125-1205
OB #2		0, 0, 100, 0			1125-1215	1215-1252
OB #3		0, 0, 0, 100			1255-1345	1345-1430
Inventory		15, 20, 28, 7				

System operating plans for terminals define train connection performance such as scheduling cars that arrive from train 1 to depart on train 2. But these assignments may not always be possible. For example, late arrival of inbound trains, canceled outbound trains, conflicting resource demands in the terminal, engineering problems in the terminal and others may all cause cars to miss their connections. So the system operating plan assumes some level of resources in the terminal and sets probabilistic or other standards for terminal performance. Terminal managers must assign specific tasks to the available resources in a specific order or sequence to attempt to achieve system specified goals.

3.3 Data Requirements

The approach presented needs detailed terminal operations data. Note, as discussed in the framework section, this data is for defining and understanding terminal processes and is collected for a certain period of time, not necessarily for every shift. In order to understand the terminal operations in detail, it is necessary to have a detailed data base to record the detailed activities at the terminal. The data should include activity information of cars or trains, resource utilization information, and other related information. *Terminal operations activity information* includes detailed activities of cars, cuts, blocks, or trains in terms of *when* and *where* these activities happened over a sufficient period of time. *Resource utilization information* includes *who* perform the activities. The detailed information includes:

(1). Each inbound train's arrival time (t1). The arrival time is defined as the time the inbound train arrives at the entrance to the terminal. The time the inbound train arrives at the receiving track and the locomotives are removed (t2). Which track in the receiving yard the inbound train is yarded on. If one part of the train sits at one track and the other part of the train sits at another track, this receiving process is called double over. For double over process, the two tracks and the number of cars in the train at each track should be recorded. The difference between t2 and t1 is this inbound train's arrival processing time.

(2). The time each inbound train is given by the transportation department to the car inspectors (t3). For a specific inbound train, the difference of t3 and t2 is the waiting time of this train for giving to inspectors. This waiting time is of the responsibility of the transportation department. The inspection flag on time (t4). The flag on time is the time when the inspectors put a safety device (blue flag or blue light) on the receiving track to prevent the movement of the cars on the track. The difference of t4 and t3 is the waiting time and some time preparing for inspection flag on. This time interval is the responsibility of the inspectors. The beginning time of inspection (t5). This is the beginning time for the inspectors to inspect the first car at the track. The end time of inspection (t6). The flag off time (t7). The release time by inspectors (t8). This is the time the inspectors finish the inspection of the track and notify terminal managers of the availability of the cars in this track for other activities. The difference of t6 and t5 is the pure inspection time not including flag on and flag off time. The difference of t7 and t4 is the pure inspection time including flag on and flag off time. The difference of t8 and t7 is the waiting time for giving back to terminal managers. This waiting time is the responsibility of the mechanical department. For each inspection process, the number of inspectors and their name should also be recorded. In many cases, only some of these times will be recorded. If the times between activities are short or are not highly variable (e.g., the time between t4 and t5), it may be sufficient to record only some of the times. It is generally more important to that consistent information be captured than that complete information be collected.

(3). The time each inbound train begins to be classified (t9). This time is defined as the time the first car is at the crest of the hump. The end time of classification (t10). This is the time the last car is going over the crest of the hump. The difference of t10 and t9 is the pure classification time. During the classification process, if trim work happens, when and how long the trim work takes should also be recorded. The engine and crew doing the classification work should be recorded.

(4). The beginning time of each outbound train's assembly process (t11). The number of blocks and cars (including car identification number) in each outbound train and the leads, tracks, and throats used in the assembly operation should also be included in the data base[8]. The end time of the assembly process (t12). The difference of t12 and t11 is the pure assembly processing time. Since the car identification number is used here, the waiting time each car spends in the bowl can be obtained from the data base.

(5). The time given to outbound inspectors (t13). Outbound inspection flag on time (t14). Beginning inspection time (t15). End time of the inspection (t16). Flag off time (t17). Released time by inspectors (t18). The difference of t13 and t12 is the waiting time for giving to inspectors. The difference of t14 and t13 is the waiting time plus the time required for preparation of outbound inspection. The difference of t16 and t15 is the pure outbound inspection time without flag on and flag off. The difference of t17 and t14 is the pure outbound inspection time with flag on and flag off time. The difference of t18 and t17 is the waiting time for giving to terminal managers. The time the road crew and engines arrive at the departure yard should also be recorded. For example, the time power on train (t19). Note that power may be attached to the train before outbound inspection.

(6). The time the signal is given (t20). The time the outbound train leaves the terminal (t21), which is defined as the outbound train moves toward the line segment from the departure yard of the terminal. The difference of t20 and t19 is the waiting time for departure process. The difference of t20 and t19 is the pure departure processing time.

The time intervals discussed above can be shown using the following figure:

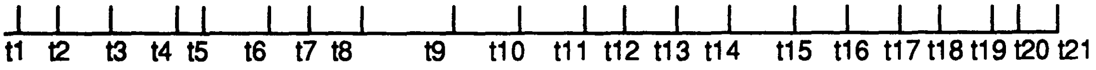

Figure 8. Terminal Detailed Activity Times

---

[8] For most railroads, the outbound consist information will include detailed car information. This can be accessed, along with blocking plans to determine some of this.

3.4 Process Behavior Model

After a data base is available, the process behavior models can be developed using econometric methods.

First, the factors affecting the pure processing times can be searched out. For example, the inspection processing time is affected by number of inspectors, number of cars, and number of bad order cars. The classification processing time is affected by number of cars, cuts, and the speed of classification engine. These kind of factors can be specified and the pure processing times can be explained by these factors.

Using econometrics methods such as regression analysis, process behavior models can be developed. For example, within a hump yard, the hump pure processing time (T) might have the following functional form:

$$T = 15.05 + 0.34^* \text{ (number of cars)} + 17.75^* \text{ (trim dummy variable)}$$

which means that there are about 15 minutes associated with each hump task (intercept term), humping each car needs about 0.34 minutes, or three cars are humped each minute, and if trim work is needed, an additional 18 minutes will be required to complete humping.

Such models reveal the relationship between pure processing times and the factors affecting the pure processing times. From the values of the factors, the pure processing times can be predicted. For example, using the above hump functional form, if a train with one hundred cars is to be humped and there is not trim work, the hump pure processing time for this train is about 49 minutes. The process behavior models can be used as inputs of other models in the approach and can also be used to better understand the processes in more detail.

The process behavior models can also be used to generate approximate times for use in development of a simple TOP. In the absence of computer-based operations research models such as that proposed in Section 3.6, these simple models may be very useful.

3.5 PMAKE analysis

This section discusses ways to predict train connections using real-time PMAKE analysis. Given ETA and ETD information and the current conditions of a terminal, train and block connections can be analyzed using train and block connection matrices. Taking into account tonnage or length constraints, which cars will be scheduled to make connections can be determined. Train connection performance can be predicted using a process PMAKE function under the assumption that processing times of the operations are normally distributed.

3.5.1 Connection Condition

As discussed in chapter two, a car from an inbound train will go through six processes before it can depart in an outbound train: arrival, inbound inspection, classification, assembly, outbound inspection, and departure. A car will usually make a connection if it is classified before the track is pulled for assembly and there is room on the outbound train for the car. That is, a necessary condition for a train connection is

$$\text{ETA} + t_i + t_{ii} + t_c <= \text{ETD} - t_a - t_{oi} - t_d \qquad (3.5.1)$$

where ETA and ETD are the expected arrival time of inbound train and expected departure time of the outbound train that car is scheduled to move on. t_i, t_{ii}, t_c, t_a, t_{oi}, and t_d are the arrival processing time, inbound inspection time, classification time, assembly time, outbound inspection time, and departure processing time, respectively. From (3.5.1), we can get:

$$t_i + t_{ii} + t_c + t_a + t_{oi} + t_d <= \text{ETD} - \text{ETA} \qquad (3.5.2)$$

which means that the available time of a car must be greater than or equal to the total processing time of the car if the car is to make the appropriate (first) connection.

Suppose that the processing times for arrival, inbound inspection, classification, assembly, outbound inspection, and departure are normally distributed with parameters μ_i, σ_i^2; μ_{ii}, σ_{ii}^2; μ_c, σ_c^2; μ_a, σ_a^2; μ_{oi}, σ_{oi}^2; and μ_d, σ_d^2 respectively and the processing times are all independent. Then the terminal (total) processing time is normally distributed with parameters μ and σ^2:

$$\mu = \mu_i + \mu_{ii} + \mu_c + \mu_a + \mu_{oi} + \mu_d$$

$$\sigma^2 = \sigma_i^2 + \sigma_{ii}^2 + \sigma_c^2 + \sigma_a^2 + \sigma_{oi}^2 + \sigma_d^2 \qquad (3.5.3)$$

Using the data from the data base, the component parameters can be estimated and hence the terminal processing parameters μ and σ^2 can be determined. The probability that condition (3.5.2) holds is:

$$\text{prob}(\mu <= \text{ETD - ETA}) = \text{prob}(\mu <= \text{AVAIL}) = \Phi((\text{AVAIL}-\mu)/\sigma) \qquad (3.5.4)$$

where $\Phi(\,)$ is the standard normal cumulative distribution function. AVAIL is the expected time available to make the connection. The process PMAKE function is of the form:

$$\text{PMAKE} = \text{PMAX}*\Phi((\text{AVAIL} - \mu)/\sigma) \qquad (3.5.5)$$

where PMAX is the maximum probability of making a connection, which will be adjusted later.

Note that this approach assumes that terminal managers assign the cars to be processed in a consistent and reasonably intelligent manner (i.e., are good at task management).

3.5.2 Predicting Train Connections Using Real-Time PMAKE Analysis

3.5.2.1 Connection Matrices

Take ETA, ETD and the terminal's blocking policy as given in a shift. Suppose during a shift, there are m trains arriving at the terminal and n trains departing from the terminal. The relationship between the inbound trains and outbound trains can be expressed as a_{ij} which means that there a_{ij} cars from the i-th inbound train which are "suitable" to depart in the j-th outbound train. By suitable, it is meant that these cars can be scheduled to make the connection from the i-th inbound train to the j-th outbound train. Since some cars to the same destination may have different priorities, the priority is denoted as k. So a_{ij} is referred to as a train connection and a_{ijk} as a block connection. Here a_{ijk} is the number of cars from the i-th inbound train to block k of the j-th outbound train. So:

$$\sum_k a_{ijk} = a_{ij} \qquad (3.5.6)$$

$$\sum_i a_{ijk} = \text{number of cars in block k of the j-th outbound train} \qquad (3.5.7)$$

Suppose also that the inventories of blocks in the terminal at the beginning of the shift can be expressed as a_{0jk} which means there are a_{0jk} cars in the inventory of the terminal to block k of the j-th outbound train. Similarly:

$$\sum_k a_{0jk} = a_{0j} \qquad (3.5.8)$$

and a_{0j} is referred to as the inventory of the j-th outbound train. The relationship between inbound block and outbound block connections (including the inventories of the blocks in the terminal) can be expressed as a matrix (A_b) as following:

$$A_b=[a_{ijk}]_{i=1,2,\dots,m\ \ j=1,2,\dots,n}=\begin{bmatrix} a_{011}\cdots a_{01k_1}\ a_{021}\cdots a_{02k_2}\ a_{031}\cdots a_{03k_3}\cdots\cdots\ a_{0n1}\cdots a_{0nk_n} \\ a_{111}\cdots a_{11k_1}\ a_{121}\cdots a_{12k2}\ a_{131}\cdots a_{13k_3}\cdots\cdots\ a_{1n1}\cdots a_{1nk_n} \\ a_{211}\cdots a_{21k_1}\ a_{221}\cdots a_{22k_2}\ a_{231}\cdots a_{23k_3}\cdots\cdots\ a_{2n1}\cdots a_{2nk_n} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m11}\cdots a_{m1k_1}\ a_{m21}\cdots a_{m2k_2}\ a_{m31}\cdots a_{m3k_3}\cdots\cdots a_{mn1}\cdots a_{mnk_n} \end{bmatrix}$$

(3.5.9)

Here, the subscripts k_1, k_2, ..., k_n denote the number of blocks in the first, second, ..., n-th outbound trains.

If an inbound block can be in either of two (or more) outbound trains, an assignment rule is needed. For example, the block can be assigned to the earliest feasible outbound train (i.e., the earliest outbound train which has an available time greater than the average processing time in the terminal).

Each row in the A_b matrix, except the first row (which is the total number of cars in the terminal at the beginning of a shift), is an inbound train. There is a row for each expected inbound train to arrive during the shift. Similarly, each column in A_b matrix contains a block of an outbound train and there are columns to represent all the outbound blocks on trains for the shift.

Similarly, the relationship between inbound train and outbound train connections (including the inventories of the blocks of outbound trains in the terminal) can be expressed as a matrix A_t as following:

$$A_t=[a_{ij}]_{i=1,2,\dots,m\ \ j=1,2,\dots,n}=\begin{bmatrix} a_{01}\ a_{02}\ a_{03}\ \cdots\cdots\ a_{0n} \\ a_{11}\ a_{12}\ a_{13}\ \cdots\cdots\ a_{1n} \\ a_{21}\ a_{22}\ a_{23}\ \cdots\cdots\ a_{2n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}\ a_{m2}\ a_{m3}\ \cdots\cdots\ a_{mn} \end{bmatrix} \qquad (3.5.10)$$

The A_t matrix is similar to A_b matrix except that each column in the A_t matrix corresponds to an outbound train rather than a block of an outbound train.

The available times of the connections in the A_t matrix using ETAs and ETDs can be expressed as:

$$T=[t_{ij}]_{i=1,2,\ldots,m\ \ j=1,2,\ldots,n}=\begin{bmatrix} t_{011}\cdots t_{01k_1} & t_{021}\cdots t_{02k_2} & t_{031}\cdots t_{03k_3} & \cdots\cdots & t_{0n1}\cdots t_{0nk_n} \\ t_{11} & t_{12} & t_{13} & \cdots\cdots & t_{1n} \\ t_{21} & t_{22} & t_{23} & \cdots\cdots & t_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ t_{m1} & t_{m2} & t_{m3} & \cdots\cdots & t_{mn} \end{bmatrix} \tag{3.5.11}$$

Note that the available time is the same for all the blocks from the same inbound train to a given outbound train. So, there is no need of another matrix to express available time for every block, except for the first row, where the time is needed for each block connection. Here, the time for the first row needs some special attention. This row specifies the available time of the inventory in the terminal at the beginning of the planning shift to make the connections. Since some cars in the inventory may have finished some processes (for example, there are some inspected cars sitting at the receiving yard at the beginning of the shift), the probability of these cars making a specific connection will be larger than that of the cars which have not finished any process for the same connection. For the cars having finished some processes, the average processing time of completed processes is added to the actual available time to obtain available time in the row. Since for the same connection, some cars may have finished some processes, while other cars may have finished different processes or even none, more than one row may be needed in the "first row" in the above matrix to record the different states of the cars in the terminal at the beginning of the shift for the same connection.

Using ETAs, ETDs and other current information, PMAX can be adjusted. Using a PMAKE function (see equation (3.5.5)), the expected block connection reliability matrix P_b

and expected train connection reliability matrix P_t can be obtained. In applying equation (3.5.5) to obtain a PMAKE function, it is necessary to use current estimates of the means and variances of the processing times. If processing times can be precisely estimated based upon current conditions, the PMAKE function will resemble a cut off. Note, in block connection reliability matrix P_b, the element p_{ijk} is the probability of making the block connection a_{ijk} from the i-th inbound train to block k in the j-th outbound train. In train connection reliability matrix P_t, p_{ij} is the probability of making train connection a_{ij} from the i-th inbound train to the j-th outbound train using the following formula:

$$p_{ij} = \frac{\sum_k a_{ijk} * p_{ijk}}{\sum_k a_{ijk}} \qquad (3.5.12)$$

$$P_b = [p_{ijk}]_{i=1,2,\ldots,m \;\; j=1,2,\ldots,n} = \begin{bmatrix} p_{011}\cdots p_{01k_1} \; p_{021}\cdots p_{02k_2} \; p_{031}\cdots p_{03k_3} \cdots\cdots \; p_{0n1}\cdots p_{0nk_n} \\ p_{111}\cdots p_{11k_1} \; p_{121}\cdots p_{12k2} \; p_{131}\cdots p_{13k_3} \cdots\cdots \; p_{1n1}\cdots p_{1nk_n} \\ p_{211}\cdots p_{21k_1} \; p_{221}\cdots p_{22k_2} \; p_{231}\cdots p_{23k_3} \cdots\cdots \; p_{2n1}\cdots p_{2nk_n} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_{m11}\cdots p_{m1k_1} \; p_{m21}\cdots p_{m2k_2} \; p_{m31}\cdots p_{m3k_3} \cdots\cdots \; p_{mn1}\cdots p_{mnk_n} \end{bmatrix}$$

(3.5.13)

$$P_t = [p_{ij}]_{i=1,2,\ldots,m \;\; j=1,2,\ldots,n} = \begin{bmatrix} p_{01} & p_{02} & p_{03} & \cdots\cdots & p_{0n} \\ p_{11} & p_{12} & p_{13} & \cdots\cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots\cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots\cdots & \cdots \\ p_{m1} & p_{m2} & p_{m3} & \cdots\cdots & p_{mn} \end{bmatrix} \qquad (3.5.14)$$

The outbound trains' tonnage limits or length constraints can be expressed as a vector:

$$D = [\, d_{01} \sim d_1,\; d_{02} \sim d_2,\; d_{03} \sim d_3,\; \ldots\ldots,\; d_{0n} \sim d_n] \qquad (3.5.15)$$

Here, d_{01} is the minimum tonnage of the first outbound train and d_1 is the maximum tonnage of the first outbound train and so on.

The values of elements in the T, P_b, and P_t matrices are highly related to line performance, since the values of the elements in the above matrices depend upon the expected arrival and departure times. From these matrices, it is clear that the line haul performance affects terminal performance directly, because the line haul performance will determine the actual arrival times of inbound trains and hence the time available for connections to outbound trains. Given inbound arrival times, estimates of the time required for terminal processes will determine if outbound trains can be departed from the terminal on time and hence affect line haul performance directly. The interaction between line haul performance and terminal performance has been shown to be a major reason of unsatisfactory performance of the rail system. The matrices and PMAKE function can be used to link line control and terminal control together to predict, understand and achieve reliable performance of the system.

3.5.2.2 Determining PMAX

After the A (A_b and A_t), T and P (P_b and P_t) matrices are obtained, which cars will make connections taking into account tonnage constraints and possibility of cancellations can be determined and PMAX can be adjusted. The following procedure can be used to predict which cars will make their connections.

Create a block connection matrix C_b with the same dimensions as the A_b matrix. The blocks in this matrix will be scheduled to make their first connections considering the tonnage or length constraints and priorities of the blocks. From the A_b matrix, select the blocks whose probabilities of making the appropriate (first) connection are greater than or equal to 0.5 in the P_b matrix. (The value can be chosen according to the opinion of terminal managers or other experts). Put these blocks in the C_b matrix. These blocks will potentially constitute outbound trains.

Before allowing connections, tonnage or length constraints should be checked. If there is room for a car in an outbound train, the maximum probability (PMAX) for the car to make a connection is 1.00. If there is no room for a car in an outbound train, the PMAX will be 0, discounting bad order cars. In the C_b matrix, the blocks which are in the first row and have high priorities for each outbound train are checked first to see if the sum of the inventories of the blocks with high priorities are at or above the maximum limit. If the condition holds, the other blocks in the same column can be deleted. If the condition does not hold, other high priority blocks in the columns for each outbound train are checked from the top (earliest arrival) to the bottom to see if the sum of the high priority blocks (including the inventory blocks with high priorities) satisfies the tonnage limits. For example, check each outbound train if:

$$d_{0j} <= \sum_i \sum_k a_{ijk} <= d_j \qquad (3.5.16`$$

Here, a_{ijk} are high priority blocks for the j-th outbound train (including high priority blocks in the inventories). If the sum is at or above the maximum limit, the other low priority blocks are deleted. If the number is too small to meet the minimum tonnage limit, the low priority blocks are chosen from the top to the bottom to make the minimum tonnage limit hold. Note that this has an implicit priority assignment routine (high priority over older low priority). In practice, this may change.

If the sum of all the blocks in the columns for a given outbound train is too small to meet the tonnage limits, other high priority blocks or even low priority blocks if needed in A_b matrix with smaller probabilities of making first connection could be added in the C_b matrix to obtain the tonnage required to run a train. The outbound train can be held until the minimum tonnage is available. Here, the ETDs and car priorities are treated as given. If ETDs change, for example, the departure time of outbound train is delayed, the probability of holding the train will decrease due to the increase of available time for making the connection. Dispatching rules or algorithms can be added to reflect the actual situation. Also, the blocks can be added as needed to make the tonnage limit hold. By doing this, as many cars as possible can be departed in outbound trains. After each outbound train is checked, all the blocks in the C_b matrix are scheduled to be in the outbound trains.

3.5.2.3 Connection Reliability Performance

The connection reliability performance is defined as the weighted average probability that cars make the planned connections. The PMAKE function and the A (A_b and A_t), T and P (P_b and P_t) matrices can be used to estimate connection reliability performance. In this method, individual operations in the terminal are not considered. Only the available times of block or train connections are considered. For this application, it is necessary to calibrate a PMAKE function based upon past performance, for example, upon typical processing times (which may vary by day of week or time of day) and the typical precision of ETAs and ETDs.

For block connection reliability performance:

$$CR_b = \frac{\sum_i \sum_j \sum_k a_{ijk} * PMAKE(t_{ij})}{\sum_i \sum_j \sum_k a_{ijk}} = \frac{\sum_i \sum_j \sum_k a_{ijk} * p_{ijk}}{\sum_i \sum_j \sum_k a_{ijk}} \qquad (3.5.17)$$

and for train connection reliability performance:

$$CR_t = \frac{\sum_i \sum_j a_{ij} * PMAKE(t_{ij})}{\sum_i \sum_j a_{ij}} = \frac{\sum_i \sum_j a_{ij} * p_{ij}}{\sum_i \sum_j a_{ij}} \qquad (3.5.18)$$

Here, a_{ijk} ---- the elements in A_b matrix that are planned to make connections;

a_{ij} ---- the elements in A_t matrix that are planned to make connections;

t_{ij} ---- the elements in T matrix that correspond to the a_{ij} in A matrix;

PMAKE( ) ---- the process PMAKE function;

p_{ijk} ---- the elements in P_b matrix that correspond to the a_{ijk} in A_b matrix;

p_{ij} ---- the elements in P_t matrix that correspond to the a_{ij} in A_t matrix

From this method, if an inbound train arrives at the terminal earlier compared with the plan, the available time for the blocks in the inbound train will be larger and hence the probabilities of making first connections will be larger. Similarly, if an inbound train arrives at the terminal late, the available time for the blocks in the inbound train will be decreased and hence the probabilities of making first connections of the blocks will also be decreased.

3.5.3 An example

Suppose each outbound train has only one block and there is no initial inventory. Suppose further that based upon recent experience, mean processing time $\mu = 6$ hours, and standard deviation of processing time $\sigma = 2$ hours. The PMAKE function is:

$$\text{PMAKE(AVAIL)} = \Phi(\frac{AVAIL-6}{2})$$

The graph is shown below:

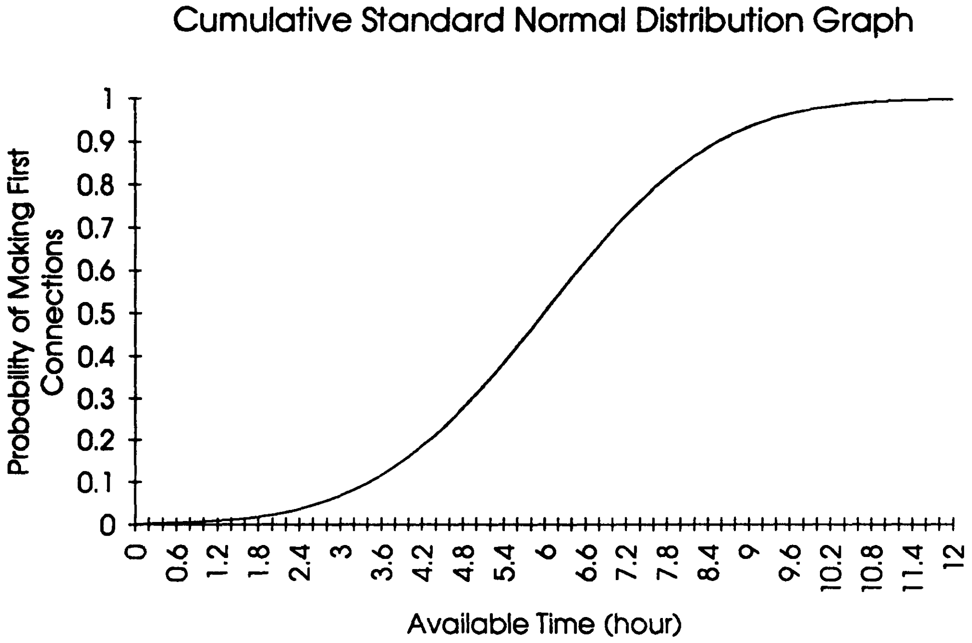

Figure 9. PMAKE Function

The arrival schedule and departure requirements for a shift are given as in Table 3 and 4.

Table 3: Arrival Schedule

ETA	IB ID	# of Cars	OB Train Connections (Number of Cars)
0130	123	60	184(15) 138(2) 185(43)
0200	186	60	135(30) 138(30)
0330	680	60	135(5) 184(50) 143(5)
0430	112	60	135(15) 138(3) 185(42)
0530	135	60	135(25) 230(30) 185(5)
0600	138	60	230(50) 138(5) 143(5)
0615	80	60	135(13) 138(47)
0730	160	60	230(10) 184(25) 143(25)
0745	143	60	135(2) 138(3) 143(55)

Table 4: Departure Requirements

ETD	OB ID	# of Cars	Connections From
1100	135	90	IB 135 143 112 680 80 186
1201	230	90	IB 135 138 160
1300	184	90	IB 680 123 160
1330	138	90	IB 138 80 112 143 186 123
1600	143	90	IB 143 160 680 138
1610	185	90	IB 135 112 123

The A, T and P matrices can be estimated as follows:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 15 & 2 & 0 & 43 \\ 30 & 0 & 0 & 30 & 0 & 0 \\ 5 & 0 & 50 & 0 & 5 & 0 \\ 15 & 0 & 0 & 3 & 0 & 42 \\ 25 & 30 & 0 & 0 & 0 & 5 \\ 0 & 50 & 0 & 5 & 5 & 0 \\ 13 & 0 & 0 & 47 & 0 & 0 \\ 0 & 10 & 25 & 0 & 25 & 0 \\ 2 & 0 & 0 & 3 & 55 & 0 \end{bmatrix}$$

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 11.5 & 12.0 & 0 & 14.67 \\ 9 & 0 & 0 & 11.5 & 0 & 0 \\ 7.5 & 0 & 9.5 & 0 & 12.5 & 0 \\ 6.5 & 0 & 0 & 9.0 & 0 & 11.67 \\ 5.5 & 6.5 & 0 & 0 & 0 & 10.67 \\ 0 & 6.0 & 0 & 7.5 & 10 & 0 \\ 4.75 & 0 & 0 & 9.25 & 0 & 0 \\ 0 & 4.50 & 5.50 & 0 & 8.5 & 0 \\ 3.25 & 0 & 0 & 5.75 & 8.42 & 0 \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 1.00 & 0 & 1.00 \\ 0.93 & 0 & 0 & 1.00 & 0 & 0 \\ 0.77 & 0 & 0.96 & 0 & 1.00 & 0 \\ 0.60 & 0 & 0 & 0.93 & 0 & 1.00 \\ 0.23 & 0.60 & 0 & 0 & 0 & 0.99 \\ 0 & 0.50 & 0 & 0.77 & 0.98 & 0 \\ 0.13 & 0 & 0 & 0.95 & 0 & 0 \\ 0 & 0.11 & 0.23 & 0 & 0.89 & 0 \\ 0.03 & 0 & 0 & 0.26 & 0.89 & 0 \end{bmatrix}$$

The expected connection reliability performance of the plan is:

$$\mathrm{CR} = \frac{\sum_i \sum_j a_{ij} * PMAKE(t_{ij})}{\sum_i \sum_j a_{ij}} = 416.22/540 = 77\%$$

If some inbound train or outbound train's schedule is changed, the adjusted connection performance can be estimated. Suppose inbound train 135 is going to arrive at the terminal one hour earlier or one hour late; the block connection probabilities will be different and the trains connection reliability performance and the percentage of cars missing their first connections will also change. The comparison showing the effects of a 1-hour delay or 1-hour early arrival for inbound train 135 is as follows:

Table 5. The Arrival Time Effects on Connection Performance

Train ID	Connection Probability (p)						Number of Cars	CR (%)	Miss (%) = 1- CR
IB 135	OB: 135	230	184	138	143	185	Making Conn.		
-1 hr	0.60	0.77	0	0	0	1.00	43	71.7	28.3
on time	0.23	0.60	0	0	0	0.99	29	48.3	51.7
+1 hr	0.11	0.23	0	0	0	0.96	14	23.3	76.7

3.6 Assignment Model

The purpose of this model is to generate a detailed terminal operating plan (TOP). A terminal operating plan is actually a realization of assigning tasks in the shift to the available resources in the terminal. The following preliminary assignment model can be used to generate a detailed TOP.

(1). Decision variables:

Indexes:

i ---- inbound train index (i = 1, 2, ..., N);

j ---- operation (1, IB receiving operation; 2, IB inspection; 3, classification; 4, assembly; 5, OB inspection; 6, departure);

k ---- outbound train index (k = 1, 2, ..., M);

t ---- time index (t = 1, 2, ..., T, T+1 for example, one shift is divided into T equal intervals and the T+1 interval is considered as a super sink to absorb unaccomplished activities)

Assignment variables:

$X_{ijt} \in \{0, 1\}$ which assigns (value of 1, otherwise 0) inbound train i for operation j at beginning of time interval t (j = {1, 2, 3});

$Y_{kjt} \in \{0, 1\}$ which assigns (value 1, otherwise 0) outbound train k for operation j at beginning of time interval t (j = {4, 5, 6});

Assignment time variables:

t_{ij} the beginning time of inbound train i for operation j;

t'_{kj} the beginning time of outbound train k for operation j;

Connection variables:

Z_{ik}, if outbound train k's assembly time is later than inbound train i's end time of classification, its value is 1, and 0 otherwise;

n_{kt}, the number of cars available for assembly at time t;

n_{k0}, the total number of cars available at the assembly time for outbound train k;

U_{ik}, if n_{k0} is within the traffic requirements for outbound train k, its value is 1, and 0 otherwise;

V_k, if n_{k0} is beyond the maximum traffic requirement for outbound train k, its value is 1, and 0 otherwise.

(2). Inputs:

For inbound train i:

arrival time (ETA), A_i;

number of cars, n_i;

number of cars for outbound train k, n_{ik}

$$(\sum_{k=1}^{M} n_{ik} = n_i);$$

and sequence of cars, from which the number of cuts, n_{ic} can be obtained;

Traffic priority:

in terms of car l, δ_{il} (1, low priority; 2, medium; 3, high priority);

in terms of train, δ_i (1, 2, 3 same as δ_{il});

For outbound train k:

departure time (ETD), D_k;

the earliest time before D_k when the outbound train k can be departed, t_{1k};

the latest time after D_k when the outbound train k can be departed, t_{2k};

minimum number of cars in outbound train k, $n_{\min k}$;

maximum number of cars in outbound train k, $n_{\max k}$;

Inventory information:

number of cars for outbound train k, n_{0k}

Processing time (here processing time is defined as "pure" processing time plus "reasonable" buffer time):

For inbound train i:

receiving processing time, P_{Ri};

IB inspection processing time, P_{IIi};

classification processing time, P_{Ci};

For outbound train k:

assembly processing time, P_{Ak};

OB inspection processing time, P_{OIk};

departure processing time, P_{Dk}

Resources:

number of receiving tracks, N_R;

number of IB inspection groups, N_{II};

number of assembly engines, N_A;

number of outbound inspection groups, N_{OI};

number of departure tracks, N_D

Connection set:

a_{ik} , which is 1 if inbound train i is to make connection to outbound train k from the system level operating plan, and 0 otherwise;

(3). Formulation

The objective function is chosen to maximize the number of cars making their first connections on the shift. There may be other objective functions such as minimizing operating cost or average yard time. As discussed in chapter two, the objective of maximizing the number of cars making their first connection is one of the most important ones to the improvement of terminal and system level service performance and has a positive effect on the improvement of operating cost and average yard time.

$$\max \sum_{i=1}^{N}\sum_{k=1}^{M} n_{ik} U_{ik} + \sum_{k=1}^{M} n_{\max k} V_k \tag{3.6.1}$$

s.t.

capacity constraints:

$$\sum_{i=1}^{N} X_{ijt} \le N_R \qquad j = 1;\ t = 1, 2, \ldots, T+1 \tag{3.6.2}$$

$$\sum_{i=1}^{N} X_{ijt} \le N_{II} \qquad j = 2;\ t = 1, 2, \ldots, T+1 \tag{3.6.3}$$

$$\sum_{i=1}^{N} X_{ijt} \le 1 \qquad j = 3;\ t = 1, 2, \ldots, T+1 \tag{3.6.4}$$

$$\sum_{k=1}^{M} Y_{kjt} \le N_A \qquad j = 4;\ t = 1, 2, \ldots, T+1 \tag{3.6.5}$$

$$\sum_{k=1}^{M} Y_{kjt} \le N_{OI} \qquad j = 5;\ t = 1, 2, \ldots, T+1 \tag{3.6.6}$$

$$\sum_{k=1}^{M} Y_{kjt} \le N_D \qquad j = 6;\ t = 1, 2, \ldots, T+1 \tag{3.6.7}$$

assignment constraints:

$$\sum_{t=1}^{T+1} X_{ijt} = 1 \qquad \forall\, i \in N,\ j \in \{1, 2, 3\} \tag{3.6.8}$$

$$\sum_{t=1}^{T+1} Y_{kjt} = 1 \qquad \forall\, k \in M,\ j \in \{4, 5, 6\} \tag{3.6.9}$$

operation sequence constraints:

$$t_{i1} \ge A_i \qquad \forall\, i \in N, \tag{3.6.10}$$

$$t_{i2} \ge t_{i1} + P_{Ri} \qquad \forall\, i \in N \tag{3.6.11}$$

$$t_{i3} \ge t_{i2} + P_{IIi} \qquad \forall\, i \in N \tag{3.6.12}$$

$$t'_{k5} \ge t'_{k4} + P_{Ak} \qquad \forall\, k \in M \tag{3.6.13}$$

$$t'_{k6} \geq t'_{k5} + P_{OIk} \qquad \forall\, k \in M \tag{3.6.14}$$

$$n_{kt} = n_{0k} + \sum_{i=1}^{N} \sum_{t(i)=1}^{t-P_{Ci}} n_{ik} X_{i3t(i)} \quad \forall\, k \in M, \forall\, t \in T+1 \tag{3.6.15}$$

$$Y_{k4t} = 0 \qquad if\ n_{kt} < n_{\min k} \ \forall\, k \in M, \forall\, t \in T+1 \tag{3.6.16}$$

possible connection variables:

Z_{ik}, its value is 1, if $t'_{k4} \geq t_{i3} + P_{Ci}$ and $a_{ik} = 1$; (3.6.17)

and 0, otherwise

possible connection volume variables:

$$n_{k0} = n_{0k} + \sum_{i=1}^{N} n_{ik} Z_{ik} \qquad \forall\, k \in M \tag{3.6.18}$$

connection variables:

U_{ik} , its value is 1, if $n_{\min k} \leq n_{k0} \leq n_{\max k}$; (3.6.19)

and 0, otherwise

V_k , its value is 1, if $n_{k0} > n_{\max k}$ (3.6.20)

and 0, otherwise

dispatching constraints:

$$D_k - t_{1k} \leq t'_{k6} \leq D_k + t_{2k} \qquad \forall\, k \in M \tag{3.6.21}$$

$$t_{ij} = \sum_{t=1}^{T+1} t * X_{ijt} \tag{3.6.22}$$

$$t'_{kj} = \sum_{t=1}^{T+1} t * Y_{kjt} \tag{3.6.23}$$

variable constraints:

$X_{ijt}, Y_{kjt}, Z_{ik}, U_{ik}, V_k \in \{0, 1\}$

$t_{ij}, t'_{kj}, n_{kt}, n_{k0} \geq 0$ and integer

The objective function (3.6.1) is to maximize the number of cars making their first connections. There are two terms in the objective function accounting two situations. One situation is that the available outbound traffic is within the requirements of the outbound train (first term). The other situation is that the available outbound traffic is beyond the maximum traffic requirement (second term). If there is not enough traffic for an outbound train, it can not enter the objective function.

Constraints from (3.6.2) to (3.6.7) are facility and resource constrains. The constraint (3.6.2) says that for inbound arrival process, the number of operations at any time in the shift can not exceed the receiving capacity in terms of number of receiving tracks. The constraint (3.6.3) says that at any time in the shift, the number of inbound inspections can not exceed the number of inbound inspection teams. The constraint (3.6.4) says that for a hump yard, the number of classification operations, at any time in the shift, can not exceed 1. For some terminals, there may be two or more humps. Under this situation, the figure here may be replaced by 2 or larger numbers depending on the number of humps and the hump operations in the terminals. Similarly, the constraints (3.6.5), (3.6.6), and (3.6.7) say that at any time in the shift, the number of assembly operations can not exceed the number of assembly engines, the number of outbound inspections can not exceed the number of outbound inspection teams, and the number of departure processes can not exceed the number of departure tracks, respectively. Here, some simplifying assumptions are made. For example, (3.6.5) says that an engine cannot assemble two trains in parallel. This is not completely accurate but is not unreasonable at this stage of modeling.

The constraints (3.6.8) and (3.6.9) are assignment constraints, which say that any task in any process is either assigned in this shift (from time interval 1 to T) or the successive shifts (the time interval T+1).

The constraints (3.6.10) to (3.6.16) are operation sequence constraints. Constraints (3.6.10) to (3.6.12) say that for any inbound train, the arrival process can be conducted only after the train arrives at the terminal, the inbound inspection process for the train can be conducted only after it finishes its arrival process, and the classification process can be conducted only after it finishes its inbound inspection process, respectively. Here, the possibility of avoiding inbound inspection process is not explicitly considered. For some urgent connections, the inbound inspections can be avoided and the corresponding constraints in (3.6.12) could be deleted. Constraints (3.6.13) to (3.6.14) say that for any outbound train, the outbound inspection can be conducted only after it finishes its assembly process, and the departure process can be conducted only after it finishes its outbound inspection process, respectively. Constraint (3.6.15) calculates the number of cars available at any time interval t in the shift for any outbound train, and constraint (3.6.16) says that for any outbound train at any time interval t in the shift, if the minimum traffic requirement is not satisfied, the outbound assembly process can not be started.

Constraints (3.6.17) to (3.6.20) formally state the connection variables Z_{ik}, n_{kt}, n_{k0}, U_{ik}, and V_k, respectively.

Constraint (3.6.21) says that for any outbound train, the departure time must satisfy the corresponding departure time requirements.

Constraints (3.6.22) to (3.6.23) are equations for obtaining assignment time variables from the corresponding assignment variables.

As the first stage, the formulation is given above. Before applying this model, some issues such as expressing the constraints in a form that a computer can process, and possible modifications to make the model correspond to the actual situation of specific terminals should be addressed. The formulation is only the beginning of the effort of using a computer-based tool to generate TOPs. More work should be done in the future. The purpose of this section is not to try to solve the problem, but simply to demonstrate that a useful formulation is possible.

Chapter 4. Achievability of the Terminal Operating Plan (TOP)

4.1 Introduction

Railroads have operating plans for their systems' operation which must maintain some service level to survive and develop in a fierce competitive environment. Similarly, each terminal has a *de facto* operating plan for the daily operations. A terminal operating plan may be useful for the improvement of terminal service performance, because it is a practical tool for the terminal manager to guide the terminal's work through a day or a shift. It specifies a set of tasks to attempt to achieve a reasonably good performance.

A terminal operating plan directs the terminal operations and directly affects the terminal performance. Sometimes the terminal operating plan can be accomplished. Sometimes it fails. What is the relationship between the TOP and performance? What is the feasibility of carrying out the operating plan to ensure some performance level? These questions have not been addressed explicitly. This may impede the application of terminal operating plans in the improvement of the terminal performance. In practice, as discussed earlier, the achievability of the terminal operating plan is very important for rail officers, especially the terminal managers, who must evaluate the feasibility of the operating plan.

As discussed in chapter two, terminal operations are complicated due to many factors affecting the terminal performance. It may be difficult or even impossible to get an optimal operating plan because of the probabilistic nature of task achievement in each process. In fact, an operating plan may be very good in terms of operating performance, but the achievability of the plan may be very low. This plan may be useless because it is unfeasible. If the uncontrollable factors, which are out of terminal control such as ETAs and ETDs, are very unfavorable for the terminal, a feasible plan may only achieve fairly poor performance but it is still a good plan. If a good terminal operating plan is available, the goal to achieve good terminal performance is transferred to realizing the plan. If the plan is accomplished, good performance will also be accomplished. From this point of view, the terminal operating plan is the core in the terminal operations. It is also the basis for the control and management of the operations in the terminal. The key issue is how to measure if a TOP is good or not. In this chapter, a new dimension to evaluate TOP is proposed, that is the achievability measure. In the following sections, the definition of the achievability is given, the usefulness of this measure is discussed, the methods to measure the achievability of TOP are presented, and how to choose a TOP from alternatives are discussed.

4.2 Definition of Achievability of TOP

As discussed earlier, a terminal operating plan is a set of tasks to be performed in a shift. The TOP should have good predicted service performance; on the other hand, the TOP should be achievable. That is, there should be a high likelihood to accomplish the tasks in the TOP.

Achievability of TOP is the probability that the TOP, the processes, and the tasks in TOP will be accomplished. There are three different levels to measure the achievability of a TOP. *Task level achievability* measures the probability each task can be accomplished within an allowed time interval. *Process level achievability* measures the probability that all the tasks in each operation (or process) can be accomplished in the shift. In chapter one, six processes in yards were introduced. *Plan level achievability* measures the probability that the whole TOP can be accomplished in the shift. All the three level achievabilities together are called achievability of TOP. In section 4.4.2, methods to calculate these probabilities are presented.

The achievability of TOP measures the feasibility or robustness of the TOP. Railroads have many system level plans such as train schedule, car scheduling, and crew and power plans. These plans sometimes or even often fail to be realized. While these failures may be due to many factors, the implementation of plans which are unlikely to succeed could be avoided if decision makers had explicitly considered their achievability.

When developing TOP, both the connection performance and the achievability of the TOP should be explicitly considered. Since TOP is a plan for the whole shift's work, it should have a high probability of being accomplished and should have good connection performance. Specifically, when developing TOP, the time interval for each task is first specified in such a way that the task is achievable within the time interval allowed. Then using the time interval for each task as constraints, the number of cars making their most appropriate connections is maximized. By doing this, the achievability of TOP and projected performance of TOP are combined together in a TOP. That is, the TOP developed has not only better projected connection performance but also higher achievability.

4. 3 Usefulness of Achievability Measure

The achievability measure developed in this chapter provides a means to measure an important dimension of terminal operating plan. First, achievability of TOP can be used by the yard manager to plan and control his work with confidence. As discussed later in this chapter, a time interval is specified for each task based on pure processing time distribution of the task, regression model of the pure processing time, or the yard manager's experience. For this specification process, the yard manager should feel confident that each individual task can be performed. After TOP is developed, the process level and overall plan level achievabilities can be used by the yard manager to evaluate each process and the whole TOP. If the yard manager is not satisfied by the process level and overall plan level achievabilities, he may change some or all of the time intervals of the tasks. And after the time intervals are changed, a new TOP can be developed. The yard manager can use achievability and projected connection performance measures to choose a TOP which has satisfactory achievability and projected connection performance. During this planning process, some kind of trade-off may be needed.

Second, achievability of TOP provides useful information for the yard crews and workers. For example, for a specific task, if the task level achievability is high, the crew or worker can do this task with confidence. On the other hand, if the achievability of the task is relatively low, the crew or worker may need to make more effort to accomplish the task in the required time interval.

Third, achievability of TOP can be used by system level managers to evaluate yards' work. If each yard has its own TOP, the system level manager can compare the TOPs of the yards and evaluate the yards' work. For example, for a specific yard, if the achievability of TOP is too small, the TOP may not be a good plan even though it has better projected connection performance. This is because the plan is not likely to be accomplished. On the other hand, if the achievability of the TOP is relatively high, but the projected connection performance is too low, the system level manager may advise the yard manager to look for a TOP with higher projected connection performance. It may be possible to sacrifice some achievability to get gains in projected connection performance. If a plan is not achieved, the consequence in connection performance can also be estimated. For example, if the plan is generated by the assignment model, a sensitivity analysis can be done to estimate the connection performance if some tasks are delayed. If the plan is developed manually, one way to estimate the connection performance is to recalculate the connection performance under the situation that some tasks are delayed and the successive tasks may also be affected.

Also, achievability of TOP can be used in choosing between different TOPs. For example, for a specific yard, alternative plans may be available and the plans may have the same or nearly same projected connection performance. Under this condition, the TOP with higher achievability may be a better plan. Even for different TOPs with different achievabilities and projected connection performances, the achievability measure can be used in the selection process. For example, consider the case where there are two TOPs, one with higher connection performance and lower achievability and the other with lower connection performance and higher achievability. The achievability of the TOPs could be a factor used to choose one of the plans.

4.4 Achievability Measurement

4.4.1 Analysis of Relationship between Terminal Processes and Achievability of TOP

A terminal operating plan specifies the tasks in each process within a shift. From the TOP, some cars will go through all the processes while others may only go through some processes before the shift ends. The achievability of operating plan is a function of all the processes in the terminal.

$$A = f(P_i, P_{ii}, P_c, P_a, P_{oi}, P_d) \qquad (3.7.1)$$

where, A is the achievability of operating plan;

P_i, P_{ii}, P_c, P_a, P_{oi}, P_d are the inbound arrival, inbound inspection, classification, assembly, outbound inspection, and departure processes respectively.

Four factors can be considered to affect each process and hence the operating plan. The four factors are:

(1). Terminal physical configuration (layout);

(2). Terminal resources;

(3). Operating policy;

(4). Terminal current conditions

The terminal physical configuration includes the number of receiving tracks, the length of each receiving track, the number of tracks in the bowl, the length of each track in the bowl, the number of departure tracks, the length of each departure track, the number of leads, the length of each lead, and so on.

The terminal resources include the number of crews, clerks, and inspectors on each shift, the number of switching engines on each shift, and so on.

The operating policy includes blocking policy (including the priority of traffic), train schedule, car scheduling, train make-up plan, crew schedule, power and empty car distribution and so on.

The terminal current conditions include traffic volume and its distribution, the terminal inventory at the beginning of each shift, and so on. The ETAs and ETDs are the basic current conditions and requirements of the terminal operations. Here, the ETAs and ETDs are used to refer to all the current conditions.

Among the four factors, the terminal managers can only control terminal resources such as number of crews, number of engines, and the method to use the resources and the terminal layout in terms of track assignment to blocks in the bowl. The terminal managers can affect the inventory of the next shift through their effort during the current shift. The terminal managers generally can not control other factors. The factors can be analyzed in detail for each process.

(1). Inbound arrival process

Terminal layout (L_i): the number of receiving tracks, the length of each receiving track, conflicts among processes to gain access to receiving yard, location of receiving tracks relative to main line;

Resources (R_i): the number of crews and engines in each inbound train;

Operating policy (O_i): the procedure and method of receiving inbound trains;

Terminal current conditions (E_i): the number of inbound trains, the number of cars in each inbound train, arrival times of inbound trains and their distribution, the priority of inbound trains, the number of inbound trains sitting at the receiving tracks at the beginning of the shift.

The above analysis shows that the inbound arrival process is a function of layout (L_i), resources (R_i), operating policy (O_i), and current conditions (including ETAs, ETDs and inventories) (E_i). That is:

$$P_i = f(L_i, R_i, O_i, E_i) \qquad (3.7.2)$$

(2). Inbound inspection process

Terminal layout (L_{ii}): the number of receiving tracks and the length of each receiving track;

Resources (R_{ii}): the number of inspection teams, the number of inspectors at each team, the number of clerks involving in paper work;

Operating policy (O_{ii}): the procedure and method of inbound inspection;

Current conditions (E_{ii}): the number of inbound trains, the number of cars in each inbound train, arrival times of inbound trains and their distribution, the priority of inbound trains, the number of bad order cars in each train, the number of inbound trains sitting at the receiving tracks at the beginning of the shift which are not inspected.

The inbound inspection process is a function of layout (L_{ii}), resources (R_{ii}), operating policy (O_{ii}), and current conditions (E_{ii}):

$$P_{ii} = f(L_{ii}\, R_{ii}\, O_{ii}, E_{ii}) \qquad (3.7.3)$$

(3). Classification process

Terminal layout (L_c): the number of leads to the hump for a hump terminal (which determines if two classification engines can cooperate for the classification operation) or the number of leads available for classification operations in a flat terminal, the maximum speed of the engine pushing the cars over the hump, which is determined by the retarding facilities at the near end of the bowl, or the maximum speed of classification for a flat terminal;

Resources (R_c): the number of classification engines, the number of crews; (the labor agreement can affect the length of working hours and hence affect the number of crews);

Operating policy (O_c): the blocking policy, the block to train assignment policy, the train schedule and so on;

Current conditions (E_c): the number of inbound trains, the number of cars in each inbound train, arrival times of inbound trains and their distribution, the priority of inbound trains, the number of cuts in each inbound train, the number of inbound trains waiting for classification at the beginning of the shift.

The classification process is a function of layout (L_c), resources (R_c), operating policy (O_c), and current conditions (E_c):

$$P_c = f\,(L_c,\, R_c,\, O_c,\, E_c) \tag{3.7.4}$$

(4). Assembly process

Layout (L_a): the number of leads available for assembly, the maximum speed for assembly switching, and grouped tracks from the design of the bowl (such as adjacent versus across several "pockets");

Resources (R_a): the number of "trim" engines, the number of assembly crews;

Operating policy (O_a): blocking policy, block to track assignment, train schedule, and block to train assignment policy;

Current conditions (E_a): the number of outbound trains to be assembled, the number of blocks and cars in each outbound train, departure times and distribution of outbound trains, the priority of outbound trains, the number of bad order cars in each outbound train, and the number of classified cars in each block at the beginning of each shift;

The assembly process is a function of layout (L_a), resources (R_a), operating policy (O_a), and current conditions (E_a):

$$P_a = f\,(L_a,\, R_a,\, O_a,\, E_a) \tag{3.7.5}$$

(5). Outbound inspection process

Layout (L_{oi}): the number of departure tracks, the length of each departure track;

Resources (R_{oi}): the number of outbound inspection teams, the number of inspectors in each team;

Operating policy (O_{oi}): train schedule, the procedure and method of outbound inspection operation;

Current conditions (E_{oi}): the number of outbound trains to be inspected, the number of cars in each outbound train, the departure times and distribution of outbound trains, the priority of outbound trains, the number of bad order cars in each outbound train;

The outbound inspection process is a function of layout (L_{oi}), resources R_{oi}, operating policy O_{oi}, and current conditions E_{oi}:

$$P_{oi} = f(L_{oi}, R_{oi}, O_{oi}, E_{oi}) \tag{3.7.6}$$

(6). Departure process

Layout (L_d): the number of departure tracks, the length of each departure track;

Resources (R_d): the number of road engines and the number of crews for each outbound train;

Operating policy (O_d): train schedule, crew assignment, power distribution policy, time required to add power to each outbound train, and the dispatching policy;

Current conditions (E_d): the number of outbound trains, the number of cars in each outbound train, the departure times and distribution of outbound trains, the priority of outbound trains, the number of outbound trains waiting for departure at the beginning of each shift.

The departure process is a function of layout (L_d), resources (R_d), operating policy (O_d), and current conditions (E_d):

$$P_d = f(L_d, R_d, O_d, E_d) \tag{3.7.7}$$

4.4.2 Methods to Measure Achievability of TOP

The previous analysis shows how the achievability of an operating plan can be considered as a function of the six processes and the four categories of factors. The effects of all the factors on each process can be expressed as the processing rate or time for individual tasks with probabilistic nature. Because the required time for a specific task in a given process can be regarded as continuous, the continuous probability function for the process is appropriate. From the data base described in section 3.3, the approximated probability density function can be obtained.

It is assumed that the pure processing times for individual tasks are mutually independent, in the sense that the later processing times are not affected by the earlier processing times. This assumption is made for all the processes and all the tasks in the terminal operating plan. For example, if there are two inbound inspection teams, it is reasonable to assume that one team's inspecting time is independent of the other team's inspecting time[9]. It is also assumed that the later classification processing time is independent of the earlier classification time. Here "independent" means that the length of earlier pure processing time does not affect the length of later pure processing time. The length of a pure processing time for a specific task is treated as a random variable rather than the time the task begins or ends. The independence assumptions of the tasks in each process appear to be reasonable. The independence assumptions between processes appear reasonable for some processes. For example, the arrival process and inbound inspection process can be regarded as independent because there is little interaction between these two processes. The independence assumption between some other processes may not be as reasonable. Instead, they should be treated as a convenient way to model the processes. For example, the inbound inspection process may not completely independent of the classification process in practice, since if some classification task is a critical task, the corresponding inbound inspection may be speeded up or even be canceled.

There are two methods to measure the achievability of TOPs. One is the PERT/CPM method, which measures the probability of performing the critical tasks for each outbound train from all its inbound connections. The achievability of performing an overall TOP is then defined as weighted average of the probabilities of all the outbound critical paths in the

---

[9] This assumption presumes that inspection time is a measure of the time to conduct or carry out the process. If the processing time is predetermined, either as a result of workers "agreeing" to take a specific time to conduct the activity, or because of a management policy, (such as allowing two hours for inspection and for repair for each track), this assumption does not hold.

TOP (for example, weighted by the number of cars or the priority of the traffic). The other method presented is called the distribution method, which applies statistical distributions of the pure processing times in the calculation of the achievability of TOPs. Note that the use of the PERT/CPM method does not allow for process level achievability, while the distribution method does. On the other hand, the PERT/CPM method uses a technique which is well known in the literature of project management.

4.4.2.1 PERT (CPM) Method

The Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) are usually used in project management. Here, PERT and CPM are applied to evaluate the achievability of a terminal operating plan. The terminal operating plan is actually a list of tasks to be performed in a shift, with time requirement for each task. Similarly, a project contains a list of activities or tasks to be performed with a time requirement for each activity or task. The terminal operating plan can be treated as a project and the achievability of the terminal operating plan can be evaluated using PERT and CPM technique.

From the data base described in section 3.3, the pure processing time distribution can be obtained for each task in each process. The achievability of an operating plan can be evaluated using probabilistic technique in the PERT and CPM method. Before presenting the technique, the basic notations are introduced in the context of a rail terminal operating plan.

An *activity* in a terminal operating plan is defined as any process which needs time and resources to perform. Each activity of the operating plan has a definable beginning and ending. It is also referred to as a *task*.

The beginning and ending points of activities are called *events*. An event is a discrete point in time. For example, the start time of an inbound train's classification is an event.

A *network* is a graphical representation of the operating plan showing the interrelationships of the various activities. Some activities or processes must be performed in a particular sequential order. Before an activity may begin, all activities preceding it must be completed and the arrows imply logical precedence only (Moder et al. 1970, pp. 25)

In a network representing an operating plan, an activity (i, j) means that there is predecessor event i and the successor event j. The following notation is used in the computation of the critical path of the network:

t(i, j): estimate of the mean duration time for activity (i, j);

$t_E(i)$: earliest occurrence time for event i;

$t_L(i)$: latest allowable occurrence time for event i;

$t_{ES}(i,j)$: earliest start time for activity (i, j);

$t_{EF}(i,j)$: earliest finish time for activity (i, j);

$t_{LS}(i,j)$: latest allowable start time for activity (i, j);

$t_{LF}(i,j)$: latest allowable finish time for activity (i, j);

$t_S(i,j)$: total slack time for activity (i, j);

$t_{FS}(i,j)$: free slack time for activity (i, j);

T(OB #i): scheduled or allocated time for the complete processing of an outbound train i.

PERT/CPM allows for forward pass and backward pass methods of calculation. In forward pass calculation, all the events that do not have predecessor events can be assigned the earliest starting time from the terminal operating plan. For example, an inspected inbound train waiting for classification at the beginning of a shift does not need inbound arrival or inbound inspection processing. The set of these events is denoted as E_0. That is:

$$t_E(i) = t(i), \quad i \in E_0 \qquad (3.7.8)$$

where, t(i) is the starting time of the event i in the operating plan.

For event i which has predecessor event(s), the earliest occurrence time is:

$$t_E(j) = \max_{(i,j)\in A}\{t_E(i) + t(i,j)\} \qquad (3.7.9)$$

where, A is the arc (activity) set of the network.

In backward pass calculation, all the events representing the ending times of outbound train operations (denoted the set as E(OB)) are assigned to be equal to the corresponding earliest occurrence time:

$$t_L(i) = t_E(i), \quad i \in \mathrm{E(OB)} \qquad (3.7.10)$$

For other events:

$$t_L(i) = \min_{(i,j)\in A}\{t_L(j) - t(i,j)\} \qquad (3.7.11)$$

After $t_E(i)$ and $t_L(i)$ are obtained for all the i's in the network, the various time introduced earlier can be obtained as follows:

$$t_{ES}(i,j) = t_E(i) \qquad (3.7.12)$$

$$t_{EF}(i,j) = t_E(i) + \mathrm{t(i, j)} \qquad (3.7.13)$$

$$t_{LF}(i,j) = t_L(j) \qquad (3.7.14)$$

$$t_{LS}(i,j) = t_L(j) - \mathrm{t(i, j)} \qquad (3.7.15)$$

$$t_S(i,j) = t_L(j) - \mathrm{t(i, j)} - t_E(i) \qquad (3.7.16)$$

$$t_{FS}(i,j) = t_E(j) - \mathrm{t(i, j)} - t_E(i) \qquad (3.7.17)$$

Note that (3.7.12) to (3.7.17) are not necessary to calculate the critical path.

Using this method, the critical paths of a terminal operating plan can be determined. Note that each outbound train has an critical path which represents the longest path from the related inbound train operations to the outbound train departure operation. The achievability of the terminal operating plan can be estimated as the averaged probability of performing the critical paths within the given scheduled or allocated times in the operating plan. Various weights can be used, such as the number of cars in the outbound trains or the outbound traffic priority.

This method can be used to calculate task level and overall plan achievabilities but can not derive the process level achievability.

Example:

Suppose that the TOP in Table 2 of chapter three is given and the pure processing times are normally distributed and are independent. Assume the pure processing times have the following forms:

IB inspection process: $t(I) = 20 + 0.10*Cars + \varepsilon(I)$ $\varepsilon(I) \sim N(0, 9)$
Classification process: $t(C) = 15 + 0.25*Cars + \varepsilon(C)$ $\varepsilon(C) \sim N(0,25)$
Assembly process: $t(A) = 30 + 0.15*Cars + \varepsilon(A)$ $\varepsilon(A) \sim N(0,25)$
OB inspection process: $t(O) = 30 + 0.15*Cars + \varepsilon(O)$ $\varepsilon(O) \sim N(0,16)$

From the TOP in chapter three (Table 2), the following network can be drawn:

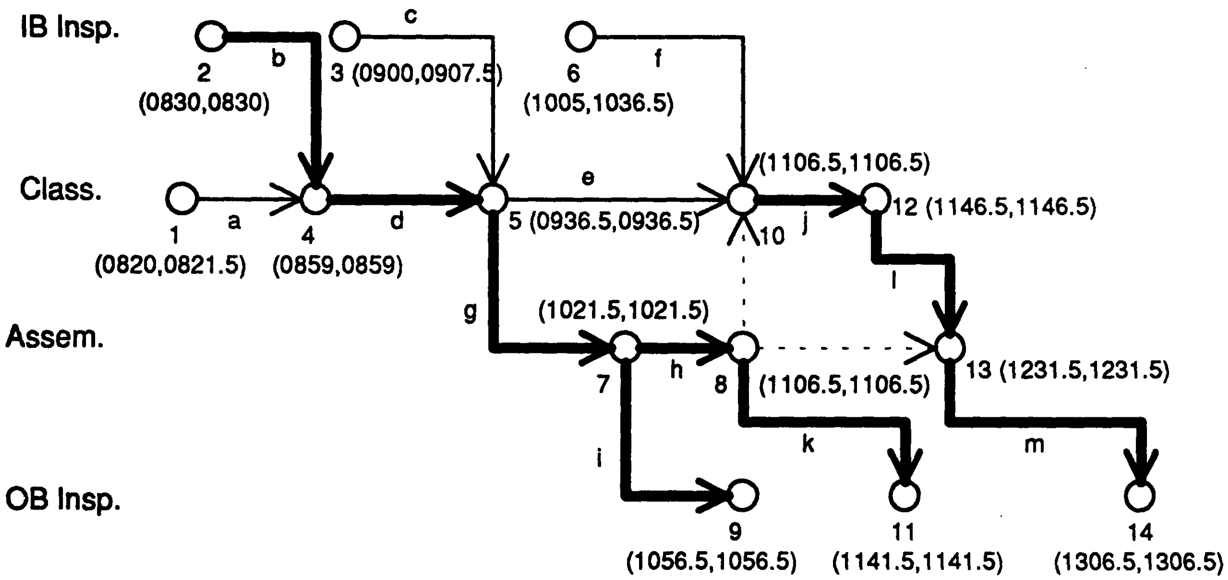

Figure 10. PERT/CPM Network Representation

The activities are defined as follows:

a: classifying inventory (expected time is 37.5 min);

b: inspecting IB #1 (expected time is 29 min);

c: inspecting IB #3 (expected time is 29 min);

d: classifying IB #1 (expected time is 37.5 min);

e: classifying IB #3 (expected time is 37.5 min);

f: inspecting IB #2 (expected time is 30 min);

g: assembling OB #1 (expected time is 45 min);

h: assembling OB #2 (expected time is 45 min);

i: inspecting and departing OB #1 (expected time is 35 min);

j: classifying IB #2 (expected time is 40 min);

k: inspecting and departing OB #2 (expected time is 35 min);

l: assembling OB #3 (expected time is 45 min);

m: inspecting and departing OB #3 (expected time is 35 min).

The expected times are obtained from the behavior model forms. In the network, the interrelationships of the processes are expressed very clearly. For example, event 4, which is the start time for classifying IB #1, can not begin until both the inbound process of IB #1 and the classification operation of the inventory train are finished.

The achievability of the operating plan is given by:

$$A = \frac{\sum a_i * P_i}{\sum a_i} = \frac{100*0.9418 + 100*0.9633 + 100*0.9995}{100 + 100 + 100} = 0.968$$

The predicted performance of the operating plan:

Total cars in the plan: 370

Number of cars making their connections: 300

Predicted PMAKE = 300/370 = 81.1%

Block a PMAKE = 65/80 = 81.3%

Block b PMAKE = 35/55 = 63.6%

Block c PMAKE = 100/128 = 78.1%

Block d PMAKE = 100/107 = 93.4%

The inbound train to outbound train PMAKE can be estimated similarly. The detailed calculation is given in Appendix A.

4.4.2.2 Distribution Method

As discussed in previous section, the probability density function of pure processing time can be obtained from the data base for each task in each process. Using this information, the three level achievabilities of an operating plan can be obtained using the following method. (The method is called the distribution method because the approximated distribution information obtained from the data base is used to calculate the achievabilities of a terminal operating plan).

The task level achievability: this level achievability measures the probability of accomplishing a specific task within the specified time window. Suppose that a task i in process j (i.e., arrival, inbound inspection, classification, assembly, outbound inspection, or outbound departure) is specified to be completed within t minutes, the assigned time interval. Suppose also that the probability density function (pdf) from the data base for this task in this process is f(x). Then the probability that the task can be performed, which is the task level achievability, is:

$$p = f(x < t) \qquad (3.7.18)$$

Note, that the results from the terminal process behavior models can also be used to estimate the pdf of the pure processing times, especially when the data base is comparatively small. For example, if the values of the factors which affect the pure processing time are known, these values can be used in the behavior model to obtain projected mean of the pdf.

For example, the probability of performing task b in the previous example is:

$$P_b = \Phi(\frac{t_b - \mu_b}{\sigma_b}) = \Phi(\frac{30-29}{3}) = 0.63$$

The process level achievability: this measures the probability that all the tasks in a process will be completed during a shift. Suppose that the total number of tasks in a specific process j in the shift from the terminal operating plan is n, and task i's mean time is μ_i, standard deviation is σ_i, and the assigned time length for the task is t_i. Since the pure processing times are assumed independent, the random variable

$$y_j = \frac{\sum_i t_i - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}} \tag{3.7.19}$$

is approximately normal distributed if n is very large or if the individual pdf in this process is normal. This random variable measures, in some degree, the whole set of tasks in a specific process. The normal cumulative probability function $\Phi(y)$ then measures the probability that the list of tasks in this process can be accomplished. That is:

$$P_j = \Phi(y_j) \tag{3.7.20}$$

where, P_j is the achievability of process j.

If the normal distribution conditions are satisfied as discussed above, the process level achievability can be calculated easily. Some approximated methods can be used to estimate the process level achievability if the normal distribution conditions are not satisfied.

Using the example in PERT/CPM section, for example, the classification process level achievabilities is:

$$P_C = \Phi(y_C) = \Phi(\frac{\sum_i t_i - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}})$$

$$= \Phi(\frac{40+40+45+40-37.5-37.5-37.5-40}{\sqrt{25+25+25+25}}) = 0.89$$

The overall or plan level achievability: this measures the probability that the overall plan can be accomplished. As discussed earlier the pure processing times from different processes are assumed independent. The overall achievability can be estimated by the product of the process level achievabilities:

$$P=\prod_{j=1}^{n} P_j \tag{3.7.21}$$

where, P is the overall plan achievability, P_j is the individual process level achievability, and n is the number of processes considered.

Using the example in PERT/CPM method section:

$$P=\prod_{j}^{4} P_j=0.99*0.89*0.96*0.98=0.83$$

As discussed earlier, the independence assumption between processes ignores the possible interaction between the processes. The results obtained can be used by terminal managers, terminal crews, and system level managers. For example, the terminal trainmaster can use the overall achievability as one means to evaluate the available TOPs. All else equal, the TOP with highest overall achievability is the best plan. Also, the system level managers can use the overall achievability as an important measure to evaluate the work of different terminals together with the consideration of the resources and facilities in the terminals. The terminal trainmaster and the yardmasters can use the process level achievability to better plan the utilization of the available terminal resources. For example, if the inbound inspection process has a much larger achievability than outbound inspection process, it may be reasonable to reallocate some inbound inspectors to do outbound inspection work. Similarly, the yardmasters and terminal crews can use the task level achievabilities to better plan their work. For example, if a task has relatively low achievability, attention can be paid to this task. In terms of work planning, higher priority traffic can be assigned more time and hence has higher achievability than low priority traffic.

A detailed example using the same data from PERT/CPM method section is presented in Appendix B.

4.5 Choosing a TOP from Alternatives

Different specifications about the times to perform individual tasks will lead to different operating plans. It is clear that several operating plans are available. The approach presented in this thesis considers not only projected performance but also the achievability of the plan. How the terminal managers to choose one operating plan from several alternatives depends on which dimension they emphasize more.

Suppose that there is another operating plan (called plan B) as in Table 6.

Table 6: Terminal Operating Plan B:

Train ID	Arrival Time	Block(a, b, c, d)	IB Inspection Time	Classification Time	Assembly Time	OB Inspection Time
Inventory		25, 10, 35, 20	(inspected)	0800-0850		
IB #1	0830	30, 15, 33, 12	0830-0850	0850-0930		
IB #3	0900	10, 10, 40, 30	0900-0930	0930-1005		
IB #2	1000	15, 20, 20, 45	1000-1020	1020-1050		
OB #1		80, 45, 0, 0			1050-1130	1130-1205
OB #2		0, 10, 118, 0			1130-1200	1200-1252
OB #3		0, 0, 10, 100			1200-1240	1240-1430
Inventory		0, 0, 0, 0				

In this plan, less time is assigned to classification and assembly operations so that all the cars in the inbound trains can make their connections to the outbound trains. Using the

PERT/CPM method, the plan level achievability is only about 50%. The comparison of the two plans is as follows:

Table 7: Comparison of Plan A and B

Operating Plan	Achievability	Connection Performance
A	0.97	81%
B	0.50	100%

Figure 11 shows the relationship between the connection performance and the achievability of the two plans.

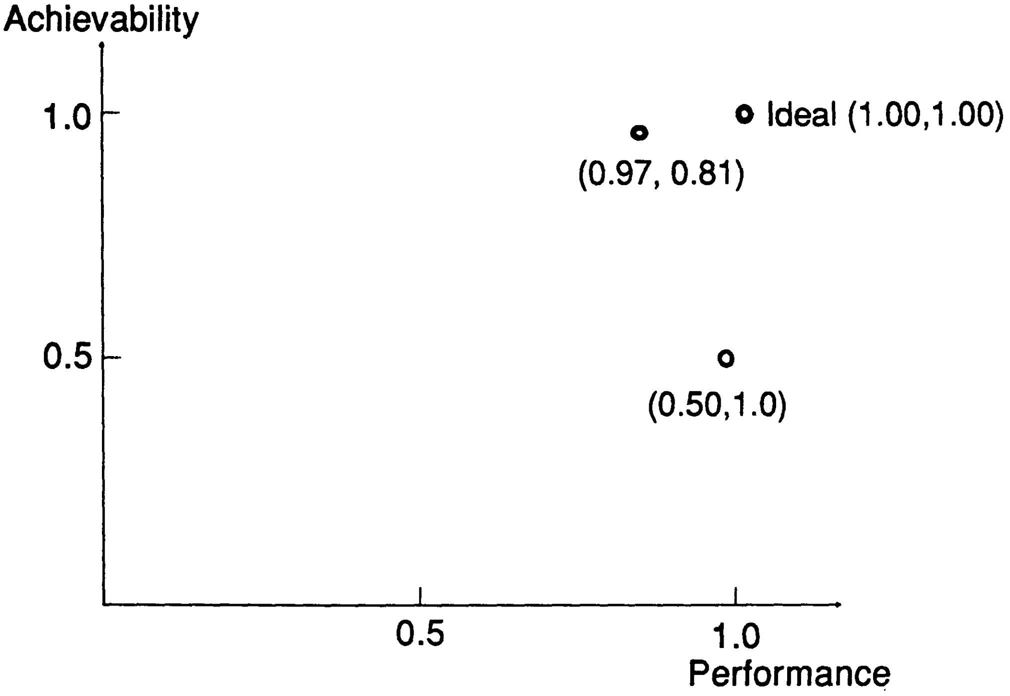

Figure 11. The Relationship between Performance and Achievability

If the terminal managers emphasize projected performance while thinking that 0.50 achievability is acceptable, they can choose plan B. If they focus more on the achievability dimension, while considering projected connection performance of 0.81 acceptable, they can choose plan A. Also, they can specify different times for individual tasks to get other plans. There is generally a trade-off between achievability and connection performance. It is possible to find a plan that both achievability and performance measures are satisfied to a certain level from the terminal managers' point of view.

Chapter 5. A Case Study

The following case study was conducted using data from Radnor Terminal of CSXT Rail Transportation Company. CSXT is a Class one railroad in the US. and the Radnor Terminal is a major yard on the Chicago-Nashville Corridor.

5.1 Radnor Terminal Description

Radnor Terminal is located in Nashville, Tennessee and is considered by CSXT to have great potential to improve service reliability of both the terminal and the system. The configuration of the terminal is in Figure 12.

The *receiving yard* of the terminal has 12 arrival tracks and 1 dedicated running track. The tracks range from 109 to 119 car lengths. The maximum speed of the arrival operation is 10 miles per hour for all the tracks. The receiving yard provides for an inventory capacity of about 1400 cars. There are 72 inbound and outbound scheduled trains daily including by-passing trains. There are 16 scheduled arrival trains which must go through the classification process.

Radnor terminal is a *hump terminal*. The hump speed is about 1 mile per hour. This average speed includes starts and stops during the hump process. Studies have shown the hump speed across the hump is greater than this average speed.

The *bowl yard* contains 56 classification tracks in 7 groups or pockets of 8 tracks each. These tracks range from 36 to 68 car lengths and provide for an inventory capacity of about 2600 cars. The assignment of tracks to blocks is dynamic.

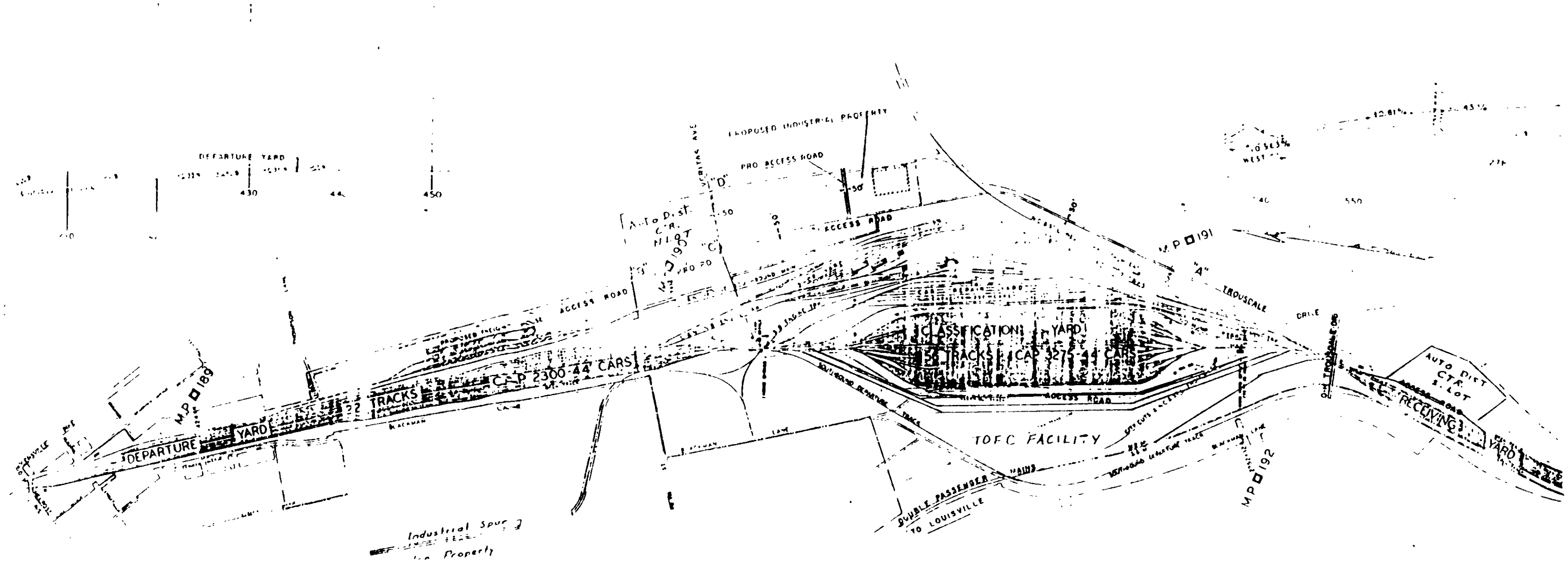

Figure 12. Radnor Terminal Configuration

There are 5 *leads or throats* between the bowl and the departure yard which are used for assembly operations. The *departure yard* contains 15 tracks (8th through 22nd tracks) that outbound trains are built on and 11 local yard tracks. The departure yard capacity for outbound trains is about 1650 cars. The outbound tracks range from 100 to 200 car lengths and the local tracks range from 147 to 238 car lengths.

The period of the case study is from September 15 through 22, 1993, which accounts for 7 days or 21 shifts. Data was collected by an interdepartmental team of CSXT. During the case study period, the MIT Railgroup was invited by the CSXT team and participated in the activities of CSXT team for three days. The MIT team also took part in the design of the case study, particularly regarding what data should be collected. After the case study was completed, CSXT made all the data available to the MIT Railgroup. This data base contains 115 inbound trains, 150 outbound trains, and about 9400 cars humped. After the MIT Railgroup received the data, a time-space diagram was drawn for the study period to record all the major activities in the four processes (IB inspection, classification, assembly, and OB inspection) and arrival and departure time of each train in a consistent manner.

The trains arriving or departing from the terminal are divided into the following categories: Q train, R train, S train, M train, and Y train in descending priority order.

5.2 Developing Process Behavior Models

5.2.1 Pure Processing Times

From the available data base, the pure processing times were estimated in Table 8.

Table 8 includes the major pure processing times and the waiting times between the successive processes in terms of means, standard deviations, minimum values, and maximum values of these times.

Table 8: Radnor Pure Processing and Waiting Times

Process	Mean	Standard Deviation	Minimum	Maximum
Receiving Time				
Waiting for IB Insp.				
Tran. Dept	3.9 hr	5.8 hr	0	24 hr
Mech. Dept	28 min	57 min	0	421 min
IB Inspection Time	122 min	25 min	55 min	280 min
Waiting for Hump	4.2 hr	2.4 hr	0.9 hr	13.1 hr
Hump Time	46 min	16 min	5 min	100 min
Arrival to Hump End	8.9 hr	3.7 hr	0.6 hr	19.2 hr
Assembly Time	2.4 hr	1. 2 hr	0.8 hr	7.7 hr
Waiting for OB Insp.				
Tran. Dept.	2 min	8 min	0	60 min
Mech. Dept	19 min	30 min	0	160 min
OB Inspection Time	1.8 hr	0.5 hr	0.1 hr	3.8 hr
Waiting for Departure	1.2 hr	1.5 hr	0	12.7 hr
Assembly to Departure	8.8 hr	3.4 hr	1.6 hr	20.4 hr

Since the arrival process was not considered explicitly by the CSXT study team in the study, this pure processing time is not available.

For inbound operations, the pure processing times are much smaller in terms of means and standard deviations than waiting times. For example, the waiting time for inbound inspection is more than 4 hours but the pure inbound inspection time is only about 2 hours (122 minutes). The standard deviation of the waiting time for inbound inspection is about 7 hours but the standard deviation of inbound inspection is about 25 minutes.

The overall inbound processing time (from arrival to end of hump) is about 8.9 hours, which includes inbound arrival, inbound inspection, and hump processes. Similarly, the outbound processing time which includes assembly, outbound inspection, and outbound departure processes is about 8.8 hours.

Comparing the inbound and outbound operations, the outbound waiting times are much smaller than the waiting times for inbound operations. For example, the waiting time for outbound inspection is only about 21 minutes. The waiting time for inbound inspection is more than 4 hours. This may indicate that the terminal managers tend to assign the outbound inspection as soon as possible after the train is assembled.

Note that the data analysis shows that the pure processing times do not have satisfactory reliability. Figure 13 shows the pure hump processing times. The results in this figure indicate that the pure hump processing times are not vary reliable. Before explicitly modeling the effects of queues of processes, the variation in arrivals and in resources, better understanding of pure processing times is desirable.

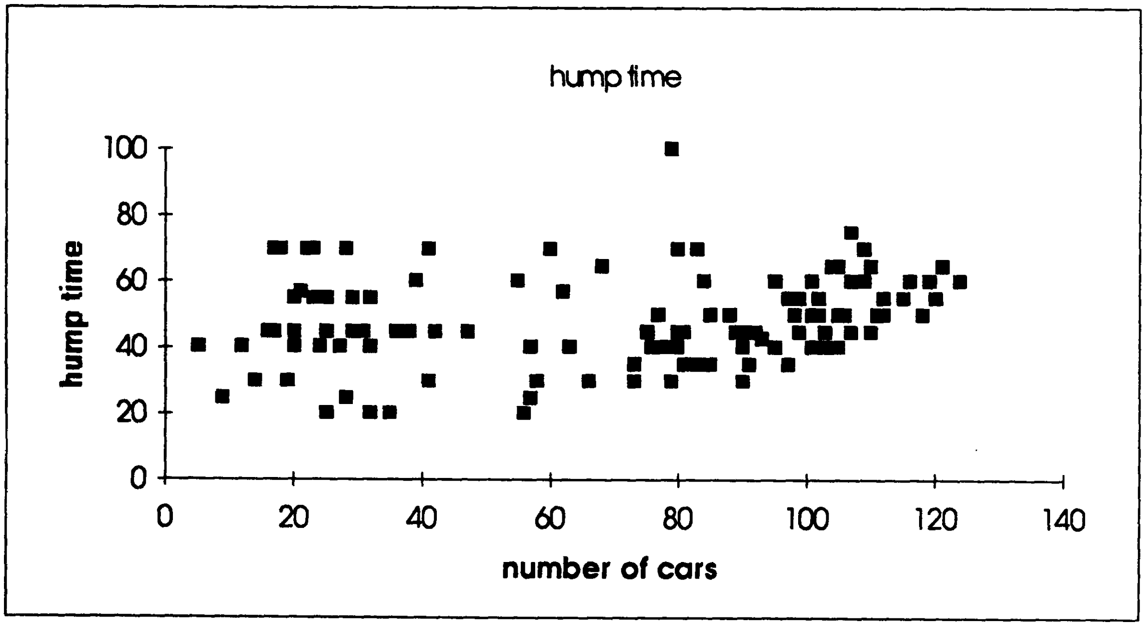

Figure 13. Pure Hump Processing Times

5.2.2 Developing Process Behavior Models

In this section, the factors affecting the pure processing times and the various waiting times are analyzed. Regression models were developed to reveal the relationship between the pure processing or waiting times and their corresponding factors. The model results can be used to predict the processing or waiting times and can provide useful information for other analysis in the development of the terminal operating plans.

5.2.2.1 Inbound Inspection Process

1. The time interval from arrival to given to inspectors (T1)

From the data base, the mean of T1 is 3.8 hours and standard deviation of T1 is 5.7 hours.

T1 is specified as a function of the following factors:

T1 = f(Q train dummy,	$\beta1, X1$
R train dummy,	$\beta2, X2$
S train dummy,	$\beta3, X3$
# of cars,	$\beta4, X4$
# of tracks waiting for inspection,	$\beta5, X5$
one hour before or after end shift dummy,	$\beta6, X6$
arrival late (two hours) dummy,	$\beta7, X7$
double over dummy)	$\beta8, X8$

It is believed that higher priority trains will be assigned an inbound inspection team earlier compared with lower priority trains if all the other conditions are the same. The larger the number of cars in an inbound train, the more likely that the inbound train is assigned an inspection team earlier. Similarly, the larger the number of tracks waiting for inbound inspection, the later the trains will be assigned to an inspection team.

The one hour before or after end shift dummy variable is 1 if the inbound train arrives at the terminal in the time window of one hour before the end of the shift and one hour after the next shift and 0 otherwise. If this variable is 1, it is believed that this train may be assigned to an inspection team later compared with the situation that the variable is 0. The reason is that at the beginning of a shift or at the end of a shift, the terminal managers are busy transferring the responsibility of the terminal operations from one shift to another and hence may not have enough time to assign inbound trains to inspection teams.

The arrival late (two hours) dummy variable is 1 if the train arrives at the terminal at least two hours later compared with its schedule, and 0 otherwise. It is believed that if this variable is 1, it may take longer time to give it to an inspection team because it is more likely to have missed its connections to outbound trains. The double over dummy variable is 1 if the inbound train is yarded on more than one track, and 0 otherwise.

Using the linear regression model, the functional form is:

$T1 = \alpha + \beta1 * X1 + \beta2 * X2 + \beta3 * X3 + \beta4 * X4 + \beta5 * X5 + \beta6 * X6 + \beta7 * X7 + \beta8 * X8$

Results (unit is hour):

Parameters		Coefficients	p-values
α	Intercept	10.20	<0.001
$\beta1$	Q train dummy	-6.86	<0.001
$\beta2$	R train dummy	-6.50	<0.001
$\beta3$	S train dummy	-5.70	0.002
$\beta4$	# of cars	-0.026	0.071
$\beta5$	# of tracks waiting	0.22	0.53*[10]
$\beta6$	one hour dummy	0.70	0.57*
$\beta7$	arrival late dummy	0.42	0.69*
$\beta8$	double over	-2.13	0.31*

---

[10] Not significant at 0.10 level

R-Square: 0.33

Adjusted R-Square: 0.29

observations: 134

All the coefficients except two hours late dummy variable have the expected signs. The insignificant late dummy variable indicates late inbound trains are not processed differently from other, all else equal.

The double over dummy variable has a negative sign, meaning that when an inbound train is doubled over, the time interval is shorter compared with non double over trains, all else equal. This is not surprising if we consider the fact that all except one of the doubled over trains were long high priority trains.

The one hour arrival at beginning or end of the shift dummy variable, number of tracks waiting for inspection variable and two hours late variable are not significant. The results show that this time interval is not significantly related to arrival time of the shift and the number of tracks waiting for inspection.

2. The time interval from given to inspectors to the start of IB inspection (T2)

This time interval is waiting time for inspection. T2 was specified as a function of the following factors:

T2 = f(# of tracks waiting for inspection,	$\beta1$, X1
# of cars,	$\beta2$, X2
one hour before or after end shift dummy,	$\beta3$, X3
Q train dummy,	$\beta4$, X4
R train dummy,	$\beta5$, X5
S train dummy)	$\beta6$, X6

The larger the number of tracks waiting for inspection, the more the waiting time might be. The larger the number of cars at a track, the shorter waiting time might be. If arrival time is within one hour from the beginning of the shift or within one hour before the end of the shift, the value is 1, and 0 otherwise. The effect of this dummy variable on T2 is similar to that on T1.

From the data base, the average waiting time is 28 minutes and the standard deviation is 6 minutes.

Results (unit is minute):

Parameters		Coefficients	p-values
α	Intercept	177.44	<0.001
$\beta1$	# of tracks waiting	48.56	<0.001
$\beta2$	# of cars	-1.64	<0.001
$\beta3$	one hour dummy	-4.16	0.89*[11]
$\beta4$	Q train dummy	24.95	0.59*
$\beta5$	R train dummy	41.23	0.37*
$\beta6$	S train dummy	150.11	0.01
R-Square:		0.35	
Adjusted R-Square:		0.31	
# observations:		93	

The results show that the shift dummy variable and train type dummy variables have unexpected signs and are insignificant.

3. IB inspection time (T3)

T3 was specified as a function of the following factors:

T3 = f(# of cars,	$\beta1$, X1
# of inspectors,	$\beta2$, X2
# of bad order cars,	$\beta3$, X3
shift 1 dummy,	$\beta4$, X4
shift 2 dummy)	$\beta5$, X5

---

[11] Not significant at 0.10 level

It is believed that the more cars in an inbound train, the more time needed to inspect the train. The more inspectors, the less time needed to inspect the train. The more bad order cars in a train, the more time needed to inspect the train because for light bad order cars, light repair is needed. The inspection time may be different for different shifts.

Results (unit is minute):

	Parameters	Coefficients	p-values
α	Intercept	91.50	<0.001
$\beta 1$	# of cars	0.55	<0.001
$\beta 2$	# of inspectors	-15.90	<0.001
$\beta 3$	# of b/o cars	0.56	0.53*[12]
$\beta 4$	shift 1 dummy	7.85	0.17*
$\beta 5$	shift 2 dummy	0.35	0.95*
	R-Square:	0.32	
	Adjusted R-Square:	0.23	
	# observations:	101	

The estimation results show that all except the two shift dummy variables have the expected signs. The bad order variable is not significant, indicating the bad order cars do not have significant effect on the inspection time. (Actually, it is the light bad order cars that would be expected to affect the inspection time).

5.2.2.2 Hump Process

1. Time interval from the end of inspection to the beginning of hump (T4)

---

[12] Not significant at 0.10 level

This time is spent waiting for hump. Generally speaking, the hump rate (or hump capacity), inspection rate and traffic volume determine the queue of the hump and hence affect the queue length and queue time. T4 is a function of the following variables:

T4 = f(# of cars,	$\beta 1$, X1
# of tracks waiting for hump,	$\beta 2$, X2
Q train dummy,	$\beta 3$, X3
R train dummy,	$\beta 4$, X4
S train dummy,	$\beta 5$, X5
shift 1 dummy,	$\beta 6$, X6
shift 2 dummy)	$\beta 7$, X7

It is believed that all else equal, the track with more cars may be humped first. The greater the number of tracks waiting for hump after inspection, the more time is spent waiting. Different train types may have different priority, and the waiting time for different type of trains may be different. Also, the waiting time may be different at different shifts.

The average waiting time for hump is 4.2 hours and the standard deviation of the time interval is 2.4 hours. Since the hump pure processing time is about 45 minutes for an average train, this waiting time is extremely large. There may be great potential for improving reliability of the yard operation and reducing yard time by seeking ways to reduce this waiting time.

Results (unit is minute):

Parameters		Coefficients	p-values
α	Intercept	134.48	0.022
$\beta 1$	# of cars	0.24	0.64*[13]
$\beta 2$	# of tracks waiting	60.48	<0.001
$\beta 3$	Q train dummy	-16.84	0.67*

---

[13] Not significant at 0.10 level

β4	R train dummy	-48.34	0.21*
β5	S train dummy	25.83	0.61*
β6	shift 1 dummy	-45.58	0.11*
β7	shift 2 dummy	11.56	0.70*

R-Square: 0.38
Adjusted R-Square: 0.33
observations: 94

The results show that the estimated coefficients have the expected signs and some of the coefficients are significant. These results suggest that trains wait for humping because other trains are utilizing the hump, and the sequence is FIFO. The t-test shows that there is no difference, all else equal, in the waiting time for the second shift and the third shift. The number of cars variable is also not significant, indicating this factor is not important when the terminal managers determine the sequence of hump when more than one tracks are waiting for hump.

4. Hump time (T5)

The average hump time is 46 minutes per train and the standard deviation is about 16 minutes, indicating a vary stable processing rate.

It is believed that T5 is a function of the following factors:

T5 = f (# of cars in train (or track),	β1, X1
# of cuts in the train,	β2, X2
# of engines working for hump,	β3, X3
trim work dummy,	β4, X4
shift dummy)	β5, X5; β6, X6

The more cars in a train, the more time is needed to classify the train. The larger the number of cuts in a train, the more time is needed to classify the train. In this case study, cut information was not available. During the study period, there were always two engines working each shift, the number of engines working variable is not applicable. The trim work is when the engines go to the bowl tracks to collect the cars and push them to the far end of the bowl. If trim work happened, more time is needed. Also, the shifts may have different effect on the hump time.

The receiving yard data set is recorded by inbound trains, and two inbound trains may be classified at the same time because they are sitting at the same track. So the data set can not be used directly. The data from a time-space diagram, which was created to record all the major activities during the study period in the yard, was used. The diagram correctly recorded the number of cars humped each time. Using these data to fit the regression model, the estimated results are as follows:

$$T5 = \alpha + \beta 1 * X1 + \beta 2 * X2 + \beta 3 * X3 + \beta 4 * X4 + \beta 5 * X5 + \beta 6 * X6$$

Results (unit is minute):

Parameters		Coefficients	p-values
α	Intercept	15.05	0.016
$\beta 1$	# of cars	0.34	<0.001
$\beta 2$	# of cuts		
$\beta 3$	# of engines		
$\beta 4$	trim dummy	17.75	<0.001
$\beta 5$	shift 1 dummy	-0.94	0.792
$\beta 6$	shift 2 dummy	4.77	0.217

R-Square:	0.40
Adjusted R-Square:	0.37
# observations:	99

The results show that both the number of cars variable and trim dummy variable have the expected signs and are significant.

Since there are only two significant variables in the model, and lacking the number of cuts variable, the ability of the model to explain hump time is modest. Deeper investigation of the factors affecting the hump time may be needed.

5.2.2.3 Assembly Process

Pure assembly processing time (T6) is a function of the following factors:

T6 = f (# of cars in the train, $\beta 1$, X1
# of blocks in the train, $\beta 2$, X2
# of throats used) $\beta 3$, X3

The greater the number of cars in a train, the more assembly time may be needed. Similarly, the larger the number of blocks in the outbound train, the more time may be needed to assemble the train. The more throats used in the assembly, the more time is needed. The time-space diagram data was used in the assembly regression model.

Results (unit is minute):

Parameters		Coefficients	p-values
α	Intercept	35.07	0.051
$\beta 1$	# of cars	0.23	0.270*[14]
$\beta 2$	# of blocks	25.96	<0.001
$\beta 3$	# of throats	2.93	0.73*

R-Square: 0.28
Adjusted R-Square: 0.26
observations: 139

---

[14] Not significant at 0.10 level

The model results show that all the coefficients have the expected signs. The effect of number of throats used is very small.

5.2.2.4 Outbound Inspection Process

Pure outbound inspection processing time (T7) is a function of the following factors:

T7 = f (# of cars in the train, $\beta 1$, X1
# of inspectors, $\beta 2$, X2
# of bad order cars) $\beta 3$, X3

As in the inbound inspection analysis, the outbound inspection time is related with the number of cars inspected, the number of inspectors employed for this inspection and the number of bad order cars. The linear regression model is as follows:

Results:

Parameters		Coefficients	p-values
α	Intercept	76.34	<0.001
$\beta 1$	# of cars	0.75	<0.001
$\beta 2$	# of inspectors	-16.41	0.003
$\beta 3$	# of bad order cars	2.01	0.62*[15]

R-Square:	0.33
Adjusted R-Square:	0.32
# observations:	128

All the estimated coefficients have the expected signs. The coefficient for the number of cars variable shows that 0.75 minutes is needed to inspect one car holding other variables

---

[15] Not significant at 0.10 level

constant. The coefficient of the number of inspectors shows that all else equal, an extra inspector may reduce the inspection time by 16.4 minutes. The coefficient of bad order cars means that one more bad order car will need 2 more minutes to handle though the estimated coefficient is not statistically significant different from zero. (The presence of bad order cars may affect the time between inspection and departure, as this is the period when the defective car would be removed from the train.)

Summary: The results of the regression models of the yard operations show that the model can only explain 25-35% variability of the operation times. The other 65-75% variability of the operation times can not be explained by the available variables. This may indicate two things. One is that the operations of the yards are fairly unreliable. The other is that the terminal operations are not understood well enough. It is believed that a reliable yard requires reliable operations at each process. More research may be necessary to understand the yard operation and find the factors affecting the processing times.

5.3 Example of Developing TOP and Measuring Achievability

In this section, the Radnor Terminal data is used to develop a simple TOP for one shift. The shift of 0800 to 1600 in September 18 is used in this example. First, from the data base, the inventory of the terminal at the beginning of the shift, and the arrival and departure traffic during the shift can be obtained. Then the regression models developed in section 5.2 can be used to estimate the amount of time to perform various tasks or activities in Radnor Terminal. From the available tasks in the terminal and the expected time to accomplish these tasks, alternative plans can be developed. Since block information such as how many cars in each inbound train or outbound train and car connection information is not available, the detailed connection information can not be predicted from the TOP. But if the assigned end time for each task is not later than the actual finishing time of the task and all the actual finished tasks are assigned in the TOP, the connection performance of the TOP should be as good as the actual connection performance.

At the beginning of the shift, six tracks at the receiving yard were occupied, which accounts for an inventory of 512 cars at the receiving yard. The detailed inventory information at the receiving yard can be expressed by Table 9.

Table 9: Radnor Receiving Yard Inventory Information

Track ID	Train ID (Number of Cars)	Inspected	Humped
A3	Q57517 (57) + Q52016 (20 D/O[16])	yes	no
A5	S52015 (102)	no	no
A6	Q52016 (91)	yes	no
A8	M71917 (43) + R53217 (24 D/O)	yes	no
A10	Q68416 (105)	yes	no
A11	Y33017 (53) + Q53617 (17 D/O)	no	no

From the behavior models developed in section 5.2, the expected inspection times can be estimated (ignoring the insignificant factors in the models) as in Table 10.

Table 10: Expected and Actual Inspection Times

Track ID	Expected Inspection Time	Actual Inspection Time
A5	116 min	130 min
A11	98 min	130 min

Note that the actual inspection time includes necessary walking time and time for preparation. The predicted time is only pure inspection processing time. Similarly, the expected hump time and actual hump times are given in Table 11.

---

[16] Doubled over train

Table 11: Expected and Actual Hump Times

Track ID	Expected Hump Time	Actual Hump Time
A3	41 min	40 min
A5	50 min	55 min
A6	46 min	35 min
A8	56 min	50 min
A10	51 min	40 min
A11	57 min	55 min

The inventory at the bowl at the beginning of each shift is on hand. Since the block and car connection information is not available, the trains for which the cars came were not available. The departure yard inventory information can be expressed by Table 12.

Table 12: Radnor Departure Yard Inventory Information

Track ID	Train ID (Number of Cars)	Assembled	Inspected
D8	Y33017 (66)	yes	yes
D12	R59618 (39)	yes	yes
D16	R53418 (99)	yes	no
D19	R53318 (120)	yes	no

From the behavior models developed in section 5.2, the expected outbound inspection time can be obtained. The results, together with the actual assembly and inspection time, are given in Table 13.

Table 13: Expected and Actual Outbound Inspection Times

Track ID	Expected Inspection Time	Actual Inspection Time
D16	118 min	135 min
D19	134 min	120 min

The arrival and departure trains during this shift are as expressed in Table 14 and 15 respectively.

Table 14: Radnor Yard Arrival Information

Track ID	Train ID (Number of Cars)	Arrival Time
A4	Q64916 (39)	10:25
	R55717 (55)	11:00
A6	S67516 (104)	14:50
A8	Q59517 (105)	13:40
A9	Q53617 (84)	09:10
A10	R53018 (81)	13:20

Table 15: Radnor Yard Departure Information

Track ID	Train ID (Number of Cars)	Departure Time
D8	Y33017 (66)	11:45
D10	M71918 (66)	09:40
D12	R59618 (39)	12:15
D14	R12018 (57)	15:30
D15	Q55618 (93)	14:00
D16	R53418 (99)	11:18
D19	R53318 (120)	09:15
D19	R18618 (33)	14:48
D21	R18518 (23)	13:55

Based on inventory information, arrival and departure information, and the expected time to perform each task in the shift, a terminal operating plan can be developed. First, suppose that the actual tasks performed in the shift constitute a terminal operating plan, designated operating plan 1. Also suppose the assigned time for each task is equal to the actual time performing the task. In this plan, the minimum time between inbound inspections for the same inspection team is 15 minutes. This figure is 10 minutes for outbound inspection process. The minimum time between humps is about 10 minutes. The minimum time between assembly is about 1 hour. For inbound inspection and hump operations, there is no lunch hour more than 30 minutes in the shift. The alternative plan developed later, plan 2, will adhere to these "constraints" and the constraints of processing sequence and available resources in the shift. There were five inbound inspectors, two hump engines, three assembly engines, and seven outbound inspectors available during the shift. Since the block and car connection information is not available (e.g., how many cars for each outbound train at any time is not available), in the alternative plan, the outbound assembly operation is not addressed explicitly and left unchanged.

Plan 1 is in given in Table 16.

Table 16: Radnor Terminal Operating Plan 1

Process	Order	Track ID	Train ID (Cars)	Actual Time	Achievability (assign = actual)
IB Inspection	1	A5	S52015 (102)	130 min	0.71
	2	A9	Q53617 (84)	85 min	0.20
	3	A11	Y33017 (53) + Q53617 (17)	130 min	0.90
	4	A4	Q64916 (39) + R55717 (55)	130 min	0.78
	5	A10	R53018 (81)	125 min	0.80
	6	A8	Q59517 (105)	125 min	0.63
Hump	1	A10	Q68416 (105)	40 min	0.25
	2	A8	M71917 (43) + R53217 (24)	50 min	0.36
	3	A6	Q52016 (91)	35 min	0.25
	4	A5	S52015 (102)	55 min	0.62
	5	A3	Q57517 (57) + Q52016 (20)	40 min	0.48
	6	A9	Q53617 (84)	60 min	0.84
Assembly	1	D15	Q55618 (93)	225 min	0.94
	2	D9	R67618 (59)	135 min	0.62
	3	D14	R12018 (57)	280 min	0.98
	4	D19	R18618 (33)	310 min	0.98
	5	D18	R68518 (106)	105 min	0.32
	6	D10	R58318 (75)	120 min	0.68
	7	D11	Q64818 (29)	45 min	0.49
	8	D12	R67418 (91)	185 min	0.84
OB Inspection	1	D16	R53418 (99)	135 min	0.72
	2	D21	R18518 (23)	140 min	0.98
	3	D9	R67618 (59)	100 min	0.66
	4	D15	Q55618 (93)	75 min	0.11
	5	D18	R68518 (106)	135 min	0.66
	6	D14	R12018 (57)	65 min	0.34

Process and overall achievability is in Table 17.

Table 17: Process and Overall Achievability of Plan 1

Process	Achievability
IB Inspection	0.88
Hump	0.42
Assembly	0.99
OB Inspection	0.73
Overall Plan	0.27

The results from Table 16 and 17 show that some tasks have very low achievability and the process achievabilities are not balanced. Note that the achievability of the tasks in hump process is generally low. The achievability calculations suggest that the hump yardmaster "assigned" time under plan 1 is greater than the actual time for several reasons. Of the 960 minutes available to the two switch engines in duty, pure processing time counts for only 29% of the time. Even allowing for lunch and coffee breaks, safety meetings and completion of the shift prior to scheduled shift end time, this figure is low. This suggests that either the engines perform other work, or that the time to go and prepare a track for humping is significant and should be included in the plan design. Similarly, the hump process level achievability is lower than other process level achievabilities. The results may suggest that either the tasks in the hump process with low achievability were accomplished, or more likely, that some extra actual processing time was not recorded in the hump pure processing time. Also, since the inbound and outbound inspectors are both odd numbers and the inspected trains are both even numbers, some inspectors only inspect one train in the whole shift. Based on the results and observation of the Radnor Yard operations from the time-space diagram, the alternative plan 2 is developed as in Table 18.

Table 18: Radnor Terminal Operating Plan 2

Process	Order	Track ID	Train ID (Cars)	Assigned Time	Achievability
IB Inspection	1	A5	S52015 (102)	140 min	0.83
	2	A9	Q53617 (84)	120 min	0.71
	3	A11	Y33017 (53) + Q53617 (17)	130 min	0.90
	4	A4	Q64916 (39) + R55717 (55)	145 min	0.91
	5	A10	R53018 (81)	135 min	0.89
	6	A8	Q59517 (105)	155 min	0.94
	7	A6	S67516 (104)	130 min	0.45
Hump	1	A10	Q68416 (105)	60 min	0.71
	2	A8	M71917 (43) + R53217 (24)	70 min	0.81
	3	A6	Q52016 (91)	60 min	0.81
	4	A5	S52015 (102)	60 min	0.74
	5	A3	Q57517 (57) + Q52016 (20)	60 min	0.88
	6	A9	Q53617 (84)	70 min	0.95
Assembly	1	D15	Q55618 (93)	225 min	0.94
	2	D9	R67618 (59)	135 min	0.62
	3	D14	R12018 (57)	280 min	0.98
	4	D19	R18618 (33)	310 min	0.98
	5	D18	R68518 (106)	144 min	0.53
	6	D10	R58318 (75)	120 min	0.68
	7	D11	Q64818 (29)	75 min	0.56
	8	D12	R67418 (91)	185 min	0.84
OB Inspection	1	D16	R53418 (99)	135 min	0.72
	2	D21	R18518 (23)	140 min	0.98
	3	D9	R67618 (59)	100 min	0.66
	4	D15	Q55618 (93)	120 min	0.59
	5	D18	R68518 (106)	135 min	0.66
	6	D14	R12018 (57)	90 min	0.55
	7	D10	R58318 (75)	110 min	0.63
	8	D19	R18618 (33)	38 min	0.16
	9	D11	Q64818 (29)	70 min	0.57

In this plan, for each inbound train, the inbound inspection time is changed in such a way that more pure processing time is added while the hump time and at least the minimum time between successive inspections are not affected. The hump time is also changed but the end time for each hump is not changed except train Q52016 which is delayed 20 minutes. The hump delay of train Q52016 does not affect assembly because all the assembly engines were working before the hump of this train. The delay of this hump does not affect successive humps much, since there were two hump engines working during the shift. The assembly times are not changed except that of the two outbound trains with low achievability (train R68518 and Q64818). The change in assembly time for the two outbound trains does not affect all the outbound inspections. For each outbound inspection, the time is changed in the similar manner as inbound inspection. By doing this, it is clear that the connection performance of this plan is as good as that of plan 1. Compared with plan 1, this plan processes one more inbound inspection and three more outbound inspections.

Process and overall achievability is in Table 19.

Table 19: Process and Overall Achievability of Plan 2

Process	Achievability
IB Inspection	0.99
Hump	0.99
Assembly	0.99
OB Inspection	0.80
Overall Plan	0.78

This plan has higher task level, process level, and overall achievabilities compared with plan 1. Since the tasks in the inbound inspection and hump processes are assigned no later than that in plan 1, (except the hump time for Q52016, 91 cars is delayed 20 minutes, which does not affect assembly time of outbound trains), the connection performance in plan 2 should at least as good as that in plan 1. But in plan 2, one more inbound train and three more outbound trains are assigned time for inspection operations. So the plan 2 is better than plan 1. This example shows that alternative plans exist and may be better than the current plans in terms of better achievability and connection performance. The models developed in previous sections can be used by terminal managers to develop simple terminal operation plans to manage and control terminal operations.

Chapter 6. Conclusions and Future Studies

6.1 Conclusions

This thesis focuses on improving *terminal service performance*. Studies show that terminal operation is a critical component for the rail system service performance. It has been shown that *terminals are significant factors to improve railroad service performance*.

Detailed *terminal operations and processes* are analyzed in this thesis. The *data issues* involved in the terminal operations and processes are stressed. Based on the terminal data, terminal *process behavior models* are developed. The purpose of this approach is to try to understand terminal processes, and to reveal the factors affecting these processes. The reliable terminal requires reliable processes, especially reliable pure processing times. By looking into process and task level performance, these behavior models can be used to manage the terminal tasks, making task level performance more reliable.

Based upon the results of the behavior models and the experience of terminal managers, the time required to accomplish individual tasks can be specified by the terminal managers. The specification can be conducted in such a way that these tasks are more likely to be accomplished, that is, these tasks are *achievable* and the task level *performance* is good. Using the assigned time for each task as input, an *assignment model* can be applied to generate a detailed *terminal operating plan (TOP)*. Developing a terminal operating plan can be considered as a planning stage. The TOP provides a whole picture in terms of what will be happening in the terminal in the planning shift.

For terminals with PMAKE functions already developed, an aggregate level planning method, named *real-time PMAKE analysis*, can be applied. Using this method, the best connection performance can be estimated and useful information provided to arrange the sequence of tasks in each process.

A major contribution of this thesis is developing the framework of TOP and the concept of *achievability* of TOP and measurement techniques. Different specifications about the time required to perform individual tasks will form different plans. It is possible that several terminal operating plans are available to the terminal managers. The conventional selection criteria may be the one with highest expected service performance. This thesis provides another measurement, the achievability of the terminal operating plan. Two dimensions, projected performance and achievability, are used to choose a plan among several alternatives. The idea is that when selecting a plan, it should not only have good projected service performance, but also be achievable. There is a trade-off between achievability and the performance, but it should be possible to choose a plan with both satisfactory performance and achievability.

The selected operating plan can be used as a tool to manage and control terminal tasks in the planning shift. The tasks in the plan can be assigned to terminal crews and inspectors with time requirements to perform the tasks.

The thesis presents a framework for improving terminal service performance. The central part of the approach is the *terminal operating plan.* The terminal operating plan can not only be used by terminal managers and terminal crews, but also be used by system level managers to evaluate the work of terminals. The following figure can be used to demonstrate the usefulness of TOP. Each major terminal in a rail system can have a terminal operating plan for each shift. The TOP can be used by the terminal manager to plan, manage, and control terminal operations in the shift. From the TOP, a list of tasks for each terminal crew can be obtained. This list may be provided to the terminal crews to make them know all the tasks and the time requirements for each task in the shift. From the terminals' TOP and the projected performance and achievability measures, the system level manager can compare and evaluate the works of these terminals and terminal managers. From a time-space point of view, the TOP provides a predicted inventory for the next shift. From the system point of view, the departure times from the TOP provides predicted train schedules. From this point of view, the TOP is connected with line haul operations and performance and the operations of connected terminals. The TOP and line haul models can be jointly used to predict ETAs for the connected terminals. The TOP can connect all the related persons in the terminal together because the different terminal tasks are performed by different persons from different departments of the terminal. Also, the TOP can connect terminals, line segments and other system components together. TOP provides a new approach to address system service performance issues.

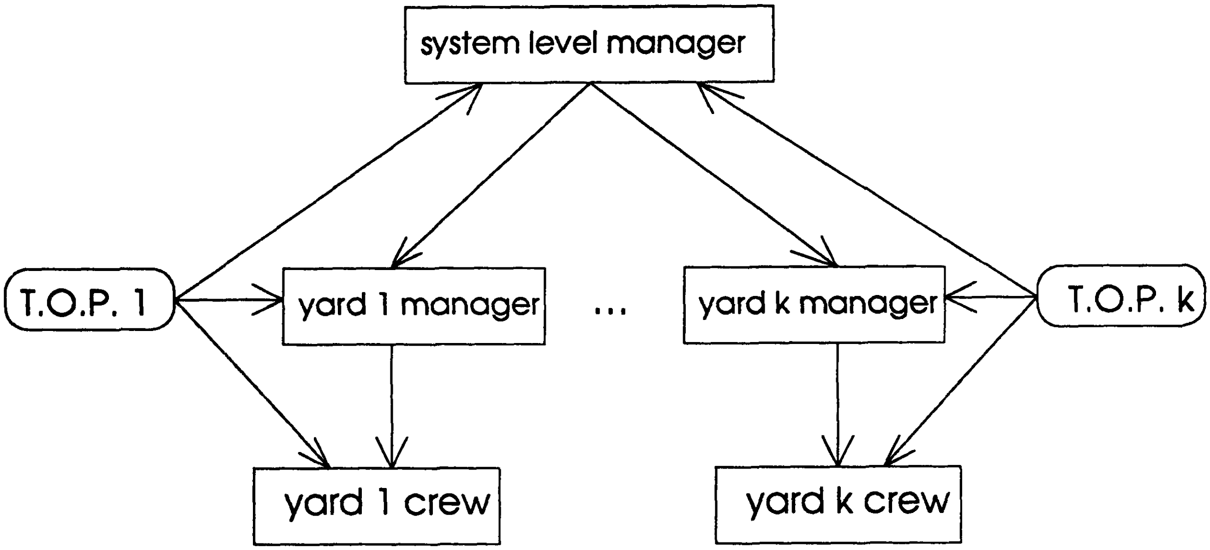

Figure 14. Applications of TOPs

6.2 Future Studies

There are many issues that have not been addressed enough which may be left for future studies. Some of them are as follows:

The first is the implementation of the assignment model that is to generate TOPs. In this thesis, only a model formulation is given. The implementation problem is not addressed.

In the case study, the block and sequence of cars information for each inbound train and outbound train were not available. This (along with time constraints) prevents an empirical application of the assignment model. In the future, this information could be collected and the assignment model could be used to compare the actual performance with the model results to see if the model gives reasonable terminal operating plans. In the assignment model, the pure processing time plus a reasonable buffer time is considered. The problem is how long this buffer time should be. To address this problem, the terminal managers' experience and more detailed data for the activities (including nonproductive activities such as walking time between two successive inspection processes) are needed.

The second direction is related with the implementation of TOP. TOP is a plan, which is to be implemented. During the implementation process, it is possible that some tasks may not be accomplished in the assigned time window. If this situation happens, what the terminal managers should then do is not addressed in this thesis. From the assignment model results, it is possible to conduct a sensitivity analysis, which may be very helpful to assist terminal managers to decide what to do next. For example, the limits within which the current TOP is still optimal may be obtained. If the time performing some tasks does not exceed these limits, the current plan is still optimal and the terminal managers can continue to implement the plan. On the other hand, if the time does exceed these limits, new plan may be needed. Under this situation, there is a need to use the assignment model to generate another operating plan according to the changed situation (since some tasks may have already accomplished). In terms of implementation, such issues may be addressed in the future.

The third direction is the calibration and field test of the models presented in this thesis. A case study is presented in this thesis. But the data applied in the case study is not enough to verify these models. In the future, more data should be collected to conduct comprehensive calibration and field test. By doing this, it provides a great opportunity to apply these models in practice.

References

Archibald, Russell D. and Villoria, Richard L., "Network-Based Management Systems PERT/CPM) Appendix C An analytical study of the PERT assumptions pp. 447-464", John Wiley & Sons, INC. 1967

Boffey, T. B., "Graph Theory in Operations Research, Chapter 6 pp121-144", The Macmillan Press LTD, 1982

Daganzo, Carlos F., "Static Blocking at Railroads: Sorting Implications and Track Requirements", *Transportation Science*, Vol. 20. No. 3, Aug. 1986

Daganzo, Carlos F., Dowling, Richard and Hall, Randolph, "Railroad Classification Yard Throughput: the Case of Multistage Triangular Sorting", *Transportation Research-A* Vol. 17A, 1983

Deming, W. Edwards, "Out of the Crisis", Cambridge, Mass: MIT, Center for Advanced Engineering Study, 1986

Duffy, Michael, "Statistical Process Control Applied to Rail Freight Terminal Performance: A Case Study of CSX's Radnor Yard", M.S. thesis, MIT, 1994

Ferguson, W. L., "Improving Railroad Terminal Control Systems: a Case Study of Southern Railway's Brosnan Yard", MIT CTS Report, 1979

Folk, Joseph F., "Some Analyses of Railroad Data", MIT Report R72-41, 1972

Folk, Joseph F., "A Brief Review of Various Network Models", MIT Report R72-42, 1972

Folk, Joseph F., "Models for Investigating Rail Trip Time Reliability", MIT Report R72-40, 1972

Haghani, Ali E., "Formulation and Solution of a Combined Train Routing and Makeup, and Empty Car Distribution Model", *Transportation Research-B* Vol. 23B, 1989

Kerr, P.A., Martland, C.D., Sussman, J.M., and Philip, C.E., "Models for Investigating Train Connection Reliability at Rail Classification Yards", MIT Report R76-44, 1976

Kulman, B.C., "A Model of Rail-Truck Competition in the Intercity Freight Market", M.S. thesis, MIT, 1974

Lang, A.S. and Martland, C. D., "Reliability in Railroad Operations", MIT Report R72-74, 1972

Little, Patrick and Martland, Carl D., "Causes of Unreliable Service in North American Railroads", Proceedings of the 1993 Transportation Research Forum, Nov., 1993

Little, Patrick, Oh Kyoung Kwon, and Martland, Carl D., "An Assessment of Trip Times and Reliability of Boxcar Traffic", MIT/AAR Affiliated Laboratory Working Paper 91-3

Martland, Carl D., "Improving Railroad Reliability: a Case Study of the Southern Railway", MIT Report R74-28, 1974

Martland, Carl D., "PMAKE Analysis: Predicting Rail Yard Time Distributions Using Probabilistic Train Connection Standards", *Transportation Science*, Vol. 16. No. 4, Nov. 1982

Martland, Carl D., H.S. Marcus, and G.B. Raymond, Jr., "Improving Railroad Terminal Control Systems: Budgeting Techniques, Probabilistic Train Connection Analysis and Microcomputer Applications", MIT CTS Report, 1983

Martland, C.D., M.G. Messner, and V. Nowicki, "Implementing an Operating/Service Plan: a Case Study of the Boston & Maine", MIT CTS Report, 1979

Martland, Carl D., Little, Patrick, Oh Kyoung Kwon, and Rajesh Dontula, "Background on Railroad Reliability", *AAR Technical Report R-803*, Association of American Railroads, March, 1992

Martland, C.D. and C.D. Van Dyke, "Transferring the MIT Service Planning Model", MIT CTS Report, 1983

McCarren, J.R. and C.D. Martland, "The MIT Service Planning Model", MIT CTS Report, 1979

Moder, Joseph J. and Philips, Cecil R., "Project Management with CPM and PERT, Chapter 11 pp.271-317 second edition", Van Nostrand Reinhold Company 1970

Morlok, Edward K. and Peterson, R.B., "Railroad Freight Train Scheduling: A Mathematical Programming Formulation", Research Report of the Transportation Center, Northwestern University, 1970

Peterson, E. R., "Railyard Modeling: Part I. Prediction of Put-Through Time", *Transportation Science*, Vol. 11. No. 1, Feb. 1977

Peterson, E. R., "Railyard Modeling: Part II. the Effect of Yard Facilities on Congestion", *Transportation Science*, Vol. 11. No. 1, Feb. 1977

Reebie Associates, "Toward on Effective Demurrage System", Prepared for U.S. Department of Transportation, Federal Railroad Administration, 1972

Reid, Robert M., O'Doherty, John D., Sussman, Joseph M., and Lang, A. Scheffer, "The Impact of Classification Yard Performance on Rail Trip Time Reliability", MIT Report R72-39, 1972

Rothberg, S. C., J.M. Sussman, and C. D. Martland, "The Design of a Management Control System for Railroad Freight Terminals", MIT CTS Report 1980

Siddiqee, M. W., "Investigation of Sorting and Train Formation Schemes for a Railroad Hump Yard", *Traffic Flow and Transportation*, G. F. Newell (ed.), American Elsevier, New York, 1972

Sussman, J.M. and Martland, C. D., "Improving Railroad Reliability: a Case Study of the Southern Railway, Executive Summary", MIT Report R74-29, 1974

Sussman, J.M., Martland, C. D., and Lang, A.S., "Reliability in Railroad Operations: Executive Summary", MIT Report R73-4, 1972

Temple, Barker, and Sloane, "Survey of 125 Shippers Using Both Truck and Rail modes", 1989, 1990

Troup, K.F., (ed) "Railroad Classification Yard Technology: An Introductory Analysis of Functions and Operations", USDOT, FRA, FRA/ORD-75-55, 1975

Tsisiklis, John, "Multilevel Advanced Train Control Systems", Research conducted for Draper Labs, Cambridge, MA, 1992-3

Tykulsker, R.J., "Railroad Terminals: Operations, Performance and Control", M.S. thesis, MIT, 1981

Vieira, Luiz, "The Value of Service in Freight Transportation", Ph.D Dissertation, MIT, 1991

Yagar, S., Saccomanno, F. F., and Shi, Q., "An Efficient Sequencing Model for Humping in a Rail Yard", *Transportation Research-A* Vol. 17A, 1983

Appendix A: PERT/CPM Calculation

The calculation steps are as follows:

(1). Earliest start time

From the operating plan:

t_E (1) = 0820;

t_E (2) = 0830;

t_E (3) = 0900;

t_E (6) = 1005

Using PERT/CPM method:

$t_E (4) = \max\{t_E (2) + t(2, 4), t_E (1) + t(1, 4)\} = \max\{0830 + 29, 0820 + 37.5\} = 0859;$

$t_E (5) = \max\{t_E (4) + t(4, 5), t_E (3) + t(3, 5)\}$

$= \max\{0859 + 37.5, 0900 + 29\} = 0936.5;$

$t_E (7) = t_E (5) + t(5, 7) = 0936.5 + 45\ 1021.5;$

$t_E (9) = t_E (7) + t(7, 9) = 1021.5 + 35 = 1056.5;$

$t_E (8) = t_E (7) + t(7, 8) = 1021.5 + 45 = 1106.5;$

$t_E (10) = \max\{t_E (5) + t(5, 10), t_E (6) + t(6, 10), t_E (8) + t(8, 10)\}$

$= \max\{0936.5 + 37.5, 1005 + 30, 1106.5 + 0\} = 1106.5;$

$t_E (11) = t_E (8) + t(8, 11) = 1106.5 + 35 = 1141.5;$

$t_E (12) = t_E (10) + t(10, 12) = 1106.5 + 40 = 1146.5;$

$t_E (13) = \max\{t_E (8) + t(8, 13), t_E (12) + t(12, 13)\}$

$= \max\{1106.5 + 0, 1146.5 + 45\} = 1231.5;$

$t_E (14) = t_E (13) + t(13, 14) = 1231.5 + 35 = 1306.5$

(2). Latest start time

$t_L(14) = t_E (14) = 1306.5;$

$t_L(11) = t_E (11) = 1141.5;$

$t_L(9) = t_E (9) = 1056.5;$

$t_L(13) = t_L(14) - t(13, 14) = 1306.5 - 35 = 1231.5;$

$t_L(12) = t_L(13) - t(12, 13) = 1231.5 - 45 = 1146.5;$

$t_L(10) = t_L(12) - t(10, 12) = 1146.5 - 40 = 1106.5;$

$t_L(8) = \min\{t_L(13) - t(8, 13), t_L(11) - t(8, 11), t_L(10) - t(8, 10)\}$

$= \min\{1231.5 - 0, 1141.5 - 35, 1106.5 - 0\} = 1106.5;$

$t_L(6) = t_L(10) - t(6, 10) = 1106.5 - 30 = 1036.5;$

$t_L(7) = \min\{t_L(9) - t(7, 9), t_L(8) - t(7, 8)\}$

$= \min\{1056.5 - 35, 1106.5 - 45\} = 1021.5;$

$t_L(5) = \min\{t_L(7) - t(5, 7), t_L(10) - t(5, 10)\}$

$= \min\{1021.5 - 45, 1106.5 - 37.5\} = 0936.5;$

$t_L(4) = t_L(5) - t(4, 5) = 0936.5 - 37.5 = 0859;$

$t_L(3) = t_L(5) - t(3, 5) = 0936.5 - 29 = 0907.5;$

$t_L(2) = t_L(4) - t(2, 4) = 0859 - 29 = 0830;$

$t_L(1) = t_L(4) - t(1, 4) = 0859 - 37.5 = 0821.5$

From the algorithm, 2 to 4, 4 to 5, 5 to 7, 7 to 8, 7 to 9, 8 to 10, 8 to 11, 10 to 12, 12 to 13, and 13 to 14 are critical paths for the outbound trains in the operating plan A (the darker block lines in the graph).

For OB #1: $\mu_{OB\#1} = t(b) + t(d) + t(g) + t(i) = 29 + 37.5 + 45 + 35 = 146.5;$

$\sigma^2_{OB\#1} = \sigma^2_{I_3} + \sigma^2_{C_3} + \sigma^2_{A_1} + \sigma^2_{O_1} = 9 + 25 + 25 + 16 = 75;$

AVAIL = 30 + 40 + 50 + 40 = 160;

$$P_{OB\#1} = \Phi(\frac{160-146.5}{\sqrt{75}}) = \Phi(1.57) = 0.9418$$

For OB #2: $\mu_{OB\#2} = t(b) + t(d) + t(g) + t(h) + t(k) = 29 + 37.5 + 45 + 45 + 35$

$= 191.5;$

$\sigma^2_{OB\#2} = \sigma^2_{I_3} + \sigma^2_{C_3} + \sigma^2_{A_2} + \sigma^2_{C_2} = 9 + 25 + 25 + 16 = 75;$

AVAIL = 30 + 40 + 50 + 50 + 37 = 207;

$$P_{OB\#2} = \Phi(\frac{207-191.5}{\sqrt{75}}) = \Phi(1.79) = 0.9633$$

For OB #2: $\mu_{OB\#3} = t(b) + t(d) + t(g) + t(h) + t(j) + t(l) + t(m)$

$$= 29 + 37.5 + 45 + 45 + 40 + 45 + 35 = 276.5;$$

$$\sigma^2_{OB\#3} = \sigma^2_{I_3} + \sigma^2_{C_3} + \sigma^2_{A_3} + \sigma^2_{O_3} = 9 + 25 + 25 + 16 = 75;$$

$$\text{AVAIL} = 30 + 40 + 50 + 50 + 40 + 50 + 45 = 305;$$

$$P_{OB\#3} = \Phi(\frac{305-276.5}{\sqrt{75}}) = \Phi(3.29) = 0.9995$$

Achievability of the operating plan:

$$A = \frac{\sum a_i * P_i}{\sum a_i} = \frac{100 * 0.9418 + 100 * 0.9633 + 100 * 0.9995}{100 + 100 + 100} = 0.968$$

Appendix B: Distribution Method Calculation

Task level achievability:

For a specific task i, if its mean is μ_i standard deviation is σ_i, and the assigned time to perform the task is $ASSIGN_i$, the achievability of performing this task is:

$$P_i = \Phi(\frac{ASSIGN_i - \mu_i}{\sigma_i})$$

From the plan of Table 2 and the pure processing times given in PERT/CPM section of chapter four, the task level achievabilities can be estimated as follows:

Inbound inspection process tasks:

$$P_b = \Phi(\frac{ASSIGN_b - \mu_b}{\sigma_b}) = \Phi(\frac{30-29}{3}) = 0.63$$

$$P_c = \Phi(\frac{ASSIGN_c - \mu_c}{\sigma_c}) = \Phi(\frac{35-29}{3}) = 0.98$$

$$P_f = \Phi(\frac{ASSIGN_f - \mu_f}{\sigma_f}) = \Phi(\frac{35-30}{3}) = 0.95$$

Classification process tasks:

$$P_a = \Phi(\frac{ASSIGN_a - \mu_a}{\sigma_a}) = \Phi(\frac{40-37.5}{5}) = 0.69$$

$$P_d = \Phi(\frac{ASSIGN_d - \mu_d}{\sigma_d}) = \Phi(\frac{40-37.5}{5}) = 0.69$$

$$P_e = \Phi(\frac{ASSIGN_e - \mu_e}{\sigma_e}) = \Phi(\frac{45-37.5}{5}) = 0.93$$

$$P_j = \Phi(\frac{ASSIGN_j - \mu_j}{\sigma_j}) = \Phi(\frac{40-40}{5}) = 0.50$$

Assembly process tasks:

$$P_g = \Phi(\frac{ASSIGN_g - \mu_g}{\sigma_g}) = \Phi(\frac{50-45}{5}) = 0.84$$

$$P_h = \Phi(\frac{ASSIGN_h - \mu_h}{\sigma_h}) = \Phi(\frac{50-45}{5}) = 0.84$$

$$P_l = \Phi(\frac{ASSIGN_l - \mu_l}{\sigma_l}) = \Phi(\frac{50-45}{5}) = 0.84$$

Outbound inspection process tasks:

$$P_i = \Phi(\frac{ASSIGN_i - \mu_i}{\sigma_i}) = \Phi(\frac{40-35}{5}) = 0.84$$

$$P_k = \Phi(\frac{ASSIGN_k - \mu_k}{\sigma_k}) = \Phi(\frac{37-35}{5}) = 0.66$$

$$P_m = \Phi(\frac{ASSIGN_m - \mu_m}{\sigma_m}) = \Phi(\frac{45-35}{5}) = 0.98$$

Process level achievability:

As discussed in chapter three, a specific process j's achievability is:

$$P_j = \Phi(y_j) = \Phi(\frac{\sum_i t_i - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}})$$

where, μ_i is task i's mean, σ_i is task i's standard deviation, and t_i is the assigned time length for task i.

Inbound inspection process:

$$P_I = \Phi(y_I) = \Phi(\frac{\sum_i t_i - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}}) = \Phi(\frac{30+35+35-29-29-30}{\sqrt{9+9+9}}) = 0.99$$

Classification process:

$$P_C=\Phi(y_C)=\Phi(\frac{\sum_i t_i-\sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}})$$

$$=\Phi(\frac{40+40+45+40-37.5-37.5-37.5-40}{\sqrt{25+25+25+25}})=0.89$$

Assembly process:

$$P_A=\Phi(y_A)=\Phi(\frac{\sum_i t_i-\sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}})$$

$$=\Phi(\frac{50+50+50-45-45-45}{\sqrt{25+25+25}})=0.96$$

Outbound inspection process:

$$P_O=\Phi(y_O)=\Phi(\frac{\sum_i t_i-\sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}})$$

$$=\Phi(\frac{40+37+45-35-35-35}{\sqrt{25+25+25}})=0.98$$

When calculating the process level achievability, all the tasks in the process are considered together. This may allow to assign extra time left from a previous task to the successive tasks. So, the estimated process level achievability is comparatively high.

Overall plan level achievability:

The overall plan's achievability is the product of the four process level achievabilities. That is:

$$P=\prod_{j} P_j = 0.99 * 0.89 * 0.96 * 0.98 = 0.83$$

Since this method assumes that the processes are independent, this method tends to underestimate the overall plan's achievability.