THE IMPACT OF OPERATIONAL STRATEGIES AND NEW TECHNOLOGIES ON
RAILROAD CAPACITY

BY

MARK H DINGLER

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Advisers:

Professor Christopher P.L. Barkan

ABSTRACT

North American freight railroads are expected to face increasing capacity constraints due to substantial, expected long-term growth in traffic. In order to meet this forecasted demand railroads must efficiently use existing capacity and effectively plan new capacity. Infrastructure to provide this capacity is capital-intensive so careful consideration of alternatives to infrastructure expansion must be considered. Consequently, railroads need to understand how operational practices and new technologies may affect rail line and network capacity.

A comprehensive literature review of previous work on railroad capacity was conducted and the various metrics used to measure it described. An assessment of the major Class 1 railroads' capacity planning methods was conducted and their research needs identified. Operational factors influencing capacity are identified and described. Rail capacity is often measured using train delay as the metric so the categories and sources of delay were evaluated.

Train type heterogeneity is a significant factor affecting railroad operating capacity. The relationship between delay, traffic volume and train type heterogeneity was investigated in a series of experiments using simulation analysis of trains operating on a single-track rail line. The specific types of conflicts and operational factors affecting delay were identified and quantified. Various operational and infrastructure methods to reduce train delay were analyzed and cost benefit analyses were conducted to determine their relative cost effectiveness. A qualitative analysis of the impact of positive train control (PTC), communications based train control (CBTC) and electronically controlled pneumatic (ECP) brakes was conducted. Each aspect of these technologies with the potential to affect capacity was identified and its effect evaluated under various implementation scenarios, including consideration of the conditions under which each technology has the potential to increase, reduce, or have no effect on capacity.

To Mom and Dad

For Pushing Me to Achieve

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The ability of the railroads to efficiently move goods is vital to the North American economy. However, freight railroads are expected to face increasing capacity constraints due to substantial, expected long-term growth in demand *(Cambridge Systematics 2007)*. In order to meet this forecasted demand railroads must efficiently use and effectively plan new capacity. To accomplish this railroads must understand how operational practices and new technologies will affect capacity on the North American freight rail network.

The U.S. Class 1 railroads moved a total of 1.7 trillion ton-miles of freight in 2008 *(AAR 2008a)* and this is expected to increase over 80% by 2035 *(AASHTO 2007)*. Additionally, the U.S. plans to introduce new, higher speed passenger services and expanded commuter rail operations across the country. Passenger operations disproportionately use capacity reducing the ability to operate existing and future freight traffic *(AREA 1921, Mostafa 1951, Harrod 2009)*. Recently, positive train control (PTC) systems have been mandated on many routes and communications based train control (CBTC) is the type of PTC system most likely to be implemented to meet the mandate. CBTC alone and in combination with another new technology, electronically control (ECP) brakes, have the potential to affect capacity.

Understanding how increased traffic, passenger trains, and new technology will impact capacity is necessary for effective capacity planning. Expanding infrastructure to meet these capacity constraints is time consuming and capital intensive. On the other hand, operational changes are more flexible and rapidly implemented *(Lai and Barkan 2009)*. Additionally, an understanding of operations is necessary to efficiently plan the investment in capacity projects that are required.

As demand for freight rail services increases, understanding how to increase capacity of single track routes will be a major focus for railroads. Currently over 80% *(AAR 2008a)* of American railroad mainlines are single track and many of these routes will require either infrastructure expansion or operational changes to meet the growing demand.

1.1 Objectives and Scope

The objectives of this research are to better understand the principal operational factors influencing railroad capacity on single track routes. Accomplishing this required consideration of the complexities of railroad capacity and the different metrics and units of capacity used by the industry. Interviews with Class 1 Railroads were conducted to understand the current capacity planning practices used by the industry and their research needs. Simulation, research was conducted to study the impact of various operational factors on railroad capacity. With this knowledge, alternatives were evaluated in order to improve operations and reduce heterogeneity. Lastly, railroads are increasingly turning to new technologies to improve efficiency. Two highly touted technologies, CBTC and ECP brakes, were evaluated for their potential to impact capacity.

1.2 Organization

This thesis is divided into three sections, an overview of railroad capacity, research on the effect of operations, and the impact of CBTC and ECP brakes on capacity. Chapters two and three offer an overview of railroad capacity and current capacity planning practices. Chapters four through seven focus on the impact of operational practices on capacity and potential methods to

improve operations. Chapter eight considers the impact of CBTC and ECP brakes on capacity. Chapter 9 discusses conclusions that can be drawn from this work and areas of future work.

1.2.1 Chapter 2

This chapter contains a discussion of railroad capacity. The different factors that influence capacity and the methods used to increase it are identified. Rail capacity can be defined and measured in a number of different ways. To provide a framework for later chapters, rail capacity is defined based on location, calculation and utilization. Finally, the metrics used to measure capacity are analyzed for their advantages and disadvantages.

1.2.2 Chapter 3

This chapter describes the current capacity analysis and planning practices of the major North American railroads. By interviewing representatives from each of the class1 railroads the specific tools and methodologies employed by each of the railroads were identified. Additionally the railroads identified future work that they consider would be most helpful to the industry.

1.2.3 Chapter 4

This chapter presents a discussion of the impact of operations on railroad capacity. The operational factors that influence capacity are identified and described. A literature review of previous research on railroad operations was conducted identifying train type heterogeneity as a significant factor influencing rail operations and an area needing additional research. Train delay

is the primary unit used to measure the impact of railroad operations. To better understand this

important metric the categories and sources of delay are discussed.

1.2.4 Chapter 5

This chapter investigates the impact of train type heterogeneity on a single track line. Using

simulation the relationship between delay, volume and heterogeneity was investigated. To better

understand the operational causes that contribute to heterogeneity caused delays, a series of

experiments were conducted using simulations to isolate the various factors and assess their

relative importance alone and in combination. A paper based on this chapter was published in

the Transportation Research Record *(Dingler et. al. 2009a)*.

1.2.5 Chapter 6

This chapter is an investigation into the specific factors causing train delay. The specific

conflicts or operational factors that result in the increased delay due to heterogeneous traffic are

not known. Using simulation, the train performance data of each train was analyzed to identify

the specific reason for each delay. The delays were categorized by type of conflict and specific

train actions. Material from this chapter was presented at the 2010 Joint Rail Conference (JRC)

and 2010 American Railroad Engineering and Maintenance-of-Way (AREMA) Annual

Conference and included in the proceedings of the latter *(Dingler et. al. 2010a)*.

1.2.6 Chapter 7

This chapter is an analysis of various methods to reduce train delay on a single track line.

Simulation software was used to evaluate multiple scenarios for their effectiveness. The

different methods reduce train delays but require additional infrastructure and equipment costs. These costs and benefits can be considered to determine the cost effectiveness of each method. Material from this chapter was presented at the 2009 Joint Rail Conference (JRC) and 2009 American Railroad Engineering and Maintenance-of-Way (AREMA) Annual Conference and included in the proceedings of the latter *(Dingler et. al. 2009b)*.

1.2.7 Chapter 8

This chapter is a qualitative analysis of the impact of positive train control, communications based train control and electronically controlled pneumatic brakes. A literature review of articles, papers, reports and regulations pertaining to each of these technologies was conducted. Each of the elements with the potential to affect capacity was identified and its effect evaluated under various implementation scenarios. Consideration was given to the conditions under which each technology has the potential to increase, have no effect or decrease capacity. Material from this chapter was presented at and included in the proceedings of the 2009 AREMA Annual Conference *(Dingler et. al. 2009c)* and a later version was published in the Transportation Research Record *(Dingler et. al. 2010b)*.

CHAPTER 2: INTRODUCTION TO RAILROAD CAPACITY

The principal service freight railroads offer is movement of goods from an origin to a destination. Railroad capacity directly affects their ability to provide this service in a timely, reliable and economical manner. The amount of capacity railroads provide is a complicated financial decision. Insufficient capacity increases travel times, which increases operating cost and reduces service quality and reliability thereby reducing demand for some commodities *(Weatherford et. al.. 2008)* and hence revenue. On the other hand, building and maintaining excess capacity is inefficient use of resources. It can weaken the profitability of a railroad and discourage outside investment. Since excess capacity serves to improve railroad service and reliability, capacity is ultimately a function of a shipper's willingness to pay for a required level of service. Consequently, without consideration of economics, the concept of capacity is meaningless *(Congressional Research Service 2007)*.

Railroad capacity is influenced by a complex relationship of many factors. These can be grouped into six categories: infrastructure, operations, motive power, rolling stock, maintenance and human resources. Infrastructure factors include the amount and quality of the trackage and the geography of the route. Operational factors include the type and scheduling of traffic (Table 2.1). Motive power and rolling stock are other important factors in railroad capacity because insufficient locomotives and cars interferes with building and operating the trains needed to handle the traffic. Maintenance of mechanical equipment and infrastructure reduces the availability of the assets; however, deferring maintenance can reduce reliability. Lastly a sufficient number of trained and qualified personnel are necessary to perform a variety of tasks and functions required for safe and efficient railroad operation.

TABLE 2.1: Infrastructure and Operational Parameters of Railroad Capacity

Factors	Parameters
Infrastructure	Number of Tracks
	Siding/Crossover Spacing
	Siding Length
	Traffic Control System
	Signal System
	Track Speed Limits
Operations	Train Speed
	Train Mix
	Train Priorities
	Stability
	Traffic Patterns
	Level of Service Requirements
	Train Size and Tonnage
	Train Power to Ton Ratios

According to the Theory of Constraints (TOC) any system is limited by a small number

of constraints *(Goldratt 1990)*. Much like a chain where the weakest link limits the strength,

constraints, often called bottlenecks, limit the system as a whole. On a railroad network there is

often sufficient capacity on most of the railroad network, but bottlenecks at certain locations or

in certain functions prevent utilization of all the available capacity *(McClellan 2007)*. On some

routes this bottleneck may be the traffic control system or an interlocking, while on others it may

be an insufficient number of locomotives to power the required number of trains. Although there

are numerous exceptions, previous research has found that congestion at terminals is often the

principal bottleneck on many freight railroad networks *(Dirnberger 2005, Laurits R. Christensen*

Associates, Inc 2009).

Understanding the available and maximum capacity of a route is important to identifying

and eliminating bottlenecks and determining the impact of additional traffic. Unfortunately, no

single metric of capacity fully captures the complexity of rail performance and therefore no

standard definition or measure of rail system capacity exists. As a result, measures of rail system capacity are ultimately a function of the assumptions made by the analyst *(Congressional Research Service 2007, Weatherford et. al. 2008).*

While it is difficult to precisely calculate the maximum capacity of a line, routes approaching capacity demonstrate certain attributes. When a route becomes capacity constrained the fluidity of the network will decline resulting in longer travel times. Increased congestion also reduces the resilience to unexpected events (e.g., bad weather, mechanical or infrastructure failures, unexpected growth in demand). Insufficient capacity can cause these disruptions to result in widespread and prolonged congestion *(Saunders 2003, Congressional Research Service 2007, Weatherford et. al. 2008).*

There are a number of ways for a railroad to address capacity constraints. Potential strategies include *(McClellan 2007, Congressional Research Service 2007)*:

1. **More efficiently use available resources.** Without adding any additional resources the railroad may be able to quickly and inexpensively increase the utilization of their current resources. Methods to improve efficiency include:

    A. *Reschedule traffic.* More efficiently scheduled traffic can reduce delays and increase density. If headways can be reduced, the density of trains on the system will increase. This can be achieved through fleeting trains or effectively using available train slots. Evenly spreading trains throughout the day reduces delays and allows more efficient use of terminal resources.

B. *Change operations.* Operational changes include changing speeds, changing priorities, distributed power and changing the power to trailing ton ratios of trains. On some routes higher speeds permit a railroad to move more traffic with the same amount of equipment, while on others this can increase conflicts and reduce capacity. Changing priorities permits more efficient meets and passes. Distributed power permits the use of longer trains, reducing the required number of trains. Finally, greater power to ton ratios allow for faster acceleration, more rapid stopping due to greater dynamic braking power and may permit higher speeds.

C. *Utilize technological improvements.* Technology has long had a role in increasing rail system capacity and productivity. Some new technologies that have the potential to increase railroad productivity include: automated condition monitoring systems, wayside detectors, cab signaling, computer aided dispatching (CAD), electronically controlled pneumatic (ECP) brakes and communications based train control (CBTC).

2. **Add resources.** Railroads can address capacity constraints by increasing the available resources. This may require long lead times, is capital intensive and depending on the resources may be inflexible to changing demands. Resources can be added in the following ways:

A. *Purchase additional locomotives or rolling stock.* Additional equipment can be used to replace or supplement aging equipment, added to current trains or used to operate additional trains. Aging cars and locomotives are often less reliable and have lower

carrying capacities and horsepower, respectively. Replacing this aging equipment can increase the carrying capacity and pulling power of a train without any change in its handling. If added to an existing train, additional locomotives allow for faster acceleration and higher speeds. Lastly, adding additional railcars to create longer trains permits the delivery of more cargo with fewer trains, releasing capacity for new traffic.

B. *Expand or improve infrastructure*. Strategically constructed infrastructure eliminates bottlenecks and increases the fluidity of operations. Some examples include adding sidings or an additional track on heavily-used rail corridors, straightening curves that cause trains to slow down, improving the track and structures to permit faster and heavier trains, and expanding or building new rail yards and intermodal terminals. Traffic control systems such as signals or centralized traffic control (CTC) can also be added to increase capacity.

3. **Shed traffic.** When facing capacity constraints railroads may shed low margin or disruptive traffic. Sometimes a small minority of trains can create widespread congestion on a network. Removing this traffic may free capacity for additional, less disruptive traffic. For example, although lucrative, highly schedule-sensitive traffic may reduce profits because of increased operating expenses resulting from the extra congestion due to its operation *(Weatherford et. al. 2008).*

4. **Reduce the quality of service.** Increased congestion results in lower levels of service; however it may be in a railroad's best interest to accept these lower service levels. This may result in some loss of traffic, but it may be unprofitable for a railroad to expand capacity for short periods of peak demand.

2.1 Capacity Definitions

Railroad capacity is a complex relationship of many factors and service demands and to date no generally accepted definition of railway capacity exists. In its most general form railroad capacity can be defined as, *the ability to move a specific amount of traffic with acceptable punctuality*. This can be more specifically defined based on location (line, network or terminal), calculation (theoretical or practical) and utilization (maximum, used or available). These terms are often used synonymously but refer to different aspects of capacity and need to be clearly defined in any discussion. The capacity definitions based on location are:

Network capacity: *The ability to move a specific amount of traffic on a given network with acceptable punctuality.* Many discussions about railroad capacity are in reference to network capacity. Rail networks include both terminals and the lines between them. Therefore, network capacity is dependent on the capacity of both of these elements.

Terminal capacity: *The amount of traffic a terminal can process in a given time interval.*

Line capacity: *The ability to move a specific amount of traffic over a defined route with acceptable punctuality.*

11

There are two commonly used terms when calculating capacity: theoretical capacity and practical capacity. These types of capacity are defined as follows.

Theoretical capacity: *The maximum traffic physically possible under ideal conditions.* Theoretical capacity assumes even temporal spacing throughout the day, with uniform train characteristics and priorities. When calculating theoretical capacity, the time for service disruptions and maintenance are ignored.

Practical Capacity: *The maximum traffic possible when accounting for actual conditions and achieving a reasonable level of reliability.* Practical capacity is the capacity experienced during day-to-day operations with sustainable operations. Practical capacity is usually around 60-75% of theoretical capacity (Kraft, 1982; Kruger, 1999).

The maximum capacity of a line is useful; however capacity planners are also interested in capacity utilization and availability. The types of capacity by utilization are defined as follows *(Kruger 1999, Abril et. al. 2007)*:

Used Capacity: *The actual traffic volume operated while accounting for normal variations in traffic and operations.*

Available Capacity: *The difference between used and maximum capacity*. It is an indication of the additional traffic volume that could be handled while maintaining a predefined performance threshold.

2.2 Metrics of Railroad Capacity

The most intuitive way to calculate railroad capacity is the maximum number of trains that can be operated over the network in a given period of time. The concept seems simple but it becomes more complex when the many factors that influence capacity are considered *(Stok 2008)*. Throughput is dependent on the specific characteristics of the trains in question. In some cases, the level of service or asset utilization may be more important metrics of capacity. Each metric has specific applicabilities and weaknesses, and analyzing trends using a single metric fails to capture the complexity of rail performance *(Weatherford et. al. 2008)*. Within each metric there are several different units that can be used to understand and quantify that metric (Table 2.2). Each metric focuses on one aspect of railroad operations, and unfortunately they are not directly convertible. Instead each metric reveals something different and is important to different groups in the railroad industry.

TABLE 2.2: Different Metrics and Units of Railroad Capacity

Metric	*Units*
Throughput	Trains, Cars, Tons, Passengers
Level of Service	Terminal Dwell, Average Velocity, Delay
Asset Utilization	Average Velocity

2.2.1 Throughput

Throughput is a measure of how much material can be transported over a route in a specific period of time. This directly measures the movement of cargo but is variable based on the traffic and operations on a route. Throughput can be measured in trains, cars, tons or passengers.

The most common unit of throughput and capacity is trains. Trains use line capacity and direct measurement of the number of trains per unit time offers the ability to estimate the used and maximum capacity of a line. Accordingly, the number of trains per day is often used when discussing capacity on a route and has the advantage of being intuitive, easily measured and understandable by railroaders and the general public alike.

Another measure of throughput is cars. Using cars as a unit of throughput is most useful when considering terminals. Cars per unit time is often used as the metric for terminal performance and capacity; much like trains per day is used for line capacity. This unit is useful to terminal designers and operators since in terminals many different car types have similar handling requirements and capacity impacts. However, car length varies widely and may need to be accounted for in some aspects of terminal capacity. Regarding line capacity it measures the amount of cargo moved while accounting for the capacity impact of empty cars. Increasing the number of cars per train is a source of increased capacity. On single track lines, train length and therefore the number of cars per train, is dependent on the siding length. Without information on car length, cars alone is an insufficient metric to offer information on this capacity constraint. In general, measuring the number of railcars is not useful as a unit of line capacity because two trains with different numbers and sizes of cars may use similar amounts of capacity.

Tonnage is often used when considering the ability to move cargo over a route and is measured in terms of gross or revenue tons. Gross tonnage measures the total weight of all the

locomotives, rolling stock and revenue and non-revenue lading, while revenue tonnage only counts the weight of paid shipments. Gross tonnage is the simplest and most commonly used metric for traffic on a line. However its meaning in terms of operations is imprecise because of the wide variations in its relationship to operating parameters. For example a coal route will move more tonnage in fewer trains than a predominately intermodal route. Revenue tonnage is important to railroad marketing and financial personnel because it provides direct information on the movement and volume of traffic earning revenue for railroads. Revenue tonnage, like gross tonnage, is imprecise measurement of capacity due to the relationship to operating parameters.

Tonnage offers different information than can be gained from knowing the number of trains or cars. The capacity of a line can be increased without increasing the number of cars or trains by using higher capacity railcars or longer trains. While tonnage is useful for comparing traffic between lines, it is not very useful in estimating the used or maximum capacity of a line.

The last measure of throughput used is number of passengers. While passengers per unit time is not relevant for freight routes this measure is important on predominately passenger routes. Planners need to determine the mix of trains (high speed, express, commuter, intercity, etc.) that will provide service to the greatest number of people. Much like tonnage, using passengers measures the movement of the items that offer revenue for the service. However, maximum ridership does not always result in maximum revenue. Many intercity trains have different classes and on overnight trains different room sizes. Consequently, not all passengers use the same amount of capacity and like the other measures of throughput the capacity is dependent on the characteristics of the trains being used.

The critical limitation of using throughput as a capacity metric is its variability depending on the specific traffic and operations. Different train types use different amounts of capacity,

15

carry different amounts and types of cargo and have different numbers of cars. For instance a route with low speed bulk trains may carry more cars and tons than a route with high speed intermodal trains but have fewer trains per day.

Throughput is a strong metric of capacity because it is a direct measure of the movement of goods across a line or through a terminal, but has limitations since it is highly influenced by the train and car types. A more complete understanding of corridor throughput requires one to look at all the units together. While railroads attempt to maximize throughput, shippers demand a high level of service. Consequently, when calculating the maximum capacity of a route, throughput cannot be determined without considering the level of service.

2.2.2 Level of Service

Level of service is a measure of the reliability and timeliness of transportation. Excessively long travel times or unreliable deliveries will be unacceptable to shippers and passengers, but the specific level of service that is acceptable will vary widely depending on price, commodity and other factors. Level of service can be measured using terminal dwell, velocity and/or train delay. Level of service is an indirect measure of capacity because it only measures the effect of insufficient or excess capacity but not capacity itself.

Most traffic has scheduled arrival windows for each customer. Variability in travel time reduces the probability in arriving within the designated time window. Terminal dwell and velocity are measures of the travel time of traffic; high variability in these values reflects a low level of service *(Laurits R. Christensen Associates, Inc 2008)*. Terminal dwell is defined by the AAR *(2010)* as the average time a car resides at a specified terminal location. The measurement

begins when a customer releases a car for shipment, a car is received in interchange, or a train arrival event, and ends with a customer placement, interchange, or train departure event.

Velocity measures the line-haul movement between terminals. It is calculated by dividing train-miles by total hours operated, excluding yard and local trains, passenger trains, maintenance of way trains, and terminal time. Since velocity and terminal dwell are dependent on many factors and not comparable between railroads, it is changes in these values that are important. In general, declining average speeds or increased terminal dwell imply problems, and increasing average speed or reduced terminal dwell imply improvements *(Weatherford et. al. 2008)*.

Another unit of capacity that measures level of service is delay. Delay is the additional travel time required to traverse a defined route due to scheduled and unscheduled events. It is a summation of the length and number of late arrivals, therefore it directly measures the magnitude of the variability in travel times. Delay is also important because it is the primary output from many simulation models of railroad operations. Change in delay is often used by railroads to determine the benefit of a project.

A weakness of using delay as a capacity metric is that different shippers have different level-of-service requirements. Intermodal customers have highly time sensitive shipments while bulk cargo deliveries may be less time sensitive. Consequently the capacity of a route in terms of delay is dependent on the traffic mix. Additionally, level of service alone does not allow for estimation of the capacity of a network and offers no information on the movement of goods.

The strength of delay as a metric of capacity is that to shippers, level of service is the most important metric of capacity. A low level of service will result in lost business and therefore it is important to the railroads to have a high level of service. Another strength of delay

is the strong delay-volume relationship. While not always true *(White 2006)*, excess capacity often results in lower delays and a higher level of service *(AREA 1931),* while insufficient capacity has the opposite effect. Consequently, delay is often used to determine the impact of operational and infrastructure changes on a route.

2.2.3 Asset Utilization

A railroad's assets include the infrastructure, rolling stock, motive power and personnel. Railcars and locomotives are costly to purchase and maintain and therefore efficient use of them is critical to economical operation. A railroad can move sufficient cargo at a high level of service but still be underutilizing its assets. Asset utilization, like level of service, is an indirect measure of capacity.

The principal metric for system-wide asset utilization is average velocity. Given a constant amount of traffic, an increase in velocity indicates shorter cycle times and thus more efficient use of assets. One major railroad estimates that an increase of one mile per hour in average velocity corresponds to an additional 250 locomotives, 5,000 freight cars, and 180 train and engine employees available to move additional traffic *(Hamburger 2006)*.

Unfortunately system velocity is a poor measure of capacity. Velocity is highly dependent on the number and type of trains being operated. The other disadvantage of using velocity is that it does not provide any information on the quality of service. Two trains may have the same average speed but the service demands of one train type may be greater than the other.

Velocity is useful for measuring asset utilization and understanding train run times. However, if a train has a scheduled average velocity of 10 mph and travels at that speed, while there is no delay, level of service will be low and the assets will be inefficiently utilized.

2.3 Discussion

Railroad capacity is a complex trade off involving a variety of different factors. Consequently it is difficult to define and measure using a single metric. Each metric reveals different factors influencing railroad capacity and must be considered together to get a complete picture of a railroad or individual rail line's ability to handle additional traffic. Railroads try to maximize each of the metrics but there are tradeoffs between them. Furthermore, the importance of each may be weighted differently on different railroads and routes due to variation in traffic, customer requirements and profitability. Railroads will tend to desire the greatest throughput in order to maximize profit but as throughput increases, level of service decreases, which is the metric most important to many shippers. If throughput is too low or if the route becomes congested, asset utilization will decrease, reducing profitability. For North American freight railroads, capacity is integrally linked with profitability. The availability of capacity and operational requirements of a line should be selected to maximize the profitability of a railroad.

CHAPTER 3: CLASS 1 RAILROAD CAPACITY PLANNING PRACTICES

In order to understand how railroads are performing capacity planning and determine topics the industry believes need research I conducted interviews with representatives from the four largest Class 1 railroads.

3.1 Purpose

Railroads perform capacity planning to protect and manage their network. This can be done by assessing the impact of train schedule changes, assisting with maintenance-of-way planning, testing different operating plans, justifying infrastructure expansion expenditures and determining the impact of passenger traffic on both mainlines and in terminals.

3.2 Organization

Capacity planning is a process performed by many people across multiple departments. Modelers are the primary people focused on understanding and modeling operations across the network but there are many other personnel involved in the process. Field transportation personnel and dispatchers identify problems and validate the models, engineering personnel determine the cost and constructability of projects, finance personnel study the economic return of projects and senior management decides which projects should be undertaken.

The location of the capacity modeling group in each railroad's organization structure, the integration and coordination between terminal and line capacity modeling and size of the groups varies between the railroads. As railroad traffic continues to increase railroads are reevaluating these organizations in order to meet these increasing demands.

3.3 Capacity Planning Process

Depending on the type of project the capacity planning process differs, but the general methodology is similar across the industry. The first step for all projects is determining the current capacity utilization. In some cases it is determined that sufficient capacity exists and no further modeling is required. If not, a model must be built and validated. Once validated the model is used to determine the benefit of multiple operational and infrastructure alternatives. In order to determine the impact of changes on future traffic volumes, railroad develop traffic forecasts. Three to five year forecasts of origin destination pairs are converted into trains, locomotives and crews for each corridor. These future traffic levels are added to the route along with any other changes to the traffic including reroutes, different train characteristics and passenger trains. Once the operational benefit and costs of each project are determined the return on investment for the each project is calculated. This information is given to a capacity planning team which gives the final approval to proceed with a project.

3.4 Models and Tools

3.4.1 Rail Traffic Controller (RTC)

The primary tool used by all of the railroads is Rail Traffic Controller (RTC) from Berkeley Simulation Software. RTC is a sophisticated software program designed to realistically simulate both freight and passenger operations over a railroad network *(Washington Group 2007, Parsons 2002).* RTC requires the user to input the infrastructure and traffic into the model. Using the built-in train performance calculator (TPC) and meet-pass logic RTC attempts to dispatch the traffic in a way similar to an actual dispatcher *(Wilson 2010).* The current generation of the software resolves conflicts using priority based dispatching; when there are conflicts, the logic

seeks alternative routes for the lower priority train *(Lai 2008)*. The software outputs multiple reports on train performance, a time-distance graph, and animations. The results from RTC have been validated with hundreds of real-world networks *(Lai 2008)* and since its introduction in 1999 it has become widely accepted by railroads, consultants and government agencies and is the de facto industry standard of the North American railroad industry.

The railroads have modeled most of their network but the data may not be up to date; therefore the initial step may be updating an older file. Once the infrastructure is updated or added, traffic is input into the model. Each railroad has a different philosophy about how to build traffic and run the model. Some railroads input traffic over multiple weeks to match actual variations in traffic. Other railroads will use a shorter period of time and more randomization to account for the variations. The model is then run and validated against actual operations.

3.4.2 Other Line Capacity Tools

While RTC is the primary tool railroads for capacity modeling, railroads also use theoretical and stringline models.

Theoretical models are used to determine bottlenecks on single-track routes. Based on train and traffic characteristics including train length, priority and TPC attributes, signal and siding spacing, siding length and length of slow orders, these models are able to identify the capacity of a segment in trains per day. The models reveal how close traffic volumes are to the capacity of a route.

Stringline tools permit a railroad to quickly determine train run time and available train slots. The use of these models for capacity analysis is limited since most do not have meet/pass

logic built in. However, one model has meet/pass logic for traffic in a single direction and shows how each train uses capacity by showing its "signal wake".

3.4.3 Terminal Capacity Tools

Most of the railroads lack a model for capacity planning of terminals. Terminals are a major contributor to the congestion and delay on the railroad network and currently most planning is based on personal experience and formal analytical tools or models. Without a model railroads are unable to test operational alternatives to increase efficiency and determine the best expansion alternatives. A terminal model is a current area of development in the industry.

3.5 Capacity Metrics

Most railroads use delay as the principal metric to measure capacity. Delay is a measure of level of service and is the primary output from RTC. Acceptable delay is variable based on specific railroad requirements, location and traffic type. Delay limits are based on costs (higher priority trains are a higher value) and geography, a more mountainous route will have higher allowable delays. However, delay is a relative measure and its correlation to capacity is complicated. For instance, current levels of delay may be sufficient now but if a passenger train is added a reduction in delay may be necessary for fluidity of operations. When deciding future investments the relative change to what is currently operating is often considered.

Besides delay railroads will also use average velocity, run times, reliability, throughput, and fuel consumption when making decisions about a project or operational change. Depending on the project each of these may be the limiting factor. Run time will become a constraint when

crews start to approach their hours of service limitations. Recently reduction in fuel

consumption has become a justification to slow trains down.

3.6 Railroad Capacity Research Needs

As a part of the interviews to understand the Class1 Railroads current capacity planning process

each railroad was asked to identify areas they felt had the greatest need for research. Most of the

suggestions were to understand the impact of outside influences on railroad capacity, passenger

rail and positive train control (PTC). Other suggestions included further terminal research and

developing a theoretical model for double track that permits rapid evaluation much like there is

currently is for single track. Finally, railroads are searching for new and improved tools to

facilitate more rapid, accurate and integrated analysis of line and terminal capacity.

CHAPTER 4: IMPACT OF OPERATIONS ON RAILROAD CAPACITY

4.1 Introduction

Railroad operations encompass all aspects of the movement of passengers and goods over the rail infrastructure between many origin and destination pairs. This includes how railroads create, operate and dispatch trains, and is influenced by a complex interaction of traffic, infrastructure, and geographical characteristics. Efficient railroad operation may require intensive planning and analysis. Decisions to improve one aspect may have unintended consequences on other aspects of railroad operations. Operations affect the use of the available rail network capacity and inefficient operations can result in lost capacity. Consequently, in order to more effectively use the available resources, a railroad must understand how their operations influence capacity.

4.2 Operational Factors That Influence Capacity

Railroad operating factors that influence capacity can be separated into two broad groups: train characteristics, and scheduling and dispatching. The key train characteristics include the length, tonnage and power and are determined when a train is made up in a yard. Scheduling and dispatching determine when and how trains are moved from origin to destination. Scheduling is the planning of traffic before leaving the terminal, while dispatching is the control of the trains while in transit. The two groups are closely linked, as train speed is influenced by the power of the train, and train length affects which sidings the dispatcher can direct a train into during meets and passes. In the following sections I will discuss each of these operational factors and their effect on railroad capacity.

4.2.1 Train Characteristics

Train length affects the required number and handling of trains. Longer train lengths reduce the number of trains required to move the same amount of cargo. Fewer, longer trains free up additional train slots for new traffic. However, longer train lengths cause longer train braking distances, resulting in longer blocks and slower speeds. Train length is often limited by yard constraints or the length of sidings on a route. Operating longer trains may also result in longer intervals between trains, with the consequent potential to reduce service quality.

Train weight, much like train length, affects the required number and handling of trains. Heavier railcars reduce the required number of trains, but increase the distance it takes for a train to stop. Depending on the traffic control system, this may require increased train spacing.

Train power affects the ability of a train to accelerate and reach its desired operating speed. A train with a higher power-to-ton ratio or tons-per-equivalent-axle can accelerate more rapidly from a stop and reach a higher maximum speed. Additional power can increase speeds, reduce time lost when a train stops due to a conflict and reduce additional delays to following trains if they must slow to wait for a train to accelerate.

4.2.2 Dispatching and scheduling parameters

Train mix indicates the degree of heterogeneity in the characteristics of trains operating on a route. When there are multiple train types with different characteristics operating on a route, capacity is reduced compared to a route with homogeneous traffic *(UIC 2004)*.

Train priorities are a non-physical characteristic assigned to each train by the dispatcher or timetable. Generally, when two trains meet, the dispatcher will direct the lower priority train to enter the siding, while the higher priority train remains on the main line and proceeds with little or no delay. This prioritization sometimes causes a train to enter a siding earlier than it otherwise would have with no priorities, increasing travel times and reducing capacity *(Krueger 1999, Abril et. al. 2008)*.

Traffic patterns include the headway between trains and how trains are distributed throughout the day. Shorter headways allow more trains but can increase the propagation of delays because the delay of one train is more likely to affect following trains *(Carey & Kwiecinski 1994)*. If trains can be scheduled evenly throughout the day conflicts between trains and the propagation of delays will be less than if the trains are scheduled in dense groups.

Stability is the ability of a network to recover from disruptions. A route may be able to operate with short scheduled headways, but if so, it is less able to recover from unexpected delays. Excess capacity results in more stable operations.

Train speed is the maximum speed at which a train is permitted to operate. Train speeds can affect traffic either positively or negatively. If all the traffic can travel at a uniformly higher speed, the throughput of a route will increase. However if the speed of some of trains increases relative to others, the resultant increase in heterogeneity may potentially reduce capacity. Train speed is influenced by the power-to-ton ratio of the trains, the geography and infrastructure of the route, and rolling stock and/or locomotive specific characteristics

4.3 Literature Review

Hallowell and Harker *(1998)* identify two scheduling strategies that railroads can use: master scheduling and real-time scheduling. Master scheduling is commonly used on European railroads. This involves developing a detailed timetable for scheduled trains and slots for unscheduled trains, and then operating with strict adherence to these schedules. With real-time scheduling, railroads use schedules more as guidelines in making decisions as to how trains should operate. Although North American railroads are becoming more scheduled, most traffic, other than passenger trains, does not conform to a precise schedule. Consequently in order to improve operations in North America research should focus on improved dispatching efficiency.

A key factor affecting railroad dispatching and operations in North America is heterogeneity in train characteristics *(UIC 2004)*. The impact of the higher speeds and priorities of passenger traffic are well documented *(AREA 1921, Mostafa 1951, Harrod, 2009)*, but there has been relatively little research on heterogeneity between freight train types. Most of the research on railroad operations comes from Europe and is focused on the impact of different passenger train speeds and classes (local, express, commuter, etc.). Several papers offer insight

28

into the impact of train type heterogeneity, and propose changes to reduce its impact. Galaverna and Sciutto *(1999)* developed a mathematical model to assess the relationship between capacity and traffic composition using running times. Similarly, Huisman and Boucherie *(2001)* developed a model to estimate running times of trains with heterogeneous traffic. The latter group focused their attention on delays due to differences in speed and showed that as the number of slow trains operating on the line increases, the performance of the faster trains decreases. Vromans et. al. *(2006)* used simulation to study heterogeneous passenger services and developed measures of heterogeneity. By giving local and long distance trains the same number of station stops they were able to homogenize the train schedule and improve operations. Abril et. al. *(2008)* used simulation to investigate different factors influencing capacity on Spanish rail lines. One of the factors they considered was trains operating at two speeds: "normal" and 50% of normal on single- and double-track lines. Their results showed that on single-track lines, capacity is more affected by the average train speed rather than the heterogeneity of train speeds.

European railroad operations are different from those used in North America. European traffic is highly scheduled and composed of higher speed trains that accelerate faster and have shorter braking distances than North American freight trains. Consequently, due to the different operating procedures and train characteristics, the particular factors and the magnitude of their impact on train type heterogeneity will be different. Bronzini and Clarke *(1985)* investigated North American operations using simulation to develop delay-volume curves for traffic with varying amounts of intermodal and unit trains on a hypothetical single-track line. They found that heterogeneity had little effect; delay was more affected by the number of slower unit trains and not heterogeneity. Harrod *(2009)* modeled traffic using mathematical integer programming. He considered the different impacts of faster versus slower non-conforming trains and found that

the slower the non-conforming train, the greater the impact on the network. Krueger *(1999)* used simulation to develop a parametric model that calculated the delay-volume curve for single track routes. The model accounted for heterogeneity by using parameters for average speed, speed ratio and priority.

When considering railroad operations the concept of train delay is one of the primary methods to quantify the efficiency and stability of operations. Mattsson *(2007)* provides a good overview of the relationship between train delay and railroad capacity and examines multiple methods of analysis. Many other papers on delay develop models to predict train delay in order to provide a tool for the operator to estimate the impact of different traffic and schedules. Gorman *(2009)*, on the other hand, used actual traffic data from the BNSF Railway in an attempt to statistically estimate delay for specific routes. His model showed that the most useful measures for predicting congestion delay are meets, passes and overtakes. Gorman's work provides a good tool for understanding the key factors that contribute to train delay in North America.

4.4 Use of Delay to Measure Railroad Operations

The most frequently used metric for analyzing railroad operations is train delay. As discussed in Chapter 2 train delay is the additional travel time resulting from various events en route. Delay is not a direct measure of capacity but rather a measure of the level of service. Delay has direct costs to the railroads due to penalties from shippers but also results in car and locomotive ownership costs, fuel costs, crew costs and lost opportunity costs due to longer cycle times *(Schafer and Barkan 2008, Lai and Barkan 2009)*.

Trains can be delayed for a number of reasons. In general these delay types can be categorized as either scheduled or unscheduled delays. Scheduled delay is the delay incorporated into the timetable as buffer time to allow for conflicts with other traffic. These delays are related to the volume of traffic on a route. With more traffic, the number of meets and passes increases, and headways are reduced, increasing the probability that a faster train will be delayed by a slower preceding train. It is generally agreed that delays increase exponentially with train volume *(Krueger 1999, Gibson et. al. 2002, Mattsson 2007)*, with the specific relationship being dependent on the infrastructure and train mix *(Bronzini and Clarke 1985, Krueger 1999)*.

Unlike scheduled delays, unscheduled delays are random and independent and vary from train to train *(Mattsson 2007)*. For this reason, they are a leading factor in unreliability and instability of a network. Unscheduled delays can be caused by: mechanical failures, malfunctioning infrastructure, weather conditions, excessive boarding times of passengers, accidents at road-railroad crossings and so on *(Carey & Kwiecinski 1999, Vromans et. al. 2006)*. Large unscheduled delays can cause crew shortages and disrupt operations on the entire network. While unscheduled delays have a low probability they can cause substantial amounts of delay. Consequently, adjustments to a train's schedule to account for potential unscheduled delays are difficult.

Unscheduled delays to one train, sometimes referred to as primary or exogenous delay, can lead to secondary or "knock-on" delays to other trains. Secondary delays are a result of the shared usage of infrastructure, rolling stock and crews *(Carey & Kwiecinski 1999, Vromans et. al. 2006)*. Delay of one train may cause a following train to stop or reduce its speed and may potentially affect subsequent trains as well. The amount of secondary delay depends not only on

31

the frequency and duration of primary delays, but also on the amount of available capacity. As a route nears its theoretical capacity the probability that primary delays will lead to secondary delays increases, while the ability to recover from these delays decreases *(Mattsson 2007, Congressional Research Service 2007, Weatherford et. al. 2008).*

The relationship between delay and capacity is not simple. The maximum capacity for a route in terms of volume is dependent on operational decisions by the railroad. Each railroad determines the maximum allowable delay based on the traffic mix, route geography and service requirements. Different types of traffic have different requirements; for instance, due to competition from trucks intermodal traffic requires higher velocities and fewer delays, than shipments of coal that are less time sensitive. If a railroad requires a more reliable schedule with fewer delays, the capacity utilization on a route will be less than if a railroad is willing to accept higher delays (Figure 4.1a). Thus surplus track capacity serves to minimize delay *(AREA 1931).* Additionally, different traffic patterns will result in different delay-volume relationships; therefore the maximum volume will be different at the same amount of acceptable delay (Figure 4.1b).
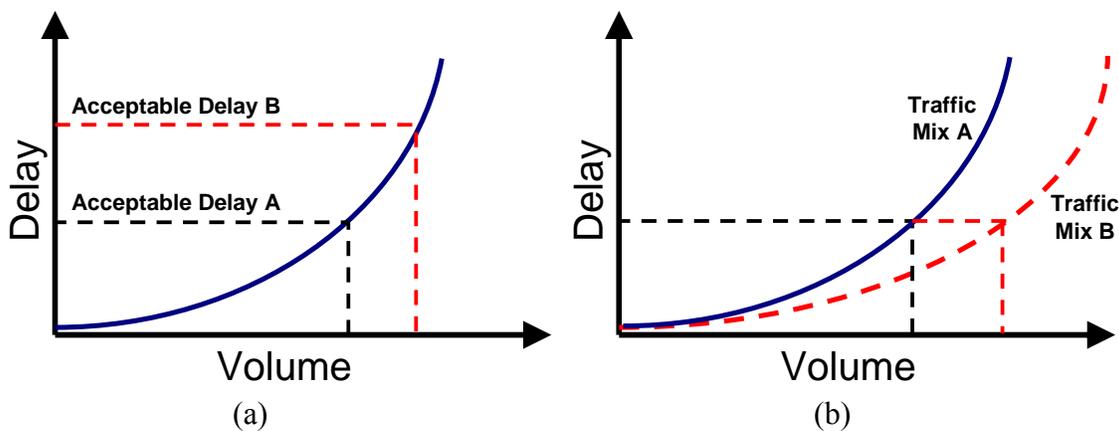
FIGURE 4.1: Maximum Volume Based on (a) Maximum Allowable Delay and (b) Different Traffic Mixes

There are two generally accepted definitions of delay. One definition for delay is the difference between the minimum, or unopposed, run time and the actual run time required to traverse the route. This includes both scheduled and unscheduled delays. Delays using this definition are directly related to the run time or average speed of the traffic. The other definition of delay only calculates the unscheduled delays. This is calculated as the difference between the scheduled and actual run time. Delays using this definition are directly related to the reliability or on-time performance of the traffic.

Delay is often used as the output in simulations and can provide a basis for decisions on infrastructure projects or operational changes *(White 2006)*. Each railroad has calculated the costs of train delay to their operations. Using this value they are able to calculate the economic value of a project or operational change.

Although delay is a useful and widely used metric, it is not without its limitations. Delay, volume and capacity are not always related *(White 2006)*. Delay can change independently of capacity depending on the train speeds and the arrangement of the sidings. Additionally, it is possible for increased traffic to cause increased delays without a negative effect on existing traffic. For example, assume that an increase in train speeds reduces travel time by twenty minutes but new traffic is added that increases delay by ten minutes. The result is an increase in volume and a ten minute reduction in total travel time, but, this improvement is not reflected in the delay measurement *(White 2006)*.

CHAPTER 5: IMPACT OF HETEROGENEITY ON DELAY ON A SINGLE-TRACK LINE

5.1 Introduction

Efficient use of existing railroad infrastructure and effective planning for new infrastructure or other capacity enhancing systems requires an understanding of how operations affect capacity. A key factor in railroad operations is the interaction of trains with different operating characteristics. In North America, intermodal, manifest, unit and local trains may all share trackage. Some lines also have intercity passenger trains and in metropolitan regions, commuter trains. Each of these train types can have considerably different characteristics and this heterogeneity can have a substantial effect on rail line capacity *(Pachl 2002, UIC 2004, Abril et. al. 2007)*.

With homogenous traffic, delays on a single track line are mostly due to meets. With heterogeneous traffic, delay is also caused by conflicts that occur as a result of differences in train characteristics, some of which increase frequency and duration of meets and passes. Additional situations causing delays with heterogeneous traffic include:

- Train delayed by a slower preceding train

- Train delayed by a preceding train with slower acceleration

- Trains experience longer meets waiting for higher priority trains

- Train delayed waiting for another train to pass

- Trains experience more conflicts due to lower average speeds resulting from other delays due to heterogeneous traffic

The magnitude of these delays is dependent on the specific train mix, volume and amount of heterogeneity.

Bronzini and Clarke *(1985)* used simulation to develop delay-volume curves for traffic

with varying amounts of intermodal and unit trains on a theoretical single-track line. To further

understand and develop their work, simulation software was used to conduct a quantitative

analysis of the impact of heterogeneity among these two train types, which are among the most

frequently operated trains on the North American railroad network. I evaluated the effects of

different volumes and percentages of each train type on a signalized, single-track route. Delay

was used as the principal metric to assess capacity impacts under different scenarios. The

objective of this research is to provide insight into the impact of train type heterogeneity and

understand the key characteristics of traffic heterogeneity that have the greatest impact on delay.

5.2 Methodology

5.2.1 Capacity Metric

Delay was used as the principal metric for capacity comparisons in this study. I define delay as

the difference between the minimum, or unopposed, run time and the actual run time required to

traverse the route. This includes the time spent stopped for meets and passes, along with the

time for braking and accelerating. The total delay was divided by train miles for a normalized

value of delay per 100 train miles.

5.2.2 Dispatch Simulation Software

I used Rail Traffic Controller (RTC) from Berkeley Simulation Software for my analyses. RTC

is a sophisticated software program designed to realistically simulate both freight and passenger

operations over a railroad network *(Parsons 2002, Washington Group 2007).* The software uses

infrastructure and traffic inputs specified by the user to resolve multi-train conflicts in a manner

intended to mimic decisions of a railroad dispatcher. I used RTC because its flexibility permits rapid evaluation of different operating scenarios and because of its widespread acceptance and use by the North American railroad industry.

5.2.3 Representative Rail Line

Specific characteristics of individual rail lines are unique and route characteristics influence railroad operations. For my research I developed a hypothetical rail line intended to represent the characteristics of a typical midwestern North America, single-track mainline subdivision (Table 5.1).

TABLE 5.1: Route Used in Analysis

Single track with 10 mile siding spacing
262 miles long
10 miles between siding centers
8,700 ft signaled sidings with #24 powered turnouts
2.75 mile signal spacing
2-block, 3-aspect signaling
0% grade and curvature

Although the attributes are somewhat idealized, the purpose is to provide a consistent basis for relative comparison of different scenarios of interest in this research under a reasonably realistic set of operating conditions. However, there is no intent to imply that the results presented here represent absolute predictive measurements for a particular set of conditions.

5.2.4 Train Types Used in Simulations

Combinations of intermodal and bulk trains were used to investigate the impact of heterogeneity. Bulk goods transported by rail include coal, ore, grain and stone typically in long, heavy, unit trains with low horsepower-to-trailing ton (HPTT) ratios (e.g. <1.0). Intermodal trains transport

trailers and containers carrying consumer goods and other high priority shipments to and from domestic and international markets. Since these shipments require fast, reliable service, intermodal trains travel at higher speeds and have higher horsepower to trailing ton ratios (e.g. >3.0). Using commodity based railroad transportation statistics it can be estimated that bulk and intermodal traffic account for roughly 60% of the US Class 1 railroad's revenue, 75% of the tonnage and 80% of the carloads *(AAR 2008a)*.

While each individual train is different, the attributes for each train type were selected to match their average characteristics (Table 5.2). The numbers of cars and units were obtained from the Cambridge Systematics National Rail Freight Infrastructure Capacity and Investment Study *(2007)* conducted for the Association of American Railroads (AAR). Typically well cars with 5 articulated units are used to transport international containers and spine cars with 3 articulated units are used to transport domestic trailers. The number of cars was determined based on the relative amounts of domestic and international intermodal traffic. Tonnages and lengths were based on averages for each car type. The power-to-ton ratios were based on experience and information from the TRB Workshop on Railroad Capacity and Corridor Planning *(TRB 2002)*.

Train priority is a non-physical characteristic assigned to it by the dispatcher. When two trains meet, priority is one factor the dispatcher will take into consideration when determining how to resolve the conflict. Dispatchers will generally try to minimize the total cost of delay *(Washington Group 2007)*. This means that the trains carrying lower value or less time-sensitive freight, will have lower priority and enter the siding, while the higher priority train holds the main and proceeds with little or no delay. In this study, intermodal trains were assigned the higher priority, as is typical in most railroad operations.

While the chosen attributes approximate actual characteristics, at the most basic level the "intermodal" trains represent freight trains with the highest maximum speeds, power-to-ton ratios and dispatching priorities, while the "bulk" trains represent those with the lowest speeds, power to ton ratios and dispatching priorities. Although I refer to them using the terms "intermodal" and "bulk" for convenience, what is actually of significance in the analyses are their specific operating characteristics, not the particular type of consist.

TABLE 5.2: Train Composition Characteristics Used in Simulations

Intermodal	Bulk
16 Three-pack spine cars 9 Five-pack well cars	115 loaded hopper cars
5,659 ft	6,325 ft
5,900 tons	16,445 tons
3.64 HPTT	0.78 HPTT
Five 4,300 HP Locomotives	Three 4,300 HP Locomotives
Maximum Speed: 70 mph	Maximum Speed: 50 mph

5.2.4.1 Train Braking and Acceleration Characteristics

When considering the impact of heterogeneity it is important to consider how the operating characteristics of the train affect its operation. Tonnage, power and length of a train have a direct impact on its ability to accelerate and brake, and consequently its effect on capacity. In typical operation a train will slow to a stop gradually using a mixture of air and dynamic brakes. The braking distances in RTC were calibrated with the assistance from a class 1 railroad to approximately match the braking distances of each type of train making a typical brake application.

To better understand the differences in operating characteristics of the two trains in the analysis their braking and acceleration distances were compared (Figure 5.1). Intermodal trains

brake at a faster rate, but due to their higher initial speeds, their braking distance is only about 2,000 feet less than the bulk train. On the other hand, there are large differences in the acceleration distances for the two train types. Intermodal trains have a much higher horsepower-to-trailing-ton ratio and therefore are able to accelerate to maximum speed more quickly.

5.2.5 Simulations

A series of simulations were developed with different traffic volumes and levels of heterogeneity. Eleven different volumes were tested from 8 to 48 trains per day simulated in increments of four. Each volume is based on an equal temporal distribution of trains in each direction over a 24 hour period. This was not intended to represent practical, sustained operation, which includes windows for inspection and maintenance, but rather to provide a basis for relative comparison of the effect of various factors of interest. The results are therefore more characteristic of the spacing or headway between trains than the actual volume. Seven different levels of heterogeneity were tested based on the percentage of the different train type. The tests were done with 12.5%, 25%, 50% and 100% of each train type. The traffic was composed of a single train of the lower percentage type, followed by the corresponding number of trains for the other train type. For example 12.5% bulk trains means that every bulk train was followed by seven intermodal trains. The ratios and traffic pattern were the same for trains traveling in both directions. For each configuration a series of twenty-five simulations were performed with the departure time of each train randomized with a uniform distribution over a 30-minute interval, up to 15 minutes before or after the scheduled departure time for that train.
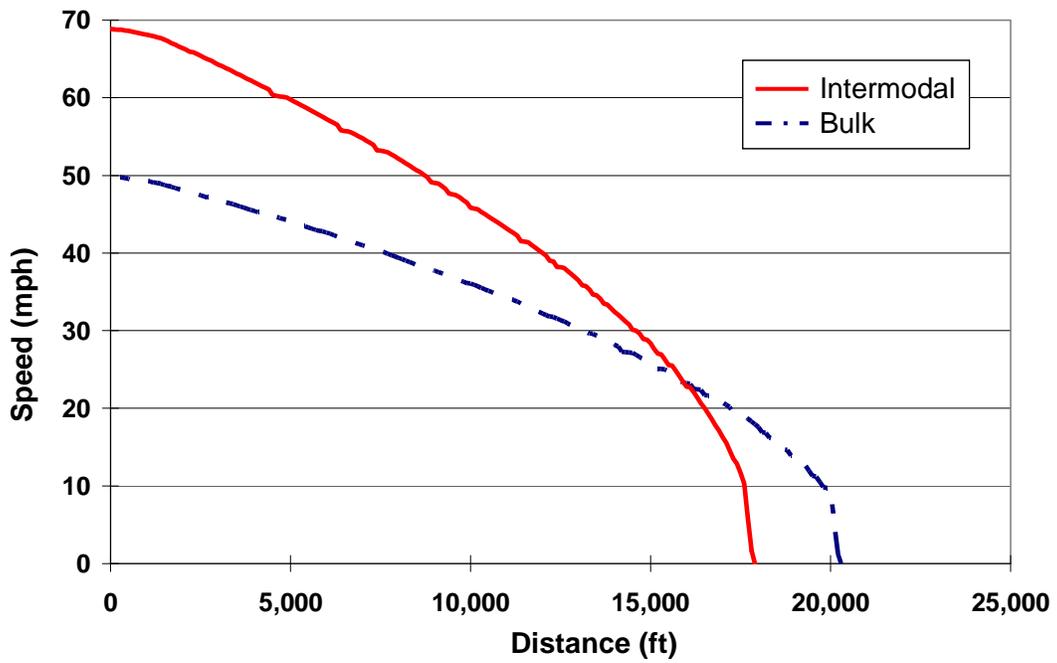
(a)
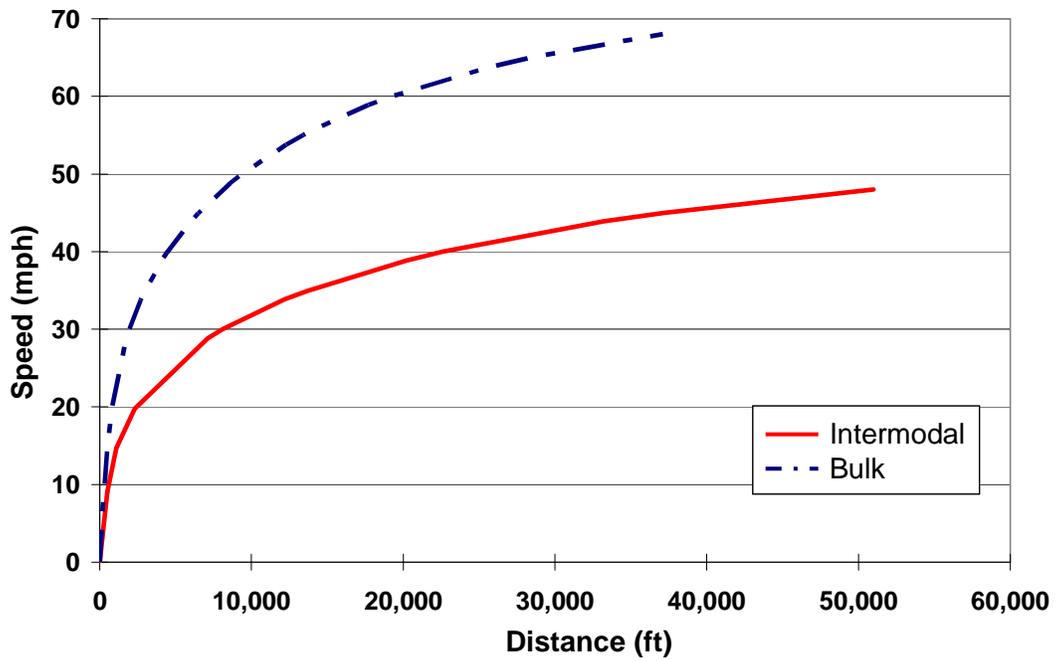
(b)

FIGURE 5.1: Bulk and Intermodal Train (a) Braking and (b) Acceleration Curves

5.3 Delay-Volume-Heterogeneity Relationship

To better understand the relationship between delay, volume and heterogeneity, simulation results were combined to create a surface over the three axes (Figure 5.2). The delay-volume-heterogeneity surface reveals two trends: delay increases with increasing volumes and is dependent on the traffic mix. To further investigate these trends, each axis was analyzed individually (Figure 5.3). This permits the investigation of the delay-volume relationship, delay-heterogeneity relationship, and volume-heterogeneity relationship.
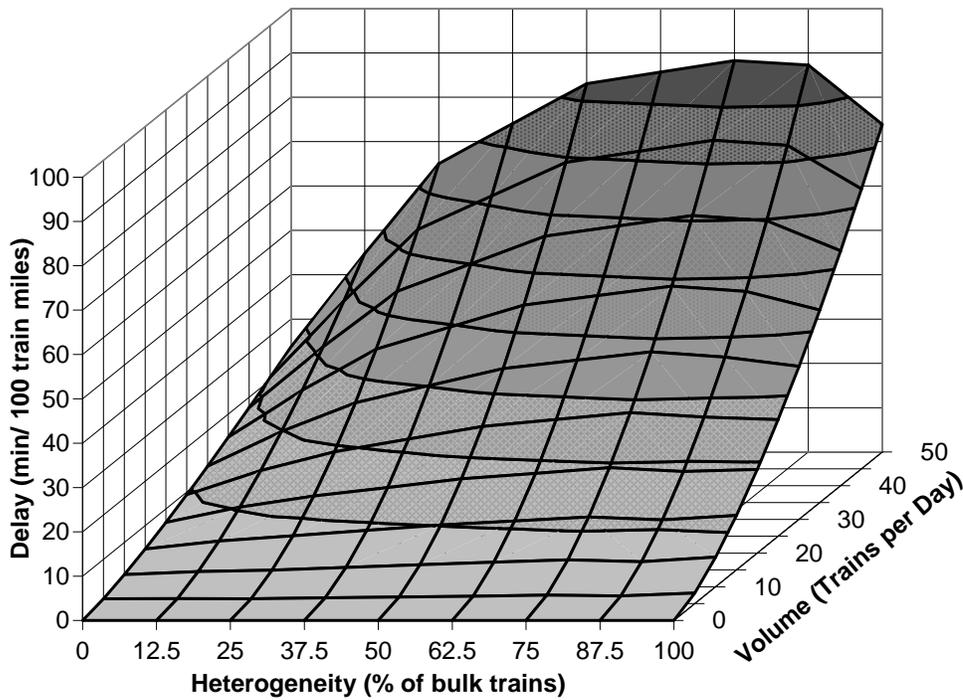
FIGURE 5.2: Delay-Volume-Heterogeneity Relationship for Intermodal and Bulk Trains
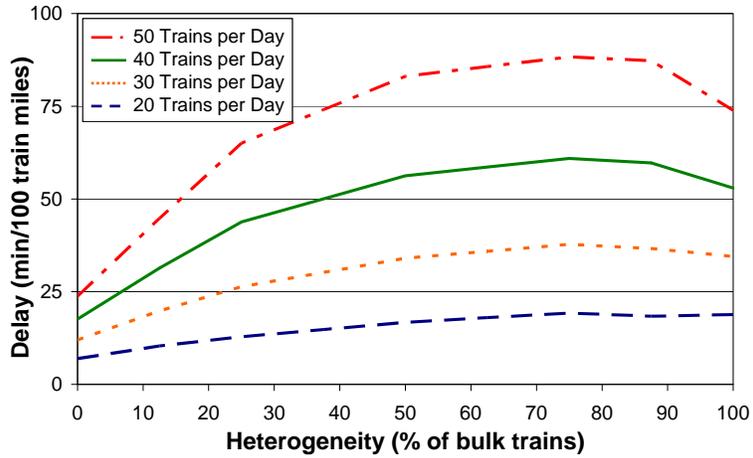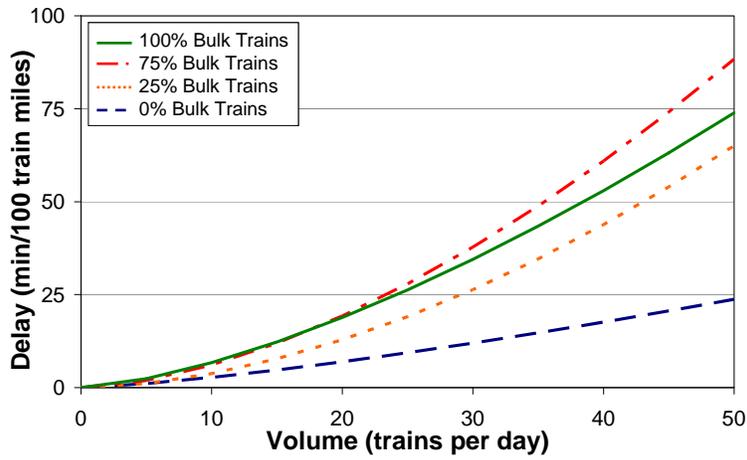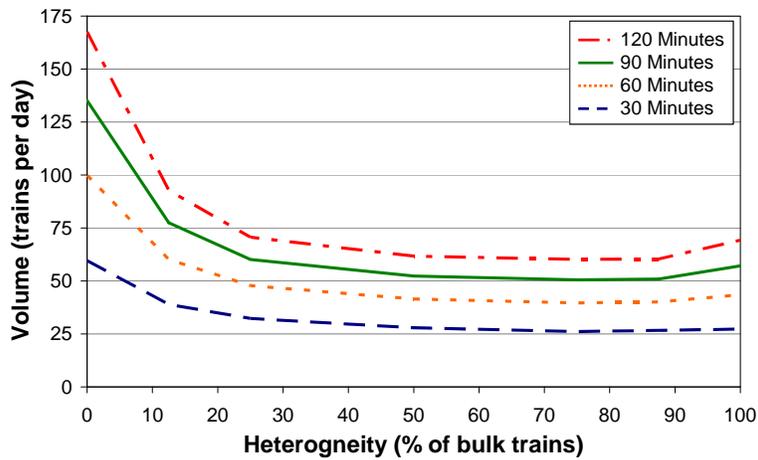
FIGURE 5.3: The (a) Delay-Heterogeneity, (b) Delay-Volume and (c) Volume-Heterogeneity Relationships for Intermodal and Bulk Trains

42

5.3.1 Delay-Heterogeneity Relationship

The delay-heterogeneity graph shows the effect of heterogeneity on delay at various volumes (Figure 5.3a). Delay is greatest when the traffic is heterogeneous. However, the maximum average delay is not when heterogeneity is largest (50%) but when the traffic is 75% bulk trains and 25% intermodal. The traffic composition resulting in the greatest delay depends on the train characteristics. In this example bulk trains experience greater delay due to their slower maximum operating speed. The combined effect of the larger number of these slower performing trains and heterogeneity is greatest when the traffic is 75% bulk trains.

5.3.2 Delay-Volume Relationship

The delay-volume graph shows the effect of volume on delay at various levels of heterogeneity (Figure 5.3b). It is evident that the effect of additional trains on delay is non linear. Delay increases at an increasing rate with higher volumes. At higher volumes there are additional meets and shorter headways resulting in greater delays. The delay-volume relationship is different for various traffic compositions. Higher levels of heterogeneity cause greater delays and the delays increase more rapidly than with homogenous traffic.

5.3.3 Volume-Heterogeneity Relationship

Another aspect of the delay-volume-heterogeneity surface that should be considered is the relationship between volume, heterogeneity and various levels of delay (Figure 5.3c). Delay is often used by the railroads to determine capacity. Capacity planners determine the acceptable amount of delay to meet their level-of-service requirements. In this example, if the maximum allowable delay is 60 minutes per 100 train miles and the traffic consists of only intermodal

trains, the maximum capacity is 100 trains per day. If the traffic changes to an even mix of intermodal and bulk trains, capacity drops to 42 trains, 21 intermodal and 21 bulk trains. Therefore, by adding 21 bulk trains the ability to run 79 intermodal trains is lost. This effect is greatest when a few bulk trains are added to a route that is primarily intermodal traffic.

5.3.4 Delays due to Heterogeneity for Each Train Type

Train type heterogeneity increases average delay but the delays are different for each train type. At 48 trains per day the delays to each train type were considered individually (Figure 5.4). In every traffic mix the bulk trains have higher delays than the intermodal traffic. The intermodal train delay is relatively constant over all traffic mixes with the greatest delays at the highest levels of heterogeneity. Bulk trains have much higher delays but the delays decline as the percentage of bulk trains increases. When the percentage of intermodal trains is lower it becomes less likely for a bulk train to meet a higher priority intermodal train. When this occurs, it often results in a meet that causes greater delays than would have occurred if a train of equal priority had been encountered. Much of the increased delay due to heterogeneity is the result of additional delays to the bulk trains, not an increase in the delays to both train types. The difference in delay between the intermodal and bulk train traffic decreases at lower volumes and as a result, delays due to heterogeneity are also lower.
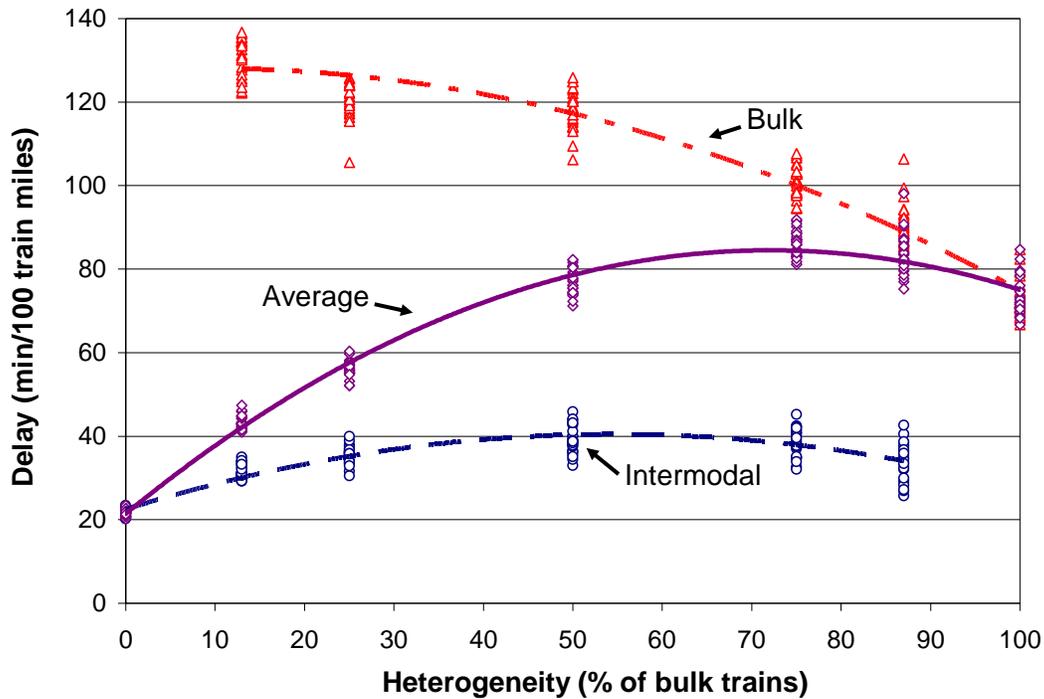
FIGURE 5.4: Delay Separated by Train Type at 48 Trains per Day

5.4 Delay-Volume-Heterogeneity Relationship for Delays due to Heterogeneity

When traffic is homogenous the average delays for intermodal and bulk trains are different.

Therefore, as the traffic mix changes, some of the delay is due to more trains experiencing a

differing delay due to their characteristics, rather than the result of heterogeneity alone. In order

to account for this the hypothetical delay that would occur for the same traffic mix in the absence

of any heterogeneity-caused sources was subtracted from the delay for each particular mixed-

traffic scenario (Figure 5.5).
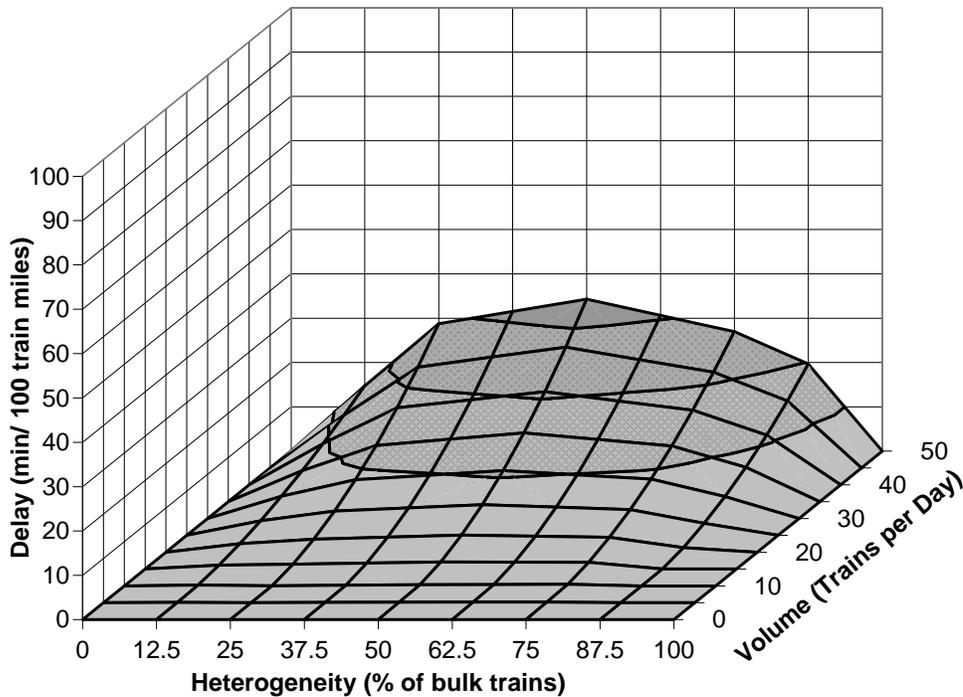
FIGURE 5.5: Delay-Volume-Heterogeneity Relationship for Intermodal and Bulk Trains with Delay due to Heterogeneity Isolated

It is clear that heterogeneity causes only some of the total delay a train experiences. The percentage of the total delay caused by heterogeneity increases with greater traffic volume and levels of heterogeneity. However, even at 48 trains per day, when the traffic is an even mix of intermodal and bulk trains, heterogeneity accounts for only 40% of the total delay. With smaller volumes and less heterogeneity this percentage is even lower.

As with total delay, the delay-heterogeneity, delay-volume relationships and volume-heterogeneity relationships are considered (Figure 5.6). The trends are different than total delay and provide further useful insights into the impact of heterogeneity. Since the delays are only due to heterogeneity the volume-heterogeneity relationship does not offer any additional insight into the effects of heterogeneity.
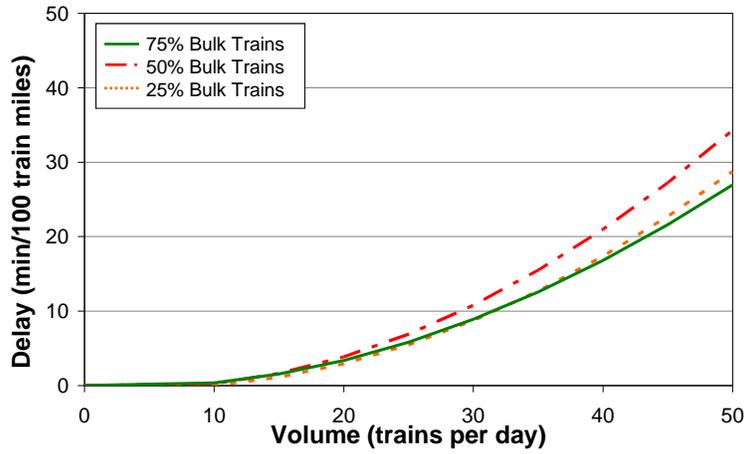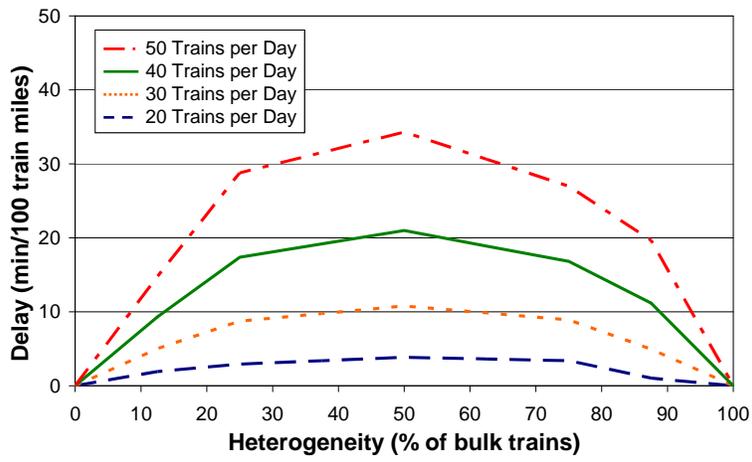
FIGURE 5.6: The (a) Delay-Heterogeneity and (b) Delay-Volume Relationships for Intermodal
and Bulk Trains with Delays due to Heterogeneity Isolated

5.4.1 Delay-Heterogeneity Relationship

Unlike total delay, the delay due to heterogeneity is when heterogeneity is greatest (Figure 5.6a). While the maximum total delay occurs at 75% bulk (25% intermodal), the maximum delay due to heterogeneity is when there is an even mix of intermodal and bulk trains. When the volumes are low there is almost no impact of heterogeneity on delay, with the delays due to heterogeneity increasing with greater volumes.

5.4.2 Delay-Volume Relationship

The delays due to heterogeneity have a non-linear relationship (Figure 5.6b). There is little effect of heterogeneity until about 15 trains per day. At low traffic levels the headways between trains are great enough that passes are limited and excess siding capacity allows for efficient meets and passes. However as traffic volume increases beyond this, meets become more frequent and headways shorter, thereby magnifying the differing characteristics and creating additional conflicts.

5.5 Analysis of Factors that Cause Delays Due to Heterogeneity

Although the effect of volume and heterogeneity on delay is clear, the specific factors causing the increased delays due to heterogeneity are not. Additional simulations were conducted to investigate the sensitivity of delay to speed, braking performance, acceleration performance and dispatching priority. While not possible in actual operations use of simulation enables each factor to be isolated to determine its effect on delay and capacity. Two scenarios were studied for each factor: in the first scenario the selected factor was the only source of heterogeneity and the delays due to this factor were considered. The second scenario is the opposite condition; the

factor was eliminated from the base scenario and the reduction in delay due to this action was evaluated.

5.5.1 Impact of Heterogeneity in Speed

Trains travel at different speeds depending on their service requirements. Higher valued commodities typically travel in higher speed trains than less time-sensitive commodities. The resultant heterogeneity in train speed can create additional conflicts and delays. With heterogeneous train speeds, a faster train may overtake a slower one. Before there is an opportunity to pass, the faster train may have to slow to the speed of the preceding train. Once a siding is reached the train being overtaken will have to reduce speed to enter the siding and stop thereby allowing the faster train to pass. In this manner heterogeneous train speeds can cause extra delay for both the faster and slower trains.

In the first scenario I considered intermodal trains traveling at 50 and 70 mph at four different traffic levels: 12, 24, 36, and 48 trains per day (Figure 5.7a). At the two lower volumes the delay due to the heterogeneity in speed was negligible. However, at 36 and 48 trains per day the speed difference caused delays. In general, delay due to speed heterogeneity is greatest at the highest volumes and levels of heterogeneity. In typical operations the higher speed trains receive a higher dispatching authority, otherwise when a higher speed train overtakes a slower train it will not result in a pass, thus increasing delays. However, in this scenario the priorities of the two trains train types were equal and therefore some additional delays are the result of faster trains being slowed.
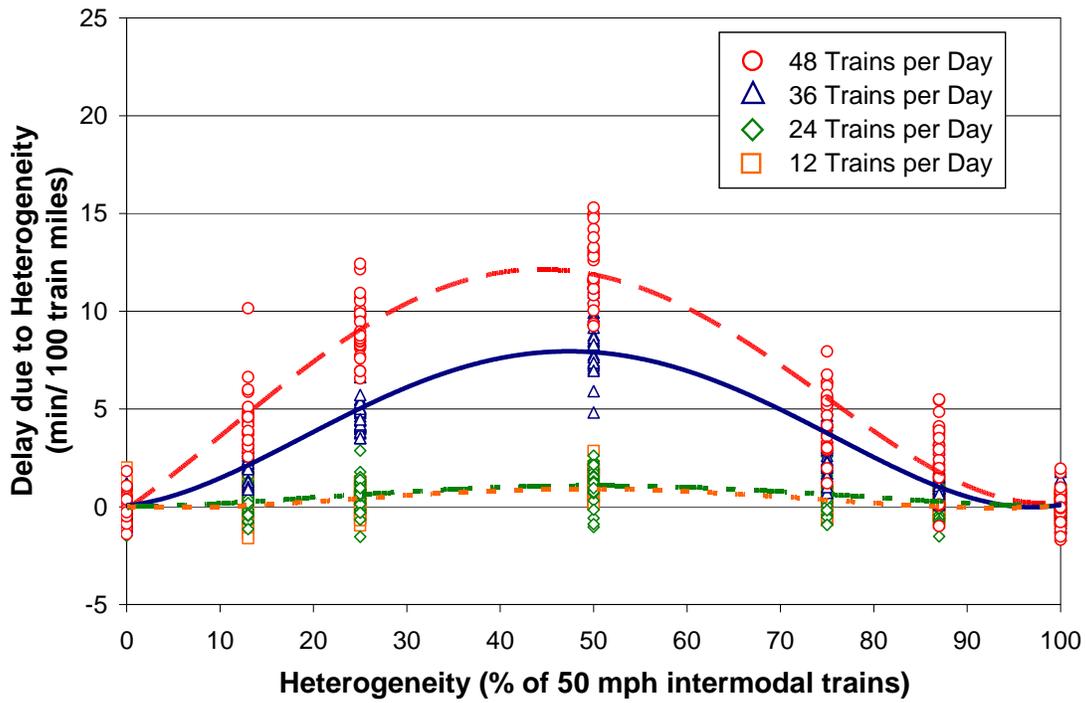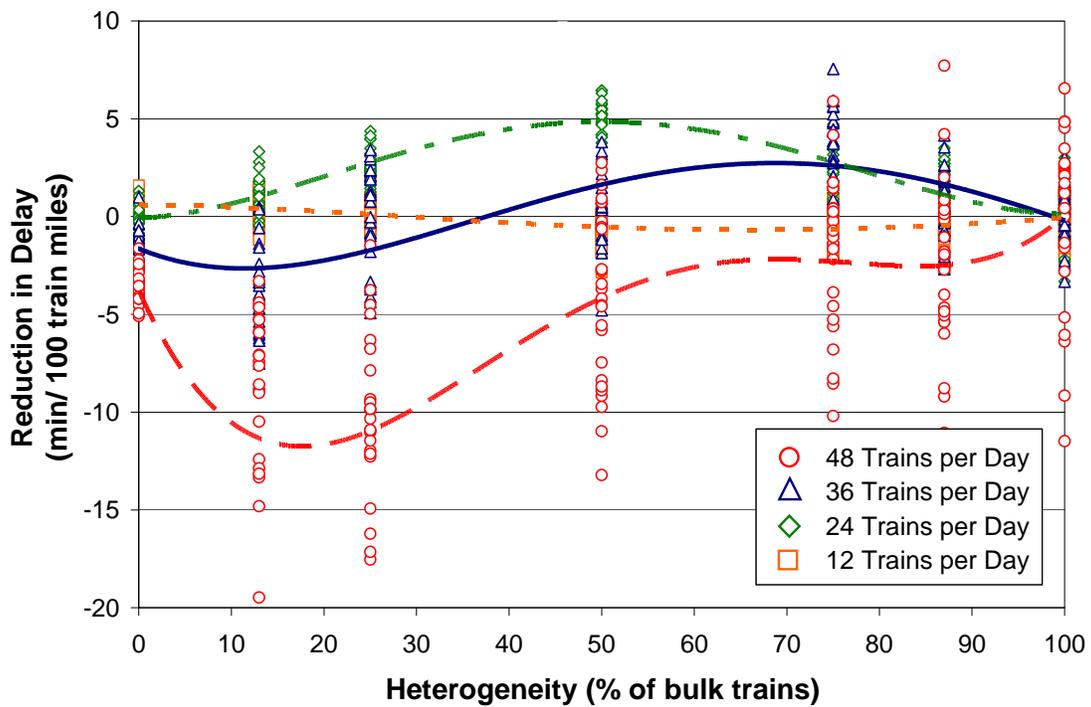
FIGURE 5.7: Delays due to Train Speed Heterogeneity (a) and Reduction in Delay from Base Scenario when Speeds of Intermodal Trains are Reduced to 50 mph (b)

50

In the second scenario the only change from the base scenario was a reduction in speed of the intermodal trains to 50 mph in order to eliminate the speed difference between trains (Figure 5.7b). The impact of this change varied, depending on the volume of traffic. At 12 trains per day the change in delay was minimal. When the volume increased to 24 trains per day delays were reduced at all levels of heterogeneity. At 36 trains per day the delays were reduced when the majority of the traffic was bulk, and increased when the traffic was a mostly intermodal. At the highest volume, 48 trains per day, delays increased with all levels of heterogeneity.

At the lowest volumes there may not be any passes due to the large headways between trains; consequently, reducing speed to eliminate meets has no effect. When volumes increase delay is reduced due to the elimination of speed difference conflicts. However, slower train speeds increase the time a train occupies the track, resulting in more meets and the associated delay. At the highest volumes the delay from the increased number of meets outweighs the savings from the elimination of passes. Additionally, the reduction in speed from 70 to 50 mph for the intermodal trains increases run times by 35.34 minutes per 100 train miles. Therefore while delay is reduced at 24 trains per day, the run time actually increases due to the slower speed.

5.5.2 Impact of Heterogeneity in Braking Performance

Trains with different lengths and tonnages require different distances to stop. Braking distance affects both signal spacing and delay during a meet. A train with poorer braking performance will take longer to reduce speed and thus require more time to resolve a conflict. Longer meet delay will increase run times and can result in additional conflicts due to longer track occupancies.

The first scenario considers bulk trains with normal braking performance, and with the braking performance of intermodal trains (Figure 5.8a). Braking performance affects the delays experienced by a train, however conflicts between trains with different braking performances did not result in additional delays. The delay that occurs while a train is braking is relatively minor compared to the time lost while a train is waiting in a siding or the time it takes the train to accelerate back to top speed. Since trains were not delayed by preceding trains with poorer braking performance; heterogeneity in braking performance has little impact.

In the second scenario heterogeneity in braking performance was eliminated by improving the braking performance of the bulk trains to match that of the intermodal train (Figure 5.8b). The changes resulted in a reduction in delay corresponding to the percentage of bulk trains. As the percentage of bulk trains increased, so did the reduction in delay. The reduction also increased with higher volumes since there are a greater number of stops for meets and passes. However, the changes did little to reduce heterogeneity caused delays.
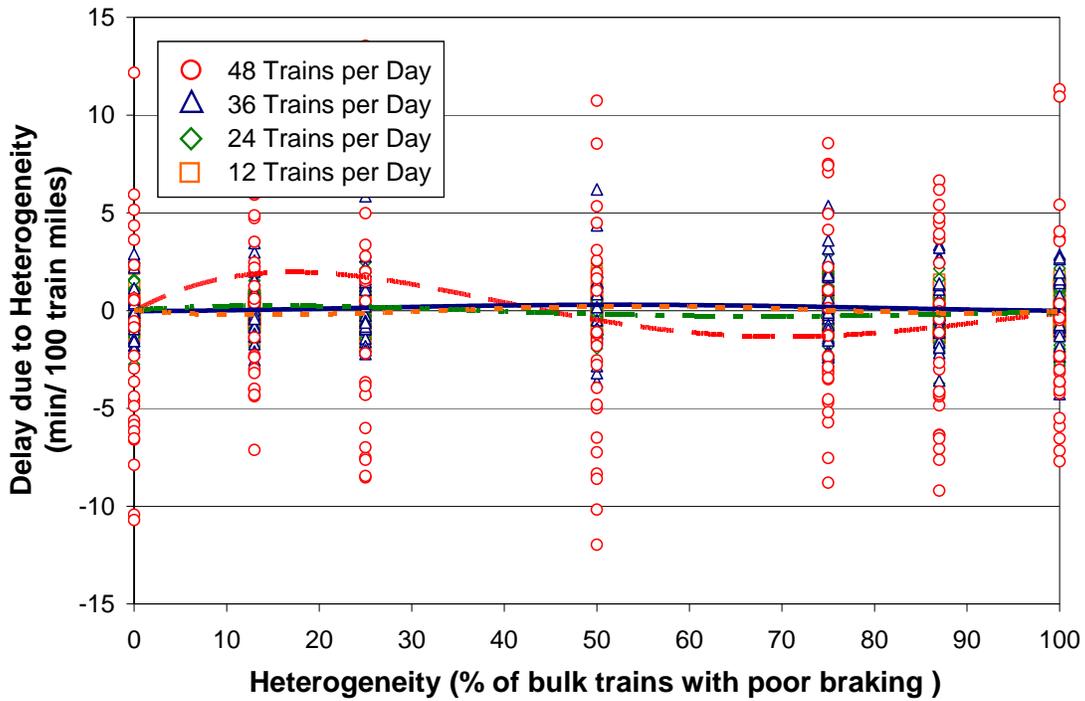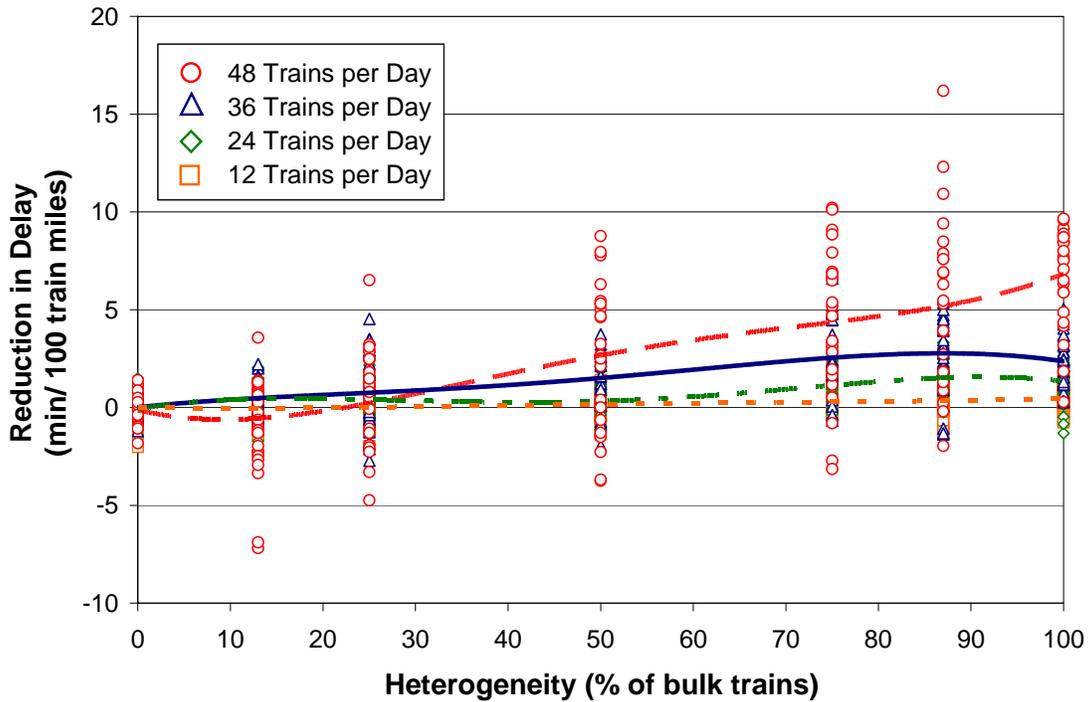
FIGURE 5.8: Delays due to (a) Braking Performance Heterogeneity and (b) Reduction in Delay from Base Scenario when Braking Performance of Bulk trains is Improved to Match Intermodal

In this research, only the impact of braking performance on delay with fixed signal blocks was considered. In actuality, braking performance affects signal spacing and allowable train speeds. Another implication of heterogeneity in train braking is the inability to optimize the signal system for a specific train type. Signal spacing is typically set based on the longest, full service stopping distance of any train that regularly operates on a line. If traffic is homogeneous signal spacing can be designed for the specific train type thereby enabling shorter headways. Alternatively, without a change in signal spacing improved braking performance can allow for faster train speeds since it is possible for a faster train to stop in the same distance as an unimproved train at a slower speed.

5.5.3 Impact of Heterogeneity in Acceleration Performance

Another type of train heterogeneity is acceleration performance. Acceleration is directly related to the HPTT ratio of a train. A higher power-to-ton ratio allows a train to accelerate faster and reach higher top speeds. There are two ways acceleration performance can influence capacity. First, heterogeneity in train acceleration performance can potentially impact capacity if a following train is delayed while a train ahead of it accelerates to maximum track speed. As traffic volume increases, the number of meets and passes increases while headways are reduced, increasing the potential for this type of conflict. Secondly, the acceleration performance of a train affects its delay because it increases the time it takes for a train to reach maximum track speed after a conflict.

The first scenario considers traffic composed of bulk trains with normal acceleration performance and the acceleration performance matching intermodal trains (Figure 5.9a). The delay due to heterogeneity in acceleration performance was dependent on traffic mix. Delays

were minimal when most trains had superior acceleration performance, but increased with the percentage of trains having inferior acceleration performance. When the majority of trains have poor acceleration performance, it is more likely for a trailing train to be slowed by a train accelerating from a meet, thereby causing additional delay.

The second scenario removes heterogeneity in acceleration performance from the base scenario by improving the acceleration performance of bulk trains to match that of intermodal trains (Figure 5.9b). This change revealed that acceleration performance is a significant factor causing delays to bulk trains and increased delays due to heterogeneity. The reduction in delay increased with greater traffic volumes and with an increasing percentage of bulk trains, however the greatest reduction is not with homogenous bulk train traffic but when 75% of the trains are bulk. Since the reduction in delay is when the traffic is heterogeneous, improved acceleration performance not only reduces delay to each bulk train but it reduces train conflicts that result in delay.

In typical operations lower priority trains often have poorer operating characteristics. This can increase congestion on a route because these trains will make more frequent stops to resolve conflicts, increasing the possibility of following trains being delayed. Without priorities, the trains with poorer operating characteristics would make less frequent stops reducing the possibility of these types of conflicts. Therefore, acceleration performance alone does not significantly influence delays but in combination with priorities it can cause the additional delays that result from heterogeneous train traffic.
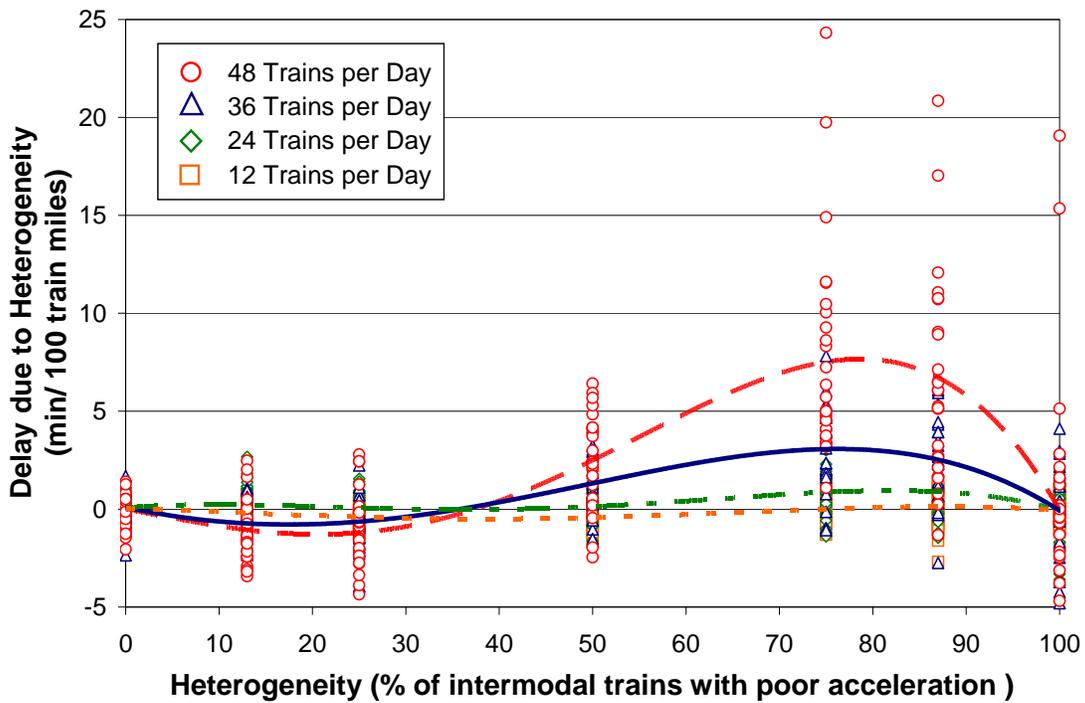
(a)
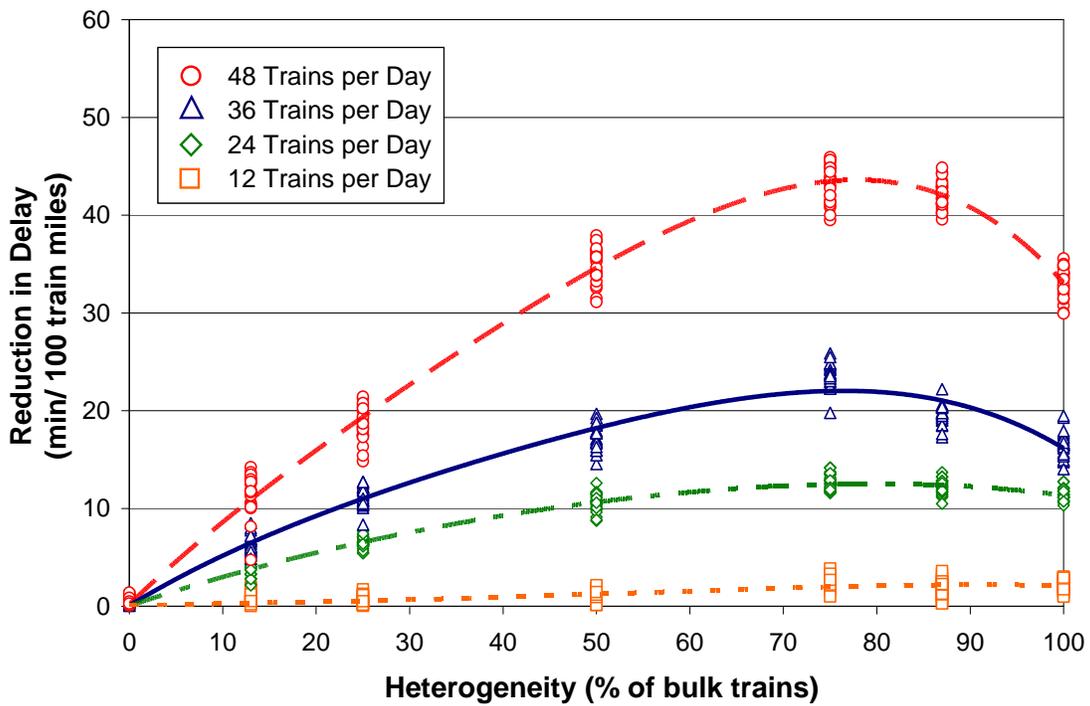
(b)

FIGURE 5.9: Delays due to (a) Acceleration Performance Heterogeneity and (b) Reduction in Delay from Base Scenario when Acceleration Performance of Bulk trains is Improved to Match Intermodal

56

5.5.4 Impact of Heterogeneity in Priority

The last characteristic considered was the dispatching priority assigned to trains. Intermodal trains with their higher value merchandise and greater customer demand for fast, reliable service are typically given the highest priority by railroad dispatchers. Bulk trains are typically less time dependent and therefore given lower priority on the network *(Congressional Research Service 2007)*. Without differential priorities during a conflict, the first train to arrive at a siding will enter it and wait for the other train. With priorities, the lower priority train will often have to stop at an earlier siding and wait so as to prevent delays to the higher priority train. Consequently priority can increase both the number and duration of meets.

The first scenario considers traffic composed of intermodal trains with high and low priorities (Figure 5.10a). When priority is considered alone the increased delays are due to more frequent and longer delays of lower priority trains during meets. As would be expected, these delays are greatest at the highest volumes and levels of heterogeneity. The second scenario considers the base scenario with all the traffic having the same priority (Figure 5.10b). With homogenous traffic all the trains are the same and have equal priority; consequently removing heterogeneity in priority did not change the delays in these scenarios. However, with heterogeneous traffic, removing priorities reduced delay with the greatest reduction at the highest levels of heterogeneity and traffic volumes.
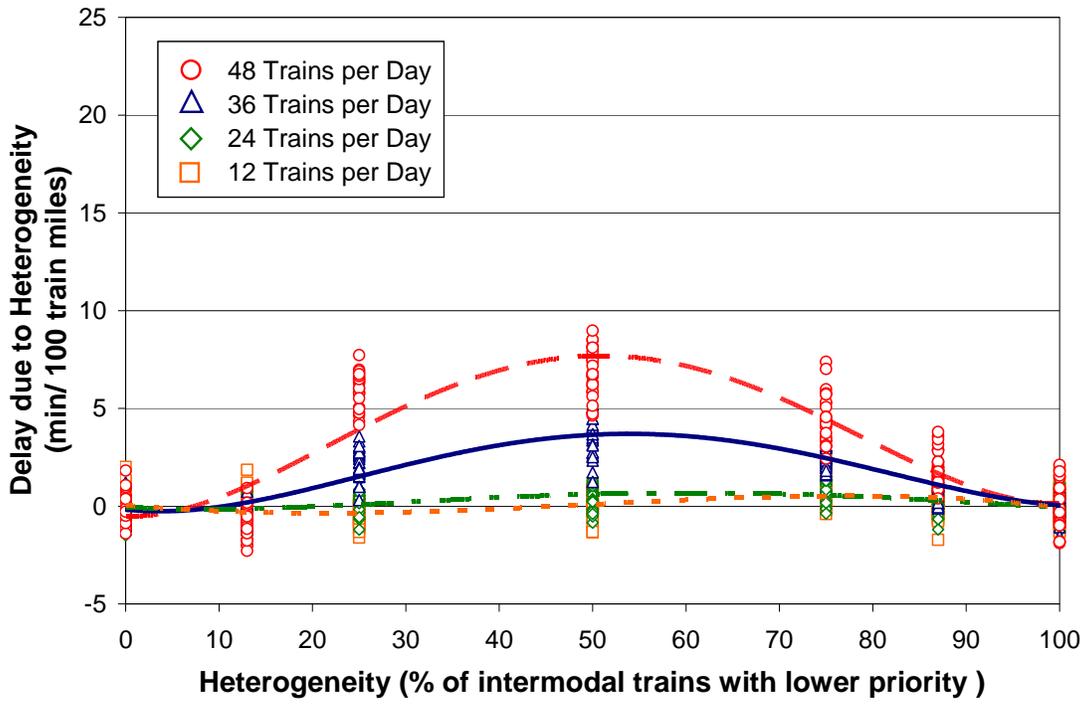
(a)
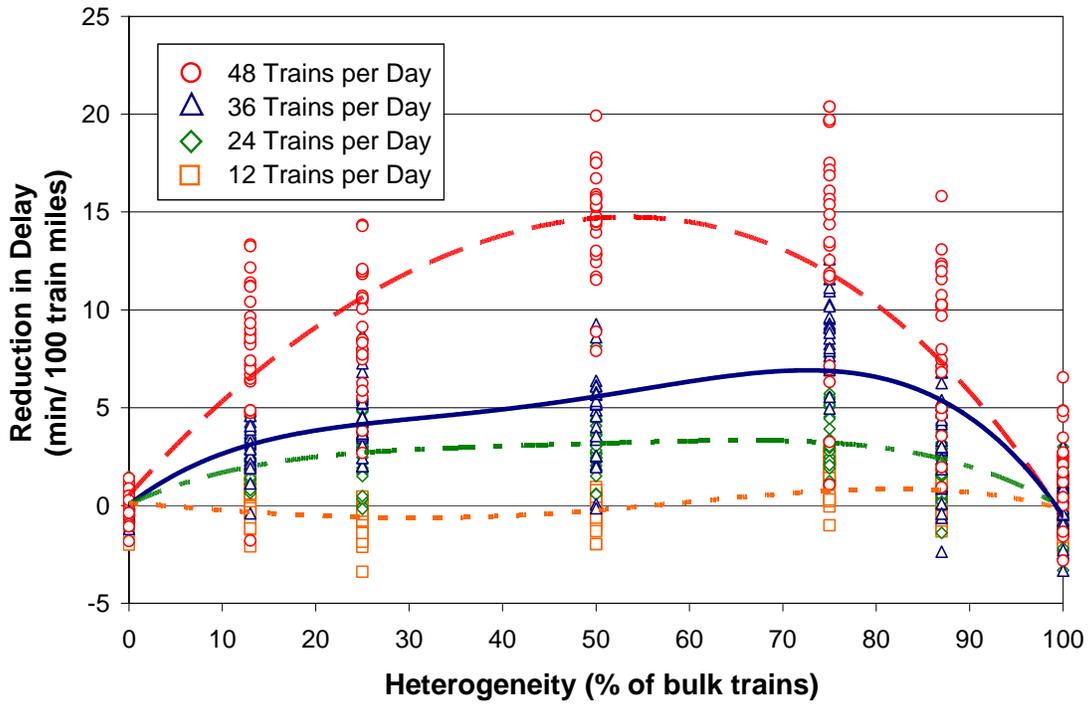
(b)

FIGURE 5.10: Delays due to Heterogeneity with only Heterogeneity in Priority (a) and
Reduction in Delay from Base Scenario when Priorities are Equalized (b)

The magnitude of the delay caused by heterogeneity in train priorities is dependent on the trains characteristics. In typical operations the trains with the lowest priority often have the poorest operating characteristics. Ironically this means that these trains with the slowest acceleration experience the most additional stops due to meets and passes. While the other factors considered contribute to delays with and without heterogeneous traffic, priority only reduces delays due to heterogeneity.

5.5.4.1 Impact of Heterogeneity in Priority by Train Type

To further investigate the effect of priority on train type heterogeneity, the delay to each type of train was individually studied (Figure 5.11). When the priorities are equal the delays to each train type are similar and the delays increase with a greater percentage of bulk trains. Unlike when intermodal is given a higher priority (Figure 5.4), in this case the intermodal trains have more delay than bulk trains. Since it is less likely that a faster train will pass a slower train when there are no priorities, there may not be enough time for the train to reach its top speed before slowing for another meet. With equal priorities, the average delays are less than when the traffic has differential priorities but the reduction comes as a result of increased delays for the higher priority intermodal traffic
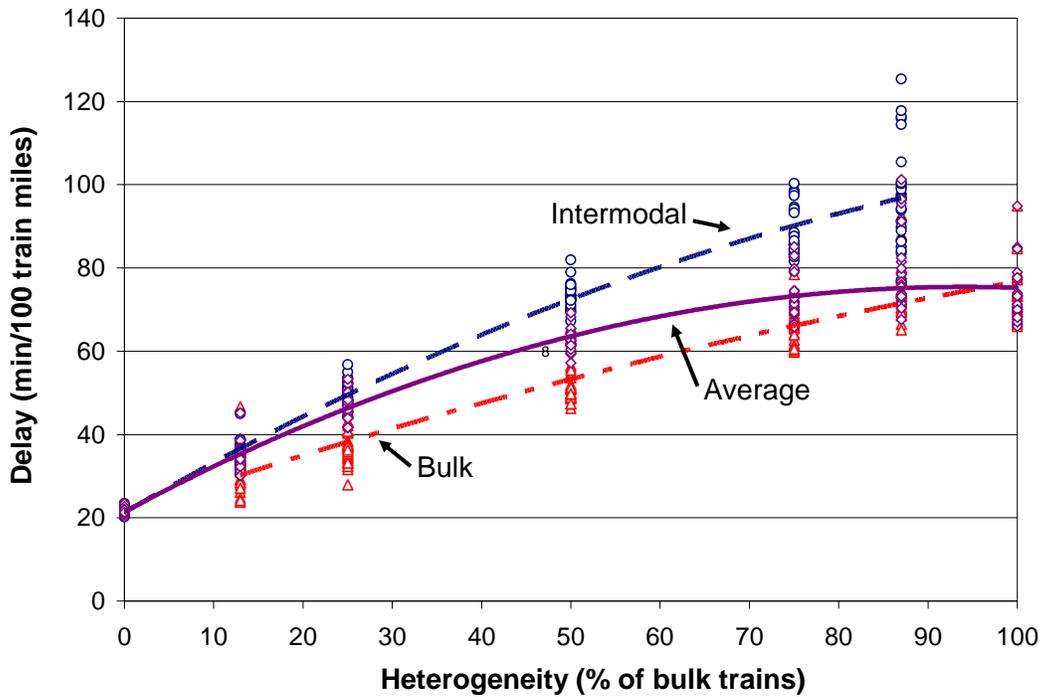
FIGURE 5.11: Delay by Train Type when Priorities are Equal at 48 Trains per Day.

5.6 Discussion

Traffic volume, heterogeneity, and delay are all closely related. Routes that have the highest volumes and are facing the greatest capacity constraints, experience the largest delays due to heterogeneity. The two principal train characteristics that reduce capacity are heterogeneity in train priority and acceleration performance. Neither causes substantial delays due to heterogeneity alone, but in combination they result in additional stops for the lower priority, which are least able to accelerate back to top speed. Other sources of heterogeneity, braking and speed difference, had little effect on delay. On single track, reducing train speed to homogenize traffic increases the occupancy time of the track between sidings thereby increasing delays and reducing capacity.

Although not considered in this study, passenger trains can introduce substantial heterogeneity on a line. They create additional delays because their pertinent characteristics are substantially different than the variation among freight trains. Passenger trains have higher maximum speeds, power-to-ton ratios and dispatching priorities, than all other freight trains. When passenger trains are added to baseline freight schedules, their impact is greater than if the same numbers of freight trains are added. This additional effect needs to be considered when additional passenger trains are proposed for a route. These trains not only take up train slots that could otherwise be used by freight, but they can also create additional delays for existing freight traffic.

Understanding the trends and causes of capacity lost due to heterogeneity among trains is important when planning for new traffic. Train type can be as important as the number of trains when considering the impact on capacity and volume should not be the sole measure of line capacity. This research has shown that even at a constant volume, traffic can experience widely varying delays, depending on the mix of trains. A route may be operating at capacity at a variety of different volumes depending on the traffic mix. Additionally, depending on the current traffic mix, additional traffic will have a different impact on capacity. For example, when the majority of traffic is intermodal the addition of a few bulk trains will have a much greater effect than adding a few intermodal trains to a network operating mostly bulk trains.

5.7 Conclusions

There is increasing demand for freight rail transport in the North America and considerable capital is being invested in new infrastructure. Investing this capital efficiently requires understanding the different operational characteristics of the intended traffic. I performed

analyses using dispatch simulation software to determine the impacts of heterogeneity in freight

traffic. This assessment reveals the relationship between volume, heterogeneity and delay.

Further work identified the key factors that contribute to the increased delays due to

heterogeneity. The train characteristics of speed, acceleration, braking and priority were

considered and the combination of priority and acceleration performance were found to have the

greatest effect.

CHAPTER 6: ANALYSIS OF SPECIFIC FACTORS CONTRIBUTING TO TRAIN DELAY ON A SINGLE TRACK

6.1 Introduction

Railroads make extensive use of simulation to determine the impact of different operations and infrastructure projects on capacity. Estimates of train delay are one of the primary outputs from the simulation analysis used to measure capacity and efficiency. Reduction in delay is often used to measure the potential benefit of a project; however the specific types of conflicts and operational factors that cause these delays are not well understood.

Gorman *(2009)* created a train run-time model from empirical data of eight BNSF subdivisions. Using these data he statistically calculated the relative amount of delay caused by various factors including: meets, passes, headway, secondary effects, priority, and horsepower-to-ton ratio. He identified meets, passes, and overtakes as the principal causes of delay.

Building on the work by Gorman, simulation analyses were conducted and the types of delays were categorized in order to understand the magnitude of delay caused by different factors. The results were studied for trends to develop a better quantitative understanding of the factors that contribute to train delay. Such understanding has intrinsic value to the study of rail capacity and will improve railroads' ability to assess different alternatives and conduct more effective capacity planning.

6.2 Methodology

In order to calculate the impact of heterogeneity and the various mechanisms that impact train delay, I used the simulation software Rail Traffic Controller (RTC). RTC was chosen because its flexibility permits rapid evaluation of a variety of different scenarios and because of its

widespread acceptance and use by the North American railroad industry. For this analysis delay was again defined as the difference between the minimum, or unopposed, run time, and the actual run time required to traverse the route. Using RTC's Train Performance Calculator (TPC) the speed, position, and acceleration data for each train were collected. These TPC data along with the time-distance diagrams were used to identify the conflicts that caused each delay for each train traversing the route. Delay time was then categorized by whether it was accumulated while the train was decelerating, traveling at a constant speed below normal, stopped, or accelerating.

6.2.1 Representative Rail Line and Train Types

The same rail line and train types described in Chapter 5 were used for this analysis. For discussion of the characteristics of the rail line and train types refer to sections 5.2.3 and 5.2.4.

6.2.2 Simulations

Simulations for this study were developed for seven different levels of heterogeneity with a constant volume of 40 trains per day. The trains were evenly distributed in each direction over a 24-hour period. The seven different levels of heterogeneity were based on the percentage of the different train types. Tests were conducted with 0%, 12.5%, 25%, 50%, 75%, 87.5% and 100% of each train type. The level of heterogeneity corresponds to the percentage of one train type relative to the other. For instance, 12.5% intermodal trains corresponds to five intermodal trains and thirty-five bulk trains. Train sequence was directly proportional to the percentage of each train type so in the example above each intermodal train would be followed by seven bulk trains. The ratios and traffic patterns were the same for trains traveling in both directions. For each

configuration a series of five simulations was performed with the exact departure time of each train randomized according to a uniform distribution over a 30-minute interval, 15 minutes before or after the scheduled time for that train. In each simulation a sample of sixteen trains was analyzed, eight in each direction with the number of trains of each type corresponding to the percentage of that train type in the scenario.

6.2.3 Factors of Delay

The delays were categorized by conflict type and sources. The conflict types considered were meets, passes, and mainline (Table 6.1). Meets were classified as any delay due to conflicts with one or more trains traveling in the opposite direction. Passes were classified as any delay due to conflicts with one or more trains traveling in the same direction, resulting in one train overtaking another. When a conflict involved multiple meets and passes the acceleration and braking delays were attributed to the first conflict while the extra dwell time required to accommodate the additional conflicts was attributed to each subsequent conflict accordingly. Mainline conflicts were classified as any delay where one train was required to slow down due to a preceding train traveling in the same direction that did not result in an overtake.

For each conflict the specific operational source of delay was also identified. These sources include the delays while a train is braking, accelerating, traveling at constant speed slower then normal or stopped (Table 6.1). By separating the delay by conflict and source, it is possible to determine which type of conflict creates the most delay, why a specific delay is occurring, and how the delay changes with changes in traffic composition.

TABLE 6.1: Categories of Delays

Conflicts	Sources
Meets	Accelerating
Passes	Braking
Line	Reduced Speed
	Stopped

6.3 Analysis of Causes of Delay

The delays due to each conflict and source of delay are combined to determine the total delays

for each traffic mix (Figure 6.1). As discussed in Chapter 5, delays are greatest with

heterogeneous traffic. However, the longest delays do not occur with the greatest heterogeneity,

but rather when the traffic is a mostly bulk trains. Additionally, delays are different when traffic

is homogenous depending on the train type: bulk trains experience greater delays than intermodal

trains.

The analysis considers a smaller sample size than used in Chapter 5 and not all delays

that occurred in the simulations had causes that could be unambiguously assigned to a particular

source. However, these delays were few and the difference between the average delays obtained

in Chapter 5 with 25 simulations and those categorized in this analysis from five simulations is

less than 5%.
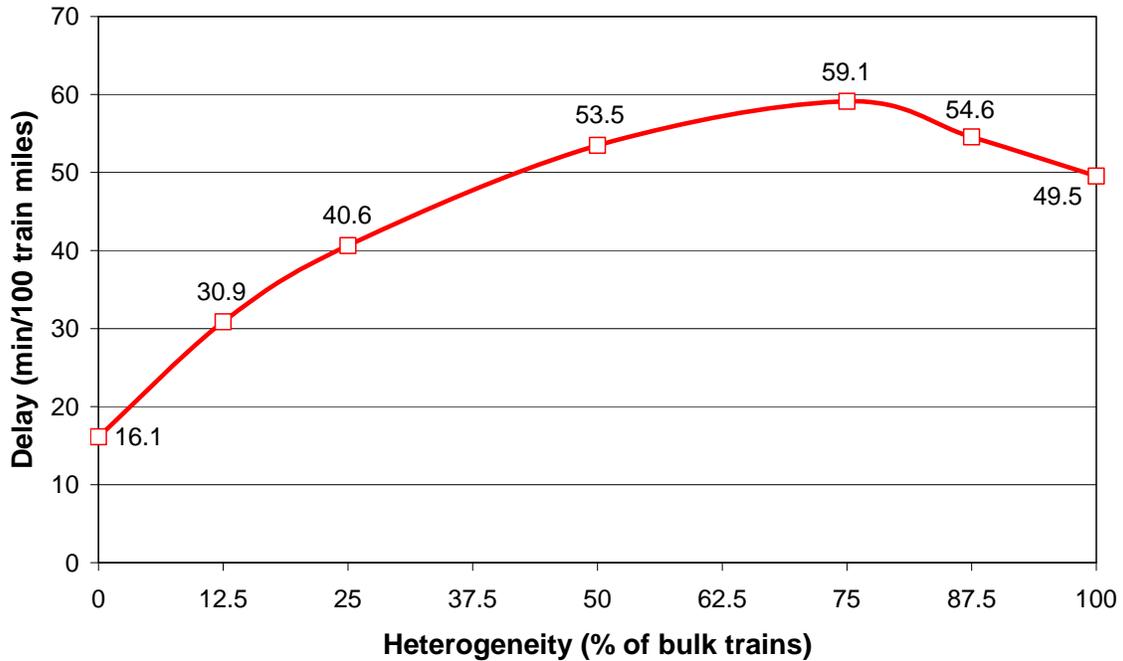
FIGURE 6.1: Average Total Delays with Different Traffic Mixes

6.3.1 Delays with Homogenous Traffic

Different train types have different operating characteristics and this difference in characteristics has a direct impact on the magnitude of delay experienced by each. 40 trains per day of homogenous bulk train traffic had an average delay of 49.5 minutes per 100 train miles (Figure 6.1). Whereas homogenous intermodal traffic at the same volume was only 16.1 minutes per 100 train miles. This difference is due to larger delays while each train is braking, accelerating, or stopped (Figure 6.2). The reduced speed delays were minor and the same for each train type. The additional delays while braking and acceleration are due to the slower braking and acceleration performance of the bulk trains. The larger delay while stopped is due to both the poorer acceleration and braking performance and the slower maximum speeds of the bulk train.

Because of the slower speed, travel times between sidings is increased thereby requiring trains to be stopped longer waiting to resolve meets.
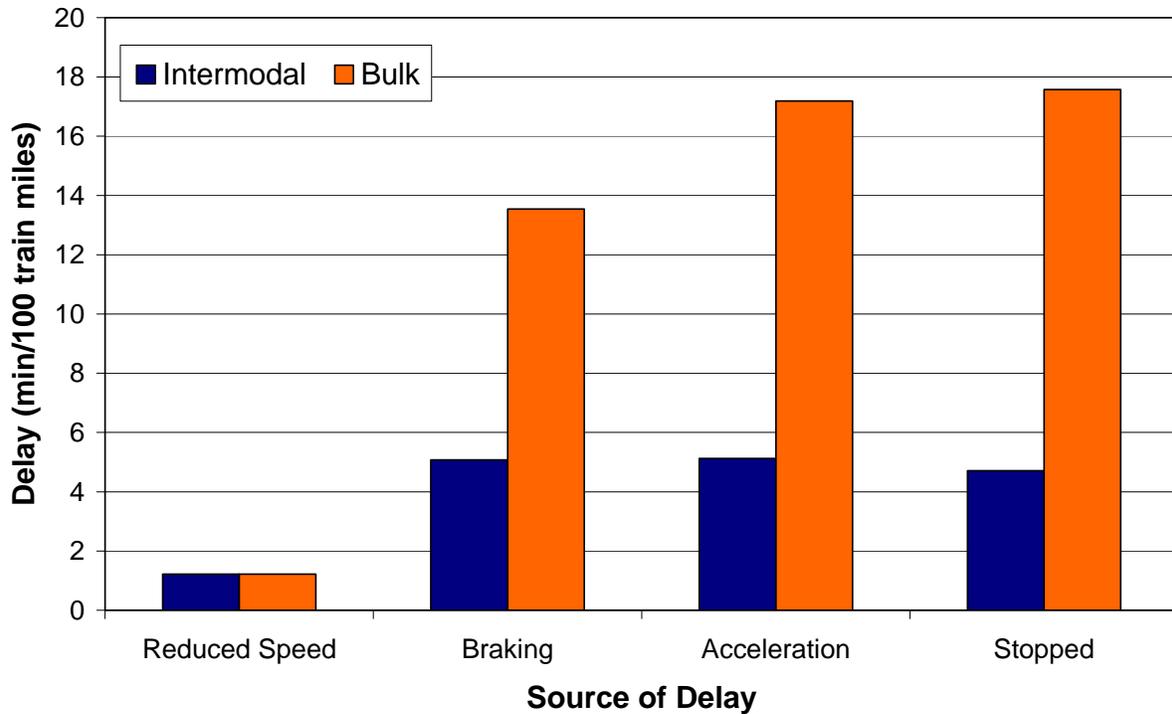
FIGURE 6.2: Delays by Source with Homogenous Traffic

The greater delays for bulk trains are due to the additional time required to accelerate and brake and the greater number of conflicts while en route. Bulk trains averaged 2.97 conflicts per 100 train miles while intermodal trains averaged 1.74 conflicts per 100 train miles. The slower average speeds of the bulk trains results in longer occupancy of the route and consequently, more meets. Additionally, while all the conflicts with intermodal traffic were simple, one-train meets, the conflicts with bulk traffic were more complex. Several conflicts with homogenous bulk traffic were passes or multiple train meets. These more complex meets increase the time a train is stopped in a siding and consequently increase delay.

Since the two train types experience, on average, a different number of meets and passes, the delays resulting from train stops were normalized for each event (Figure 6.3). The delays

while braking, accelerating and stopping per conflict are larger for bulk trains. On average it takes over 7 minutes longer for a bulk train to resolve a conflict compared to an intermodal train. The delays do not directly correspond to the characteristics of the trains (characteristics of each train type are discussed in section 5.2.4). The braking distance of a bulk train is 15% longer than an intermodal train, however, delay while braking is 55% greater. Likewise, intermodal trains take 326% longer to reach 30 mph and 476% longer to reach 45 mph, however the bulk train delay while accelerating is only 96% greater. These relationships will vary depending on the specific traffic and route characteristics.
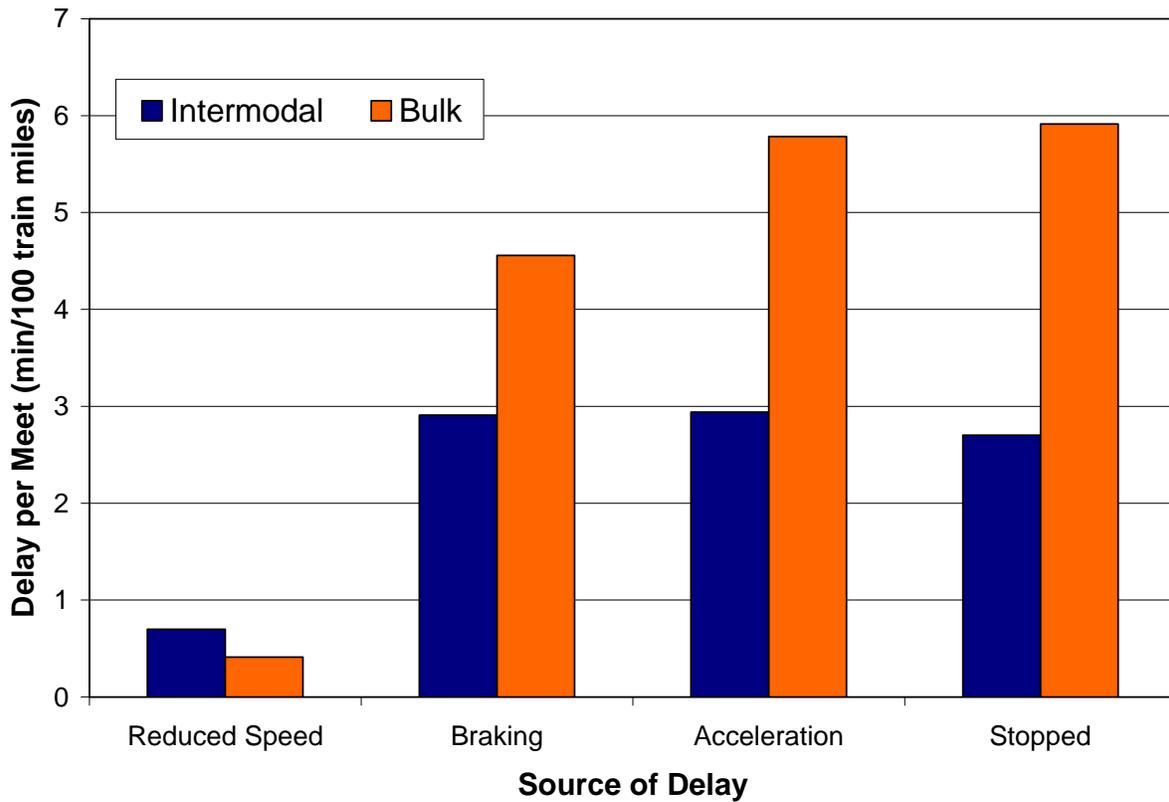
FIGURE 6.3: Delays by Source per Meet or Pass Conflict with Homogenous Traffic

6.3.2 Delays with Heterogeneous Traffic

6.3.2.1 Conflicts that Cause Delays with Heterogeneous Traffic

When separated by conflict type the results show that each type of delay changes differently with changing traffic (Figure 6.4). The delays due to passes are greatest at the highest levels of heterogeneity, and mainline delays are the same with all levels of heterogeneity. The delays from meets closely follow the trend of the average delays. Consequently, the increased delay due to heterogeneity is primarily from increased meet delays.
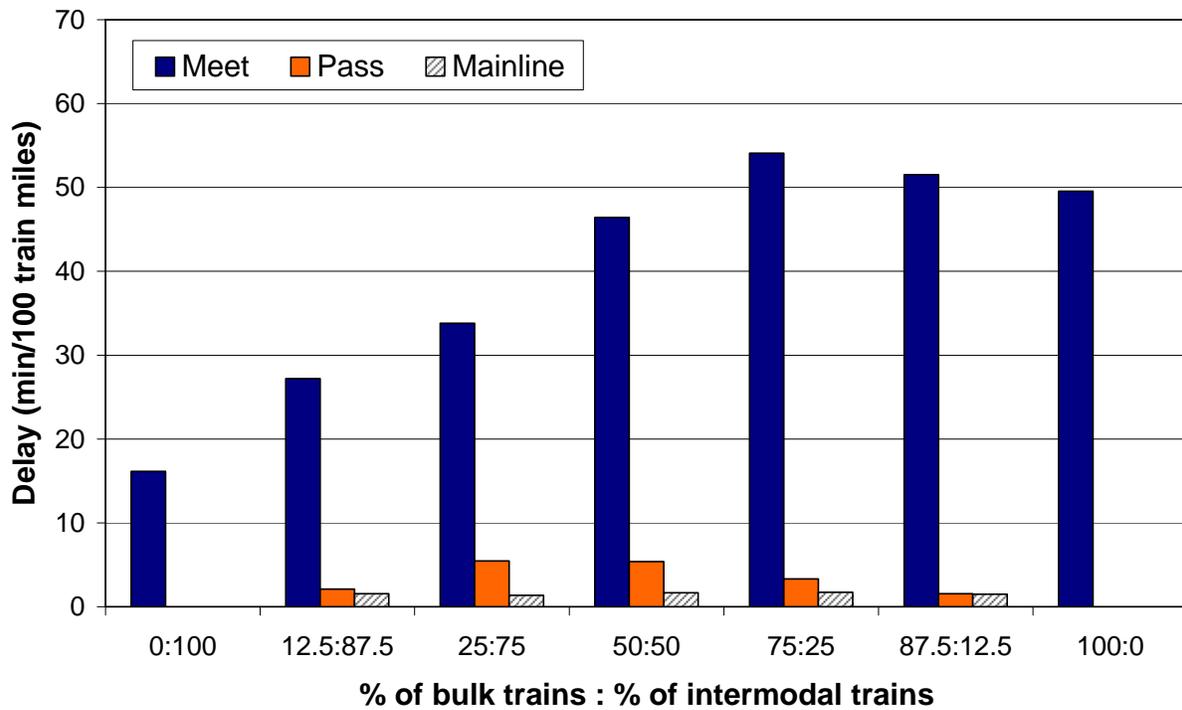
FIGURE 6.4: Average Delay by Conflicts

In addition to the delays, the number and type of conflicts per 100 train miles were recorded (Table 6.3). Nomenclature for the meet and pass types is included in Table 6.2. The results provide interesting information on the complexity of operations as heterogeneity increases. The number of passes increases with increasing heterogeneity, however there is not a

clear trend for meets. The number of meets increases until there are 75% bulk trains, drop when

the traffic is 87.5% bulk trains, but reach a maximum when the traffic is 100% bulk trains.

When the traffic is mostly bulk with a few higher priority intermodal trains the number of meets

is actually less; however, the meets are longer and more complex.

TABLE 6.2: Nomenclature for Types of Conflicts

Nomenclature:
M1P0: Meets 1 train
M2P0: Meets 2 trains
M1P1: Meets 1 train and is passed by 1 train
M2P1: Meets 2 trains and is passed by 1 train
M0P1: Passed by 1 train
M3P0: Meets 3 trains
M3P1: Meets 3 trains and is passed by 1 train
M2P2: Meets 2 trains and is passed by 2 trains
M3P2: Meets 3 trains and is passed by 2 trains
M1P2: Meets 1 train and is passed by 2 trains
M4P0: Meets 4 trains
M0P2: Passed by 2 trains

The complexity of the meets increases with increased heterogeneity. Most of the

conflicts are a single train meeting a single train, referred to as a simple meet. However, as

heterogeneity increases both the number of meets or passes per conflict and delay per conflict

increase (Table 6.4). With each additional train involved in a meet, the delay increases

disproportionally. A one-train meet results in an average of 16.15 minutes of delay, a two-train

meet, 36.02 minutes of delay and a three-train meet, 52.42 minutes.

TABLE 6.3: Number of Conflicts per 100 Train Miles

Bulk:Intermodal	M1P0	M2P0	M1P1	M2P1	M0P1	M3P0	M3P1	M2P2	M3P2	M1P2	M4P0	M0P2
0:100	1.743											
12.5:87.5	1.806	0.067	0.034	0.005	0.101	0.034		0.010		0.005		
25:75	1.541	0.158	0.115	0.058	0.115	0.034	0.010		0.005			0.005
50:50	1.551	0.355	0.144	0.034	0.154	0.019	0.005					
75:25	1.806	0.394	0.062	0.024	0.110	0.034	0.005				0.005	
87.5:12.5	1.801	0.235	0.010	0.005	0.058	0.010						
100:0	2.876	0.072	0.010		0.014							

TABLE 6.4: Delay per Conflict

Bulk:Intermodal	M1P0	M2P0	M1P1	M2P1	M0P1	M3P0	M3P1	M2P2	M3P2	M1P2	M4P0	M0P2
0:100	9.01											
12.5:87.5	11.11	28.90	42.98	58.05	16.48	40.62		121.07		63.70		
25:75	12.64	34.05	40.96	70.05	15.57	55.31	103.57		130.32			29.15
50:50	15.82	34.68	52.56	84.13	15.27	51.73	86.97					
75:25	17.14	37.54	49.51	73.28	19.03	60.29	91.73				86.52	
87.5:12.5	17.74	38.30	50.15	82.00	17.33	70.90						
100:0	18.48	37.78	46.73		33.49							
Average	14.92	36.02	47.41	74.46	17.00	53.42	96.46	121.07	130.32	63.70	86.52	29.15

6.3.2.2 Sources that Cause Delays with Heterogeneous Traffic

Each source of delay has a different trend with regard to traffic mix (Figure 6.5). Delays due to

trains traveling at a reduced speed were minor and relatively constant over all traffic mixes. The

delays while a train is braking and accelerating increased with higher percentages of bulk trains.

These delays do not increase with heterogeneity and are therefore due to greater numbers of bulk

trains, which have longer accelerating and braking delays per meet. As discussed previously,

bulk trains experience more delay during each type of conflict. Consequently, more bulk trains

will result in greater acceleration and braking delays.
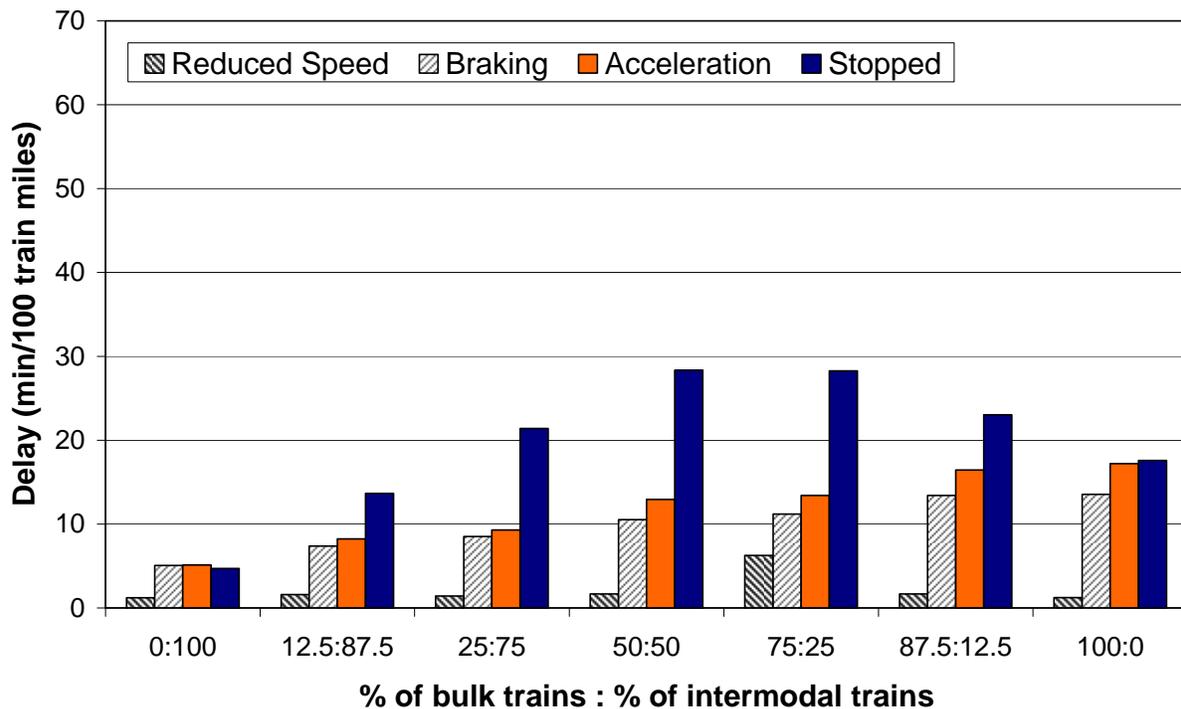
FIGURE 6.5: Average Delay by Source

The delay while stopped is the only source of delay that increased with heterogeneity.

Consequently, the delay increase associated with greater heterogeneity is due to a longer amount

of time the trains are stopped waiting in sidings. This is potentially due to two reasons. First, at

73

higher heterogeneity levels there is a greater likelihood for less efficient meets when two trains of different priorities meet. Sometimes when two trains of different priorities meet the siding that would result in the least delay is not used in order to reduce delays to the higher priority train. The lower priority train will enter an earlier siding and wait for the higher priority train, resulting in greater overall delays. Secondly, meets tend to be more complex when traffic is more heterogeneous. If a train is met or passed by more than one train, the time stopped in the siding is longer.

6.4 Discussion

By categorizing and analyzing delays, the specific conflicts and sources that contribute to delay can be identified and measured. This enables a better understanding of why heterogeneous traffic causes more delay. The simulation analyses conducted in this study showed that increased delay was the result of trains waiting longer in sidings to resolve additional and more complex meets when traffic heterogeneity was highest.

In Chapter 5 (Section 5.1) five situations with heterogeneous traffic that caused additional delays are listed. The analysis conducted in this chapter investigated the impact of each of these situations on delay. Delays due to a slower preceding train and time lost during a pass are minor. While these situations cause delays, the increase in the number and length of meets is the primary cause of the additional delays due to heterogeneity.

The specific train type characteristics affect the amount of delay they experience. The bulk trains, with their poor acceleration and braking performance and slower speeds, experience more delays. Improving any of those characteristics has the potential to reduce delays and improve capacity.

Time spent stopped in a siding for meets was the leading cause of delay suggesting that any efforts to reduce train delay should focus on this. Possible methods of reducing this delay include increasing train speed, reducing siding spacing, equalizing priorities, and adding a second track. Increasing speed will reduce the time needed for one train to pass another, either during a pass or a meet. Reducing siding spacing will allow trains to stop closer to the conflict. Removing priority makes meets more efficient: instead of forcing the lower priority train to stop, the first train to arrive will enter the siding. Lastly, adding a second track will completely eliminate delay time from meets. Of course each of these has different capital and/or operating cost implications that would need to be accounted for in a cost/benefit analysis.

6.5 Conclusions

The primary output from the simulation analyses was train delay. Reduction in delay is often used to determine the benefit of a project or operational change. Although delay is used to make decisions the specific causes of the delay are not well understood. Using the internal TPC of the simulation software the delays were categorized by the specific conflict or operational situation that caused the delay. Results showed that the only source of delay that substantially increased due to heterogeneity was the time trains spent stopped in sidings to resolve meets. Using this information, the best method of improving operations is to make changes that either reduce the number of meets or reduce the time a train is stopped while in a meet.

TABLE 6.5: Amounts of Delay per 100 Train Miles for Each Source and Conflict

	Bulk:Intermodal	0:100	12.5:87.5	25:75	50:50	75:25	87.5:12.5	100:0
Meet	Accel	5.1	7.1	7.6	11.1	12.0	15.4	17.2
	Brake	5.1	6.4	6.9	8.9	9.8	12.5	13.5
	Reduced Speed	1.2	1.2	1.1	1.1	5.7	1.3	1.2
	Stopped	4.7	12.5	18.2	25.4	26.6	22.3	17.6
Pass	Accel	0.0	0.5	1.2	1.2	0.8	0.5	0.0
	Brake	0.0	0.4	1.0	1.0	0.6	0.4	0.0
	Reduced Speed	0.0	0.1	0.2	0.2	0.3	0.1	0.0
	Stopped	0.0	1.1	3.1	2.9	1.6	0.7	0.0
Mainline	Accel	0.0	0.6	0.5	0.6	0.6	0.6	0.0
	Brake	0.0	0.6	0.6	0.7	0.7	0.5	0.0
	Reduced Speed	0.0	0.3	0.1	0.4	0.3	0.3	0.0
	Stopped	0.0	0.1	0.2	0.0	0.0	0.0	0.0
	Total	16.1	30.9	40.6	53.5	59.1	54.6	49.5

TABLE 6.6: Delay per Meet or Pass Conflict

Bulk:Intermodal	Acceleration	Braking	Reduced Speed	Stopped	Total
0:100	5.1	5.0	1.2	4.6	15.8
12.5:87.5	8.1	7.3	1.6	13.5	30.5
25:75	9.2	8.4	1.4	21.2	40.2
50:50	12.8	10.4	1.7	28.0	52.9
75:25	13.3	11.1	6.2	27.9	58.5
87.5:12.5	16.3	13.3	1.7	22.8	53.9
100:0	17.0	13.4	1.2	17.4	49.0

TABLE 6.7: Meets, Passes and Conflicts per 100 Train Miles

Bulk:Intermodal	Meets	Passes	Conflicts
0:100	1.74	0.00	1.74
12.5:87.5	2.11	0.17	2.06
25:75	2.23	0.32	2.04
50:50	2.55	0.34	2.26
75:25	2.84	0.20	2.44
87.5:12.5	2.32	0.07	2.12
100:0	3.03	0.02	2.97

CHAPTER 7: MITIGATING TRAIN TYPE HETEROGENEITY ON A SINGLE-TRACK LINE

7.1 Introduction

Demand for freight railroad transportation is increasing *(AASHTO 2007)*. In order to accommodate this new traffic, changes to railroad operations and infrastructure will be required. Infrastructure expansion requires long lead times and is capital intensive. Alternatively, a less expensive and faster means of creating additional capacity may be possible through changes in operations. Building on the previous chapters, various methods to reduce train delay and increase capacity were considered for their effectiveness and economic benefit. Simulation software was used to investigate the impact of various operational changes with different traffic mixes and volumes on a hypothetical, signalized, single-track rail line.

7.2 Methodology

Multiple operational and infrastructure scenarios were considered in order to calculate the effectiveness of various methods of reducing delay. Dispatch simulation software was used to simulate multiple traffic scenarios, with train delay being the primary metric to measure capacity and the cost of train operations. This chapter uses the same software and definition of delay used in previous chapters.

Simulations were conducted using three different representative rail lines in order to determine the comparative benefit of changing operations as opposed to modifying or expanding infrastructure. The base simulations and subsequent operational changes were made on a route with 20 miles between sidings. These changes were then compared to the improvements obtained when the siding spacing is reduced or a second track is added.

77

7.2.1 Representative Rail Lines

Specific characteristics of individual rail lines are unique and route characteristics influence the study of railroad operations. For my research I developed three hypothetical rail lines intended to represent the characteristics of typical North America midwestern mainline subdivisions (Table 7.1).

TABLE 7.1: Routes Used in Analysis

Single track with 20 mile siding spacing
262 miles long
20 miles between siding centers
8,700 ft signaled sidings with #24 powered turnouts
2.55 mile signal spacing
2-block, 3-aspect signaling
0% grade and curvature

Single track with 10 mile siding spacing
262 miles long
10 miles between siding centers
8,700 ft signaled sidings with #24 powered turnouts
2.75 mile signal spacing
2-block, 3-aspect signaling
0% grade and curvature

Two tracks with 10 mile crossover spacing
260 miles long
10 miles between crossovers
Universal crossovers with #24 powered turnouts
2.75 mile signal spacing
2-block, 3-aspect signaling
0% grade and curvature

Although the attributes are somewhat idealized, the purpose is to provide a consistent basis for relative comparison of different scenarios of interest in this research under a reasonably realistic set of operating conditions. Track maintenance and other factors that can affect capacity were

not considered and there is no intent to imply that the results presented here predict absolute measurements for a particular set of conditions.

7.2.2 Train Types

The same train types described in Chapter 5 were used for this analysis. See Section 5.2.4 for discussion of the characteristics of these trains.

7.2.3 Simulations

Simulations were conducted at different traffic volumes and levels of heterogeneity. Eleven different volumes were simulated ranging from 8 to 40 trains per day, simulated in increments of four. The simulations at each volume used a uniform average temporal distribution of trains in each direction over a 24-hour period. The results are intended to provide a basis for relative comparison of the effect of various factors of interest; however, inspection, maintenance and other factors that affect capacity were not considered. Consequently, the results are more representative of the possible spacing or headway between trains than the actual train volume on a line. Seven different levels of heterogeneity were tested based on the percentage of the different train types. The tests were done with 12.5%, 25%, 50% and 100% of each train type. The level of heterogeneity corresponds to the percentage of one train type relative to the other. Train sequence was directly proportional to the percentage of each train type; for instance when the traffic is 12.5% intermodal, each intermodal train would be followed by seven bulk trains. The ratios and traffic patterns were the same for trains traveling in both directions. For each configuration a series of twenty-five simulations was performed with the exact departure time of

each train randomized according to a uniform distribution over a 30-minute interval, 15 minutes before or after the scheduled time for that train.

7.3 Analysis of Operational and Infrastructure Changes to Improve Capacity

One of the implications of understanding the factors that influence delay is that this knowledge can be used to determine the most effective methods of reducing delay. On a single-track route meets are the principal source of delay and consumer of capacity; therefore, reducing the number and duration of meets will have the largest effect on capacity. Several methods are possible to reduce delay due to meets. Increased power (e.g. using additional locomotives) increases the acceleration rate thereby reducing the time to leave a siding after a meet, increased train speeds reduce the dwell time of a train waiting in a siding to meet other trains, and removing dispatching priorities reduces the number of meets. Each method may provide benefits due to reduced delays and increased capacity. Infrastructure expansion will also reduce delay time and improve operations. If sidings are more closely spaced it reduces the time trains must wait for oncoming traffic, and adding a second track eliminates meets all together. Although infrastructure expansion is an effective method of increasing capacity, it is also expensive. One of the objectives of the analyses described here is to develop a better understanding of how operational changes compare to infrastructure in terms of capacity. Both operational and infrastructure changes were simulated and compared to the base scenario with 20 miles between sidings.

7.3.1 Base Route

The base route used for this analysis was a single track segment with 20 miles between sidings.

At a traffic level of 32 trains per day or less, the delays are greatest when heterogeneity is highest

(Figure 7.1). However, when traffic volume increases to 40 trains per day the delays increase

substantially especially when the traffic is mostly bulk trains. In this case, the delays are greater

with homogenous bulk trains than with heterogeneous traffic. The sharp increase in delays for

bulk traffic is symptomatic of a route nearing or reaching its capacity. At this point railroads

would begin to consider alternatives to improve the capacity in order to reduce delays to current

and future traffic.
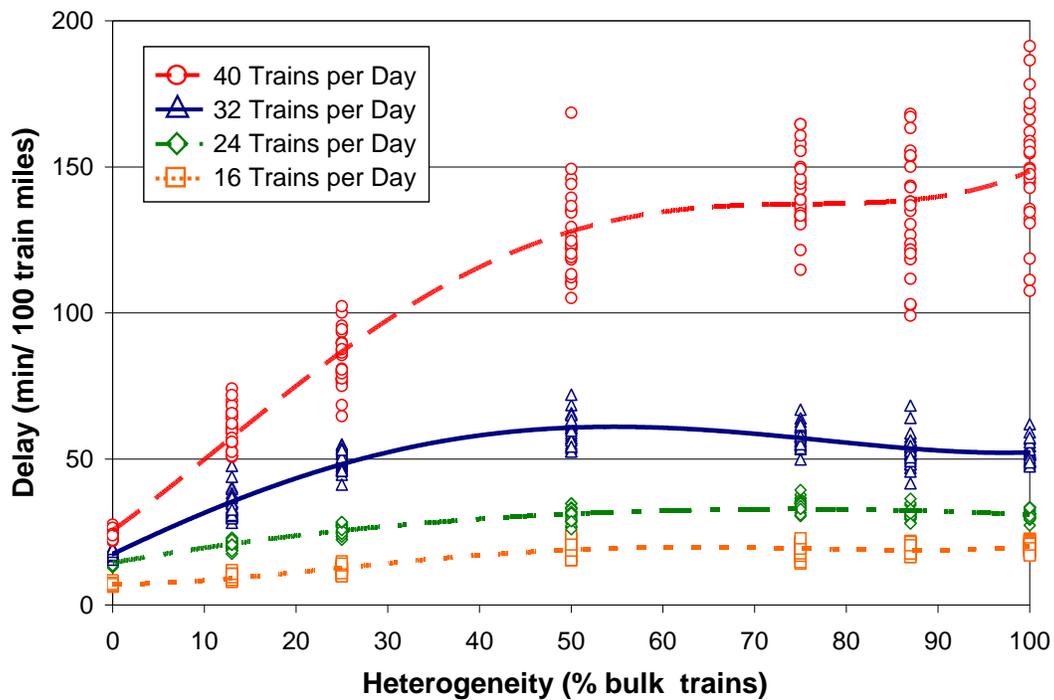
FIGURE 7.1: Delays on Baseline Route and Train Types

7.3.2 Additional Locomotive for Each Bulk Train

The first alternative considered for its potential to reduce delays is to add an extra locomotive to each bulk train. The additional locomotive increased the horsepower-to-trailing-ton (HPTT) ratio for the bulk trains from 0.78 to 1.05, thereby reducing acceleration distance (Figure 7.2). In Chapter 5 I showed that acceleration distance has a significant impact on train delay for heterogeneous traffic. Faster acceleration reduces the time that trains occupy the mainline traveling below normal maximum speed. This reduces the time lost in meets, thereby reducing overall run time. However additional locomotives require major capital investment as well as additional maintenance and fuel costs.
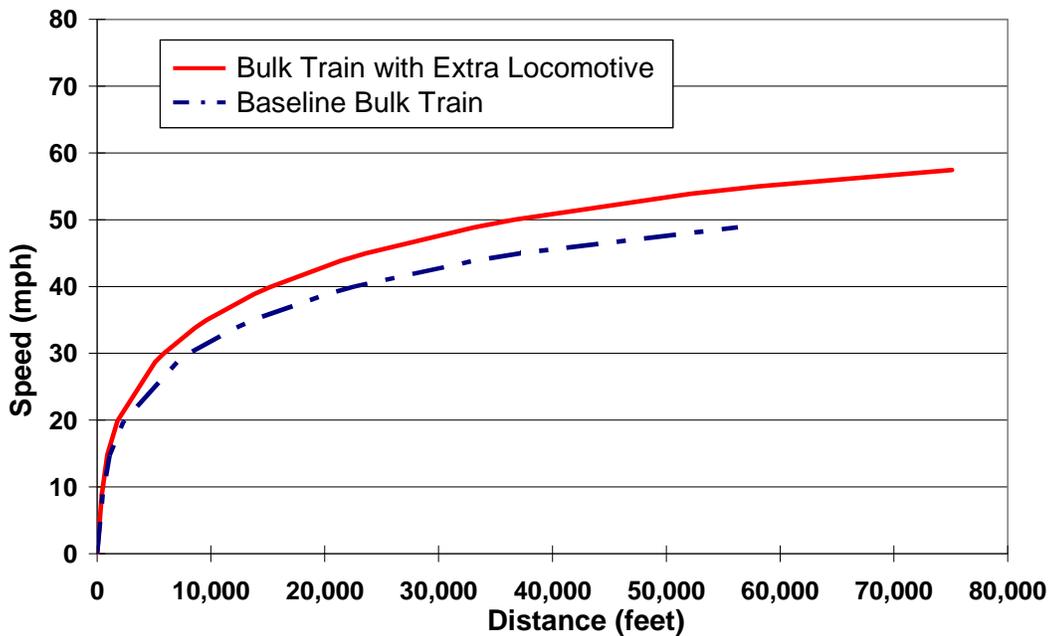
FIGURE 7.2: Acceleration Curve of Baseline Bulk Trains and with an Additional Locomotive with the Maximum Speed Increased to 60 mph.

The reduction in delay from adding locomotives increases with the percentage of bulk trains (Figure 7.3). This is expected because I showed previously (section 6.3.2) that delays due to acceleration and braking performance are directly related to the traffic mix. As the number of

trains with poorer braking and acceleration performance increases, the delays due to these activities also increases. The reduction in delay due to additional locomotives is also related to traffic volume. The benefit at 40 trains per day is much greater than with 32 trains per day. When the volume is 40 trains per day the route is heavily congested and nearing its maximum capacity. Consequently, trains are more frequently delayed by trains ahead of them that have not yet reached their normal maximum speed. Under these circumstances the benefit of improved acceleration performance is more pronounced. Furthermore, at higher volumes there are more conflicts in which a train must stop and then accelerate back to full speed, so the improved acceleration performance helps here as well.

7.3.3 Additional Locomotive and 10 mph Speed Increase for Each Bulk Train

Additional locomotives not only improve acceleration performance but also permit trains to reach higher top speeds. Consequently, the effect of a 10 mph speed increase for bulk trains was tested. Increasing the speed of slower moving bulk trains will reduce conflicts due to the speed difference between train types. More importantly, increased speeds reduces the time to travel between sidings, thereby improving capacity. However, an increase in train speed increases train stopping distance, so this must also be accounted for.

The combined effect of increased train speeds and additional power leads to a greater reduction in delay than additional power alone (Figures 7.3 and 7.4). At higher speeds the trains can traverse the route faster and reduce potential conflicts. As in the previous case, without the speed increase, the delays at the highest volumes are greatest when bulk trains are the majority. However there are greater reductions in delay with heterogeneous traffic. Consequently, the speed increase reduces conflicts between trains.
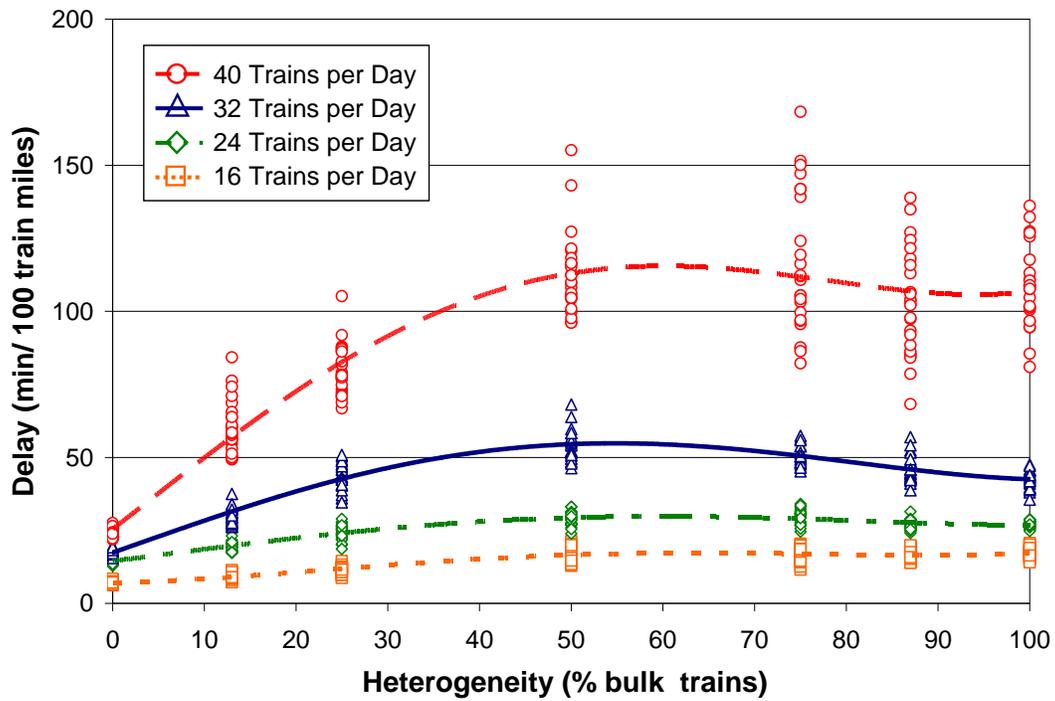
(a)
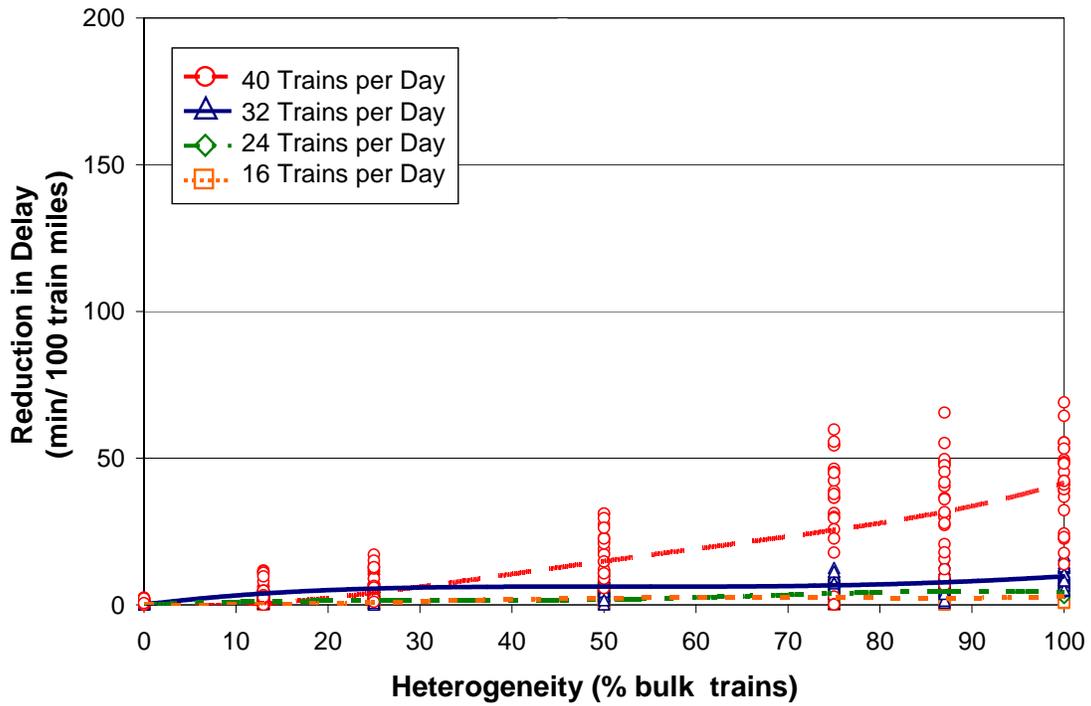
(b)

FIGURE 7.3: Delays (a) and Reduction in Delay (b) When all Bulk Trains Have One Extra Locomotive
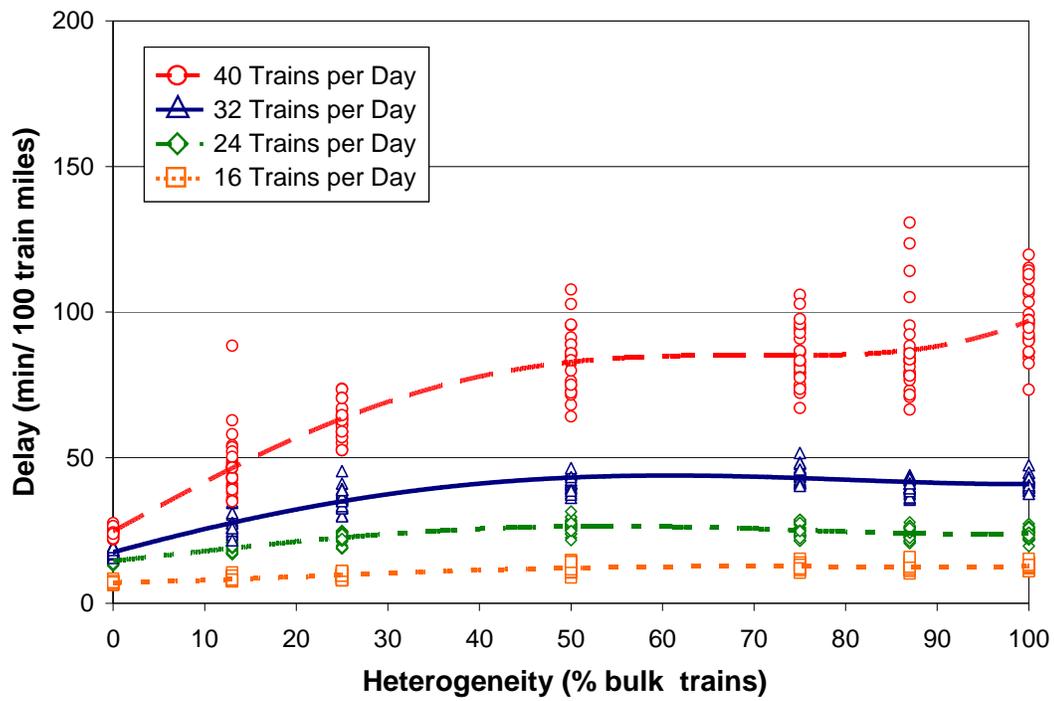
(a)
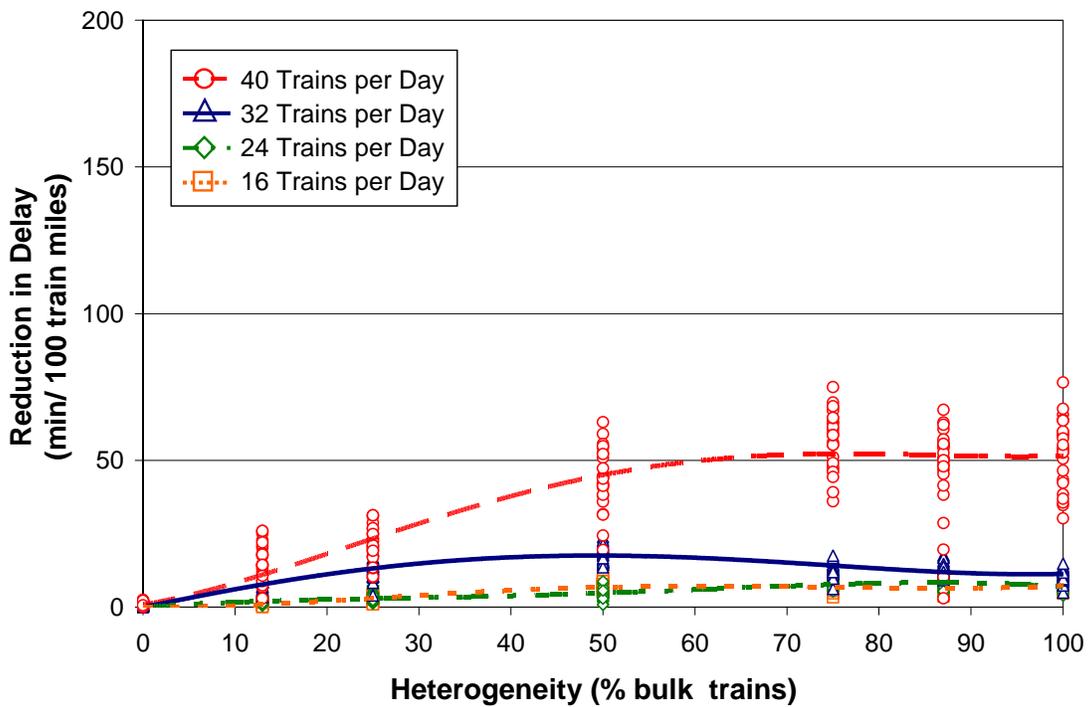
(b)

FIGURE 7.4: Delays (a) and Reduction in Delay (b) When All Bulk Trains Having an Extra Locomotive and 10 mph Speed Increase

Reduction in delay does not explain all the benefits of higher speed. By defining delay as the difference between the minimum and actual run times, the benefit of a reduction in minimum run time is not included in delay calculations. Increasing the average speed of the bulk trains from 50 to 60 mph reduces the minimum run time by 21.13 minutes per 100 train miles. Consequently, the benefit of increased speeds is due to both a reduction in delay and a reduction in run time.

7.3.4 Equalizing Priorities

Chapter 5 identified priority as a significant factor influencing delay on a single track rail line. Consequently, I analyzed equalizing priorities as a potential operational method to reduce delay. In typical operations, the lower priority train will stop and wait at an earlier siding in order to prevent delays to the higher priority train. Consequently, heterogeneous priorities increase both the number and duration of meets. When traffic has the same priority during a meet the first train to arrive at a siding will enter the siding and wait for the oncoming train. In the preceding examples the higher priority trains also had a higher maximum speed, but in this analysis the faster trains have the same priority as the slower ones. Therefore, when a faster train overtakes a slower train it will not result in a pass unless the preceding train is already stopping for a meet.

One of the advantages of equalizing priorities is the limited additional cost, since no additional infrastructure or equipment is required. However, as discussed in section 5.5.4.1 equalization of priorities increases delays to the train type that previously received the higher priority. Thus, there is a trade off between greater delays to the trains whose customers expect higher service quality versus reduction in delays to trains with less demanding schedule requirements.

Equalizing priorities reduces the delays due to heterogeneous traffic but has no effect when the traffic is a single train type (Figure 7.5). The reduction in delay was greatest when the traffic was 25% bulk and 75% intermodal. As the percentage of bulk trains increased and volumes decreased, the benefit of the equalization in priorities was reduced. At lower volumes the delays due to heterogeneity are less; consequently, the potential benefits of equalizing priorities are also less.

7.3.5 Adding Extra Sidings

Reducing siding spacing on a single track line increases capacity because it allows there to be more meets and passes *(Krueger 1999)*. In order to understand the benefit of adding sidings the siding spacing was reduced from 20 miles to 10 miles. This permits a comparison of the relative benefits of operational changes to infrastructure expansion. However, closer siding spacing also reduces delay because it allows for meets or passes to be resolved at more ideal locations, reducing the time trains must wait in sidings. The complexity of meets is also reduced because it is possible for a train to advance to a siding instead of waiting for another train to meet or pass at the current siding.

Construction of additional sidings requires both upfront capital cost for construction and operating expense for maintenance. On many rail lines, sidings are unevenly spaced leading to bottlenecks. These are easily identifiable and additional sidings can be built at the appropriate locations. The route used for this analysis however, has evenly spaced siding and no inherent bottlenecks. Therefore, when adding sidings in these simulations, they were placed between each pair of sidings in order to not create new bottlenecks.
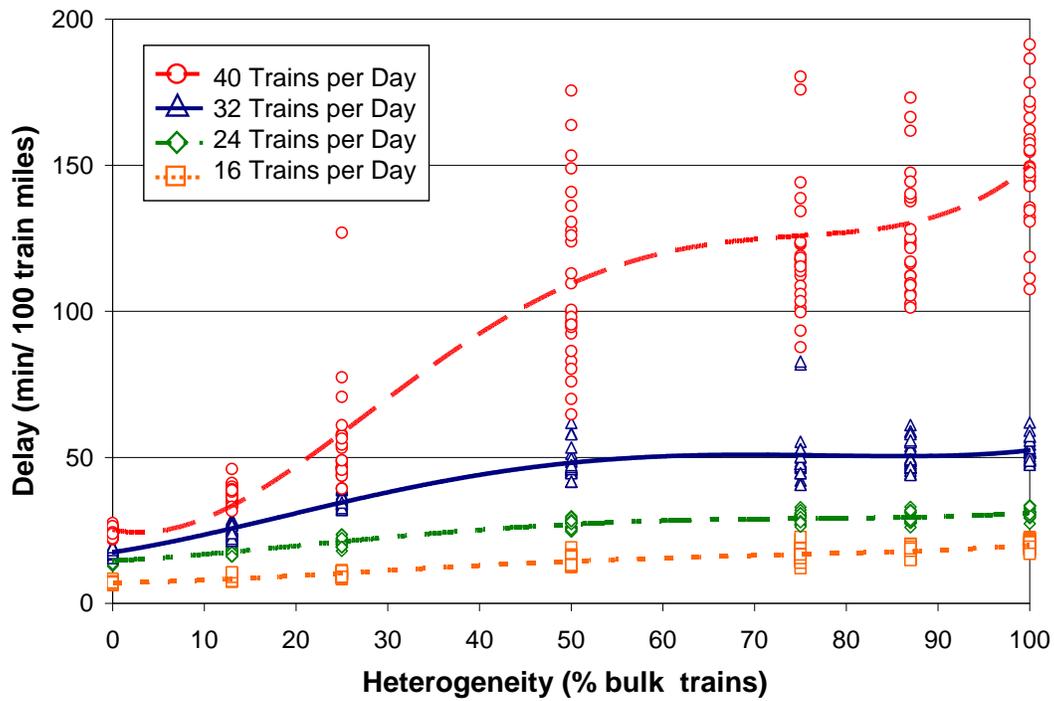
(a)
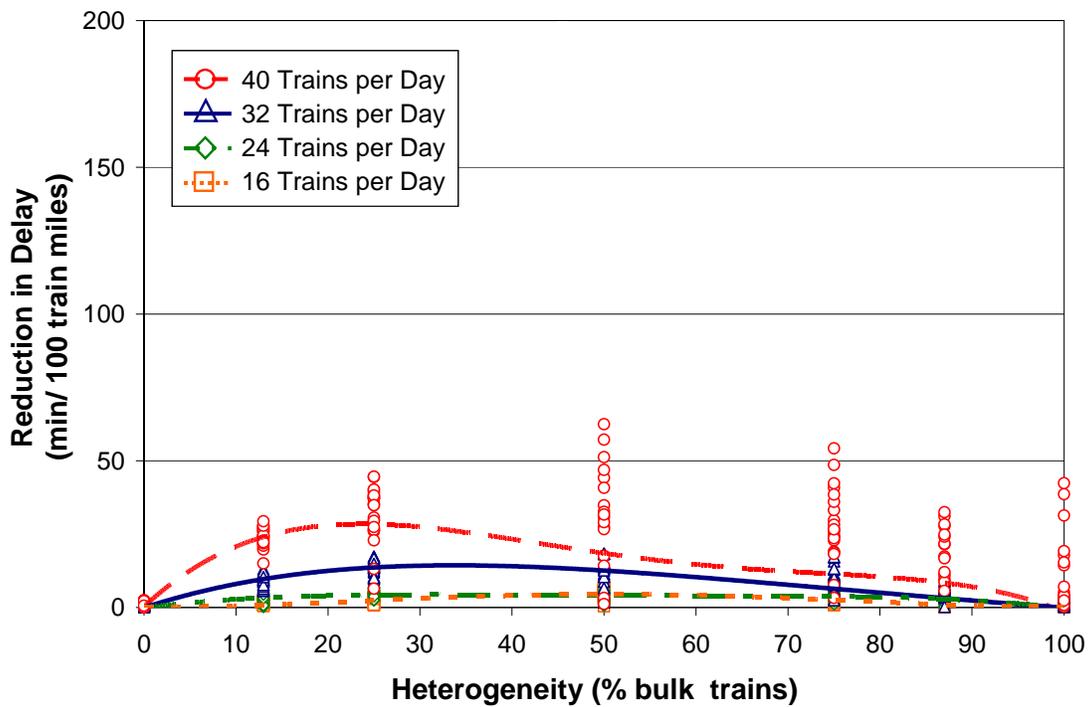
(b)

FIGURE 7.5: Delays (a) and Reduction in Delay (b) When all Trains are Given the Same Dispatching Priority

When the sidings are spaced 10 miles apart the greatest delays occur with heterogeneous traffic (Figure 7.6). The greatest reduction in delay is at the higher volumes and all the reduction in delay is greater than observed for any of the operational changes. Since the maximum capacity of the route has been increased, the excess capacity serves to minimize delays at higher volumes.

7.3.6 Adding a Second Track

The second infrastructure change considered is the construction of a second main track. Operations are vastly different with multiple tracks than with one track. With multiple tracks trains can be separated by direction, referred to as directional running, eliminating meets. It also reduces the impact of heterogeneous train speeds. If there is sufficient capacity, and a suitable traffic control system, a faster train can easily crossover to the other track to pass a slower train and then return to the original track at a later crossover. This reduces delays to both the faster and the slower trains. However, a second track requires significant construction and maintenance costs and crossovers require twice as many turnouts compared to passing sidings, further increasing capital and maintenance costs.

By adding a second track the delays with homogenous traffic are eliminated, and the delays with heterogeneous traffic are few (Figure 7.7). When all the trains are traveling at the same speed the trains can keep a constant headway between the trains, preventing delays. When the traffic is heterogeneous and the traffic volumes are low enough, the higher speed, higher priority intermodal trains pass the slower bulk trains with minimal delays to the traffic. It is only at higher volumes and traffic with heterogeneous speeds that there will be delays with multiple tracks. At higher volumes faster trains are unable to crossover due to opposing traffic on the other track.
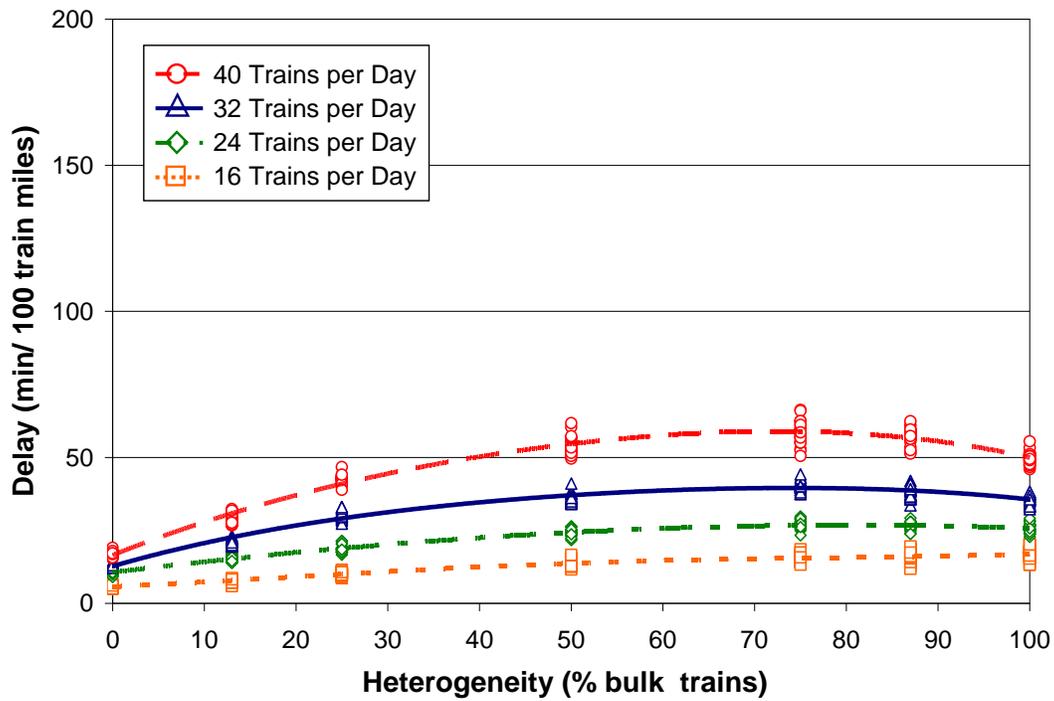
(a)
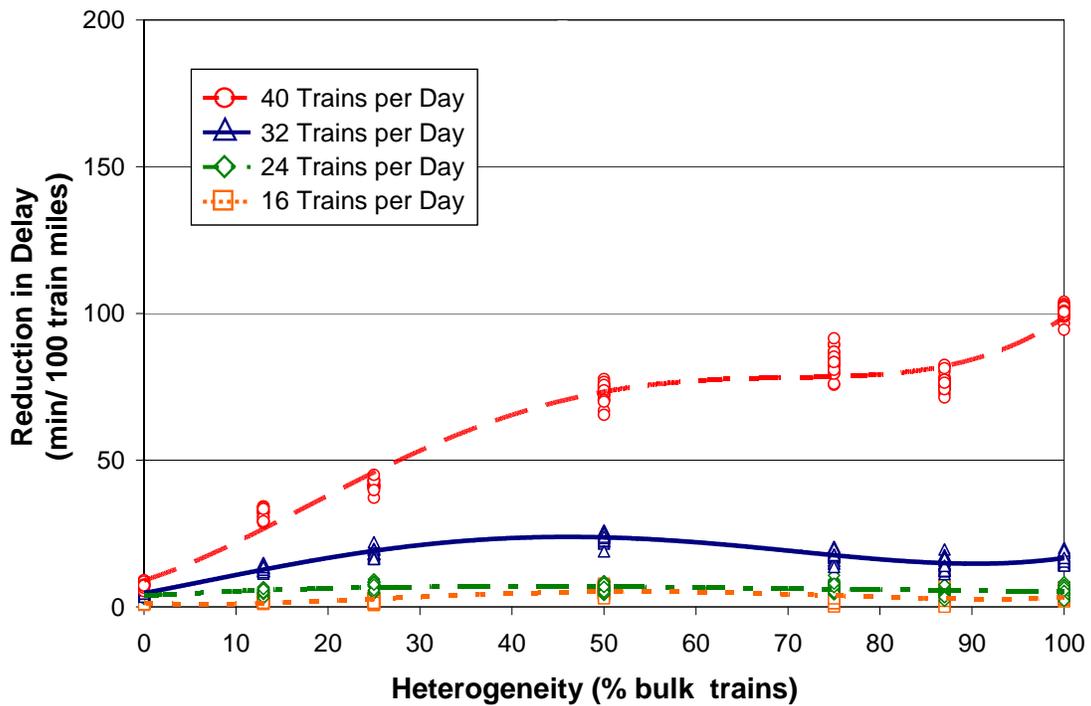
(b)

FIGURE 7.6: Delays (a) and Reduction in Delay (b) when Additional Sidings are Added Reducing the Siding Spacing from 20 miles to 10 miles
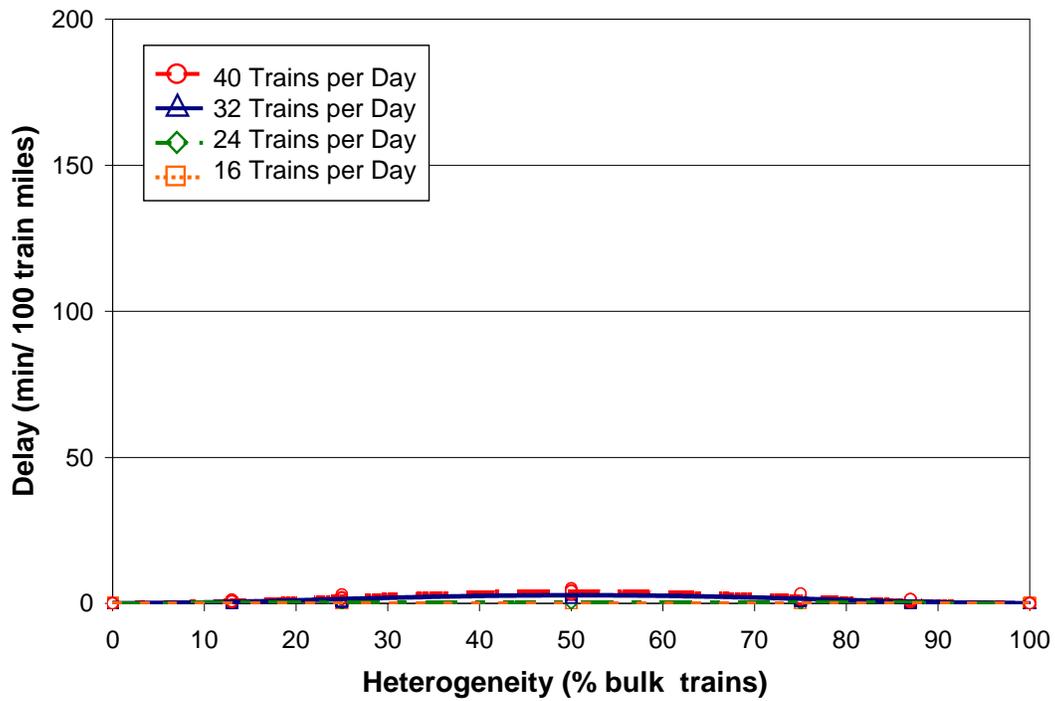
(a)
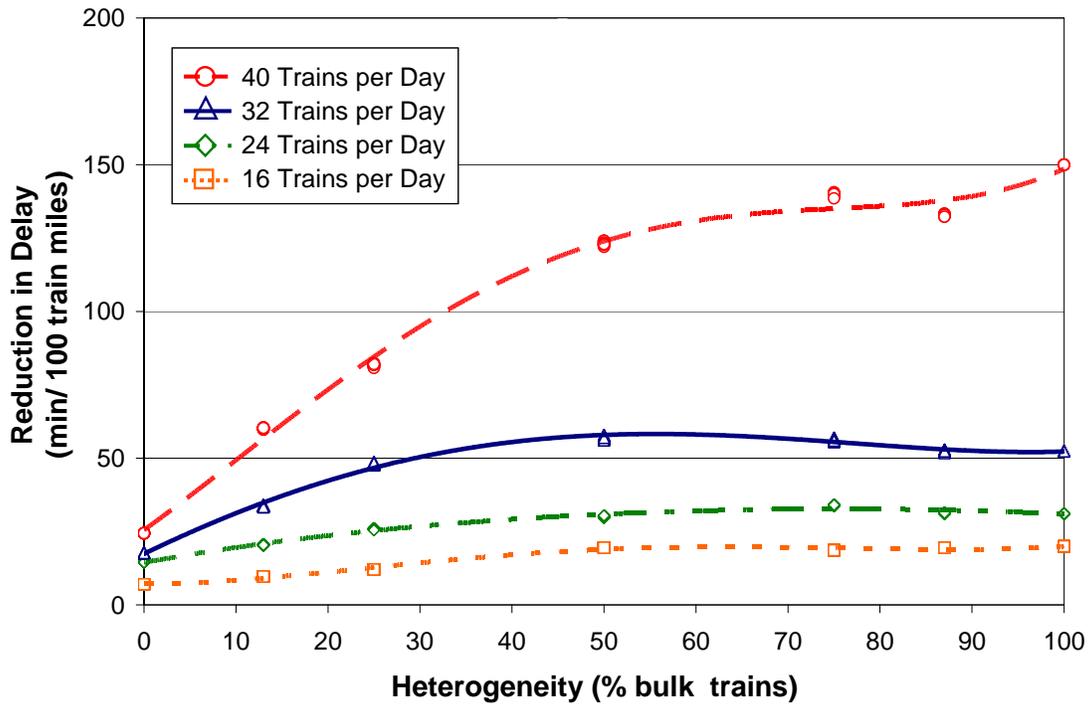
(b)

FIGURE 7.7: Delays (a) and Reduction in Delay (b) When a Second Track is Added

7.4 Economic Analysis of Operational and Infrastructure Changes

In each of the scenarios considered the reduction in delay is due to an operational change or an investment that incurs some additional expenditure. Deciding which, if any, of these approaches is most appropriate requires understanding both the costs and benefits. Consequently, I developed a framework to consider each of the scenarios I analyzed. For each scenario, the reduction in delay was considered to be the benefit of the project, while any increases in delay, additional locomotives, increased fuel consumption, increased track maintenance costs, and additional infrastructure were considered to be the expenses. In order to quantify these costs, estimates for the costs of train delay, new infrastructure, and equipment were developed.

7.4.1 Train Delay, Equipment and Infrastructure Costs

7.4.1.1 Train Delay Cost

Train delays affect the railroads through additional expenses and lost revenue. Each hour a train is delayed represents at least some degree of lost opportunity to transport cargo and increase revenue. Delays cause railcars and locomotives to be used less efficiently, and additional crew hours are also accrued. If a railroad can reduce train delay on its network, average train velocities will increase, resulting in potentially significant savings to the railroad through improved utilization of railcars, locomotives and crews. According to one railroad, an average system velocity increase of one mile per hour can free up 250 locomotives, 5,000 freight cars, and 180 train crews to move more traffic with an approximate annual savings of $200 million *(Hamberger 2006).*

  Several previous studies have quantified train delay cost. Schaffer and Barkan *(2008),* evaluating the economic impact of broken rails, estimated that the average railroad cost of train

delay was $213.52 per hour. Lai and Barkan *(2008)* created a network capacity planning model

to consider the tradeoff between higher transportation costs and infrastructure expansion. To

quantify the transportation costs they used a value of $261 per train-hour. The report of the

Railroad Safety Advisory Committee (RSAC) to the FRA on the Implementation of Positive

Train Control (PTC) Systems *(FRA 1999a)* used train delay value of $250 per train hour for

freight and $7,125 per train-hour for passenger traffic. The studies of freight train delay

considered average cost, independent of the type of cargos; however, not all trains have the same

delay cost. For example, intermodal traffic is more time sensitive than other train types and

consequently has a higher train delay cost. Train specific delay cost was used previously by

Smith et. al. *(1990)* in their work determining the benefits of more efficient meet/pass planning

expected from the Advanced Railroad Electronics System (ARES) (Table 7.2). They considered

both the lading and equipment delay costs for empty and loaded bulk trains, mixed freight, and

intermodal trains. They calculated delay values ranging from $162.98 per train-hour for mixed

freight to $266.47 per train-hour for intermodal traffic (in 1990 dollars).

TABLE 7.2: Average Delay Cost by Train Type (From Smith et. al.. 1990)

Train Type	Lading Delay Cost ($/Train-hr)	Equip. Delay Cost ($/Train-hr)	Total Delay Cost ($/Train-hr)
Loaded Bulk	10.37	172.00	182.37
Empty Bulk	0.00	172.00	172.00
Mixed Freight	35.98	127.00	162.98
Intermodal	136.47	130.00	266.47

Schaffer and Barkan *(2008)* calculated train delay using four components: (1) car/equipment cost; (2) unproductive locomotive cost; (3) idling fuel cost; and (4) crew cost. However, these factors do not include costs incurred due to delays to the lading. When a railroad's cars and locomotives are fully utilized, each delay causes the cycle time of a train to increase, and the potential revenue from additional shipments is lost. The shipment will be moved by a different railroad or mode of transportation. Fuel costs are considered separately and are not considered in this train delay calculation. For the current analysis, the train type specific delay cost has four components: (1) car cost; (2) locomotive cost; (3) crew costs; and (4) lading costs.

Car delay cost is estimated using a time-based metric for the cost of rail car ownership. A new coal hopper can cost between $72,000 and $82,000, while a new 3-well or 5-well intermodal articulated car costs between $175,000 and $275,000 *(Murray 2008)*. When leased, car costs are determined by hourly car hire rates. Car hire is dependent on the car's age and value *(ORER 2009)* and can be used to approximate the cost per car-hour of owning and operating a railcar. Car hire rates for TOFC/COFC flat cars can vary from $0.29 to $2.01 per hour and $0.07 to $1.63 per hour for open top hopper in special service *(ORER 2009)*. Since trains are often composed of cars with a wide range of ages and values, an average value of $1.00 per car-hour was used for intermodal and $0.58 per car-hour for bulk. The National Freight Capacity Study performed by Cambridge Systematics *(2007)* contains statistics for the average number of cars or units per train for automobile, bulk, general merchandise, and intermodal traffic for both eastern and western U.S. railroads. According to the study, bulk cargo moved in trains with an average of 99.2 cars, while intermodal units move in trains carrying on average 137.5 units. In 2006 77% of intermodal units were containers *(AAR 2008b)*

94

therefore assuming double stack ability for containers the average number of cars for intermodal was calculated to be 84.9. It is common that railroads will not own or lease all the cars in a train; however, railroads pay compensation for each car they use. By multiplying the cost per car hour and the average number of cars, the car delay cost per train-hour was calculated to be $84.90 for intermodal and $57.54 for bulk trains (Table 7.3).

TABLE 7.3: Car Delay Cost

	Intermodal	Bulk
Avg. Cost per Car Hour	$1.00	$0.58
Avg. Cars per Train	84.9	99.2
Car Cost per Train-Hour	$84.90	$57.54

Similar to car delay cost, locomotive delay cost is also a time-based metric, the hourly cost of owning a locomotive. This cost can be calculated from the economic life of a locomotive using the purchase cost, expected usable life, discount rate, and salvage value. The purchase cost of a new locomotive varies by type and additional features, however, Murray *(2008)* and Railway Age *(2008)* provide recent estimates. Using these values the cost of a new locomotive is assumed to be $1,750,000 with a salvage value of $200,000. The number of locomotives per train corresponds to the number used in the simulations. Using an economic life of 25 years and a discount rate of 7% the estimated locomotive cost per train-hour is $87.52 for intermodal and $52.52 for bulk (Table 7.4).

TABLE 7.4: Locomotive Delay Cost

	Intermodal	**Bulk**
Cost of New Locomotive	$1,750,000	$1,750,000
Economic Life	25	25
Discount Rate	7%	7%
Salvage Value	$200,000	$200,000
Units per Train	5	3
Locomotive Cost per Train-Hour	$87.52	$52.51

In addition to equipment expenses, when a train is delayed additional crews are required. Typical North American freight train operations use two person crews. The average straight time pay of a transportation (Train & Engine) employee is $24.68 per hour *(STB 2008)*. In addition to their base salary, employees get fringe benefits amounting to an additional 35% above their base salary *(AAR 2008a)*. Since both train types have the same size crew, the delay cost due to labor for both intermodal and bulk trains is $66.64 per train-hour (Table 7.5).

TABLE 7.5: Crew Cost

	Intermodal	**Bulk**
Crew Members per train	2	2
Average Hourly Pay	$24.68	$24.68
Fringe Benefits	35%	35%
Crew Cost per Train-Hour	$66.64	$66.64

The last component of the delay cost is for the lading. When a train is delayed the cycle time of the route increases and the train is unavailable to move additional cargo. Instead of starting a new cycle with its new revenue the train continues to transport cargo that has already been paid for. The lading cost per hour is calculated using the revenue per loaded train, cycle time, return ratio and availability rate. Based on data from several Class 1 railroad annual reports *(BNSF Railway Company 2008, CSX Transportation, Inc 2008, Norfolk Southern*

Combined Railroad Subsidiaries 2008, Union Pacific Railroad 2008), the average revenue per

intermodal unit is $933. The revenue per carload of bulk cargo was calculated by dividing the

revenue by carloads originated for each of the cargo types typically transported in bulk unit

trains *(AAR 2008a)* giving a value of $2,048 per car. Multiplying the revenue per car (or unit) by

the average number of units or cars per train gives a revenue of $128,246 per completely loaded

intermodal train and $203,182 per completely loaded bulk train.

In previous research, cycle times of various train types were calculated *(Kwon et. al.*

1995). During a cycle, bulk trains are loaded in only one direction, returning empty to complete

the cycle and reload. Alternatively, intermodal trains generally are able to reload and transport

goods from the destination back to the origin. The efficiency of this can be calculated by finding

the ratio of total miles to loaded miles. These values, known as return ratios, are 2.03 for bulk

trains and 1.14 for intermodal trains *(AAR 2008a)*. This means that an intermodal train will, on

average, be composed of 88% loaded units and 12% empty units. Lastly, it is assumed that a car

is available 75% of the time, with the remaining time the car is unavailable due to maintenance,

lack of demand, repositioning, etc. Using this method the lading cost per train-hour was

estimated to be $1,153 for intermodal and $410 for bulk. While bulk trains make more revenue

per loaded train the longer cycle times and inability to return loaded cars means that the lading

delay cost for a bulk train is only about one-third that of intermodal. (Table 7.6)

TABLE 7.6: Lading Delay Cost

	Intermodal	Bulk
Avg. Cars/Units per Train	137.5	99.2
Revenue per Car/Unit	$933	$2,048
Revenue per Loaded Train	$128,246	$203,182
Cycle Time in Days	6.15	15.27
Empty Return Ratio	1.13	2.03
Availability Rate	75%	75%
Lading Cost per Train-Hour	$1,153.38	$409.67

Summing the four components yields a total delay cost of $1,392 for intermodal and $586 for bulk (Table 7.7). Although this provides a good estimate it does not account for all the costs of delay. Shipper delay penalties, extra crew costs due to hours of service limitations, fuel costs or possible additional maintenance costs were not considered in this calculation. The lading delay is an opportunity cost, based on the assumption that if a train is delayed, the ability to move more cargo is lost. This assumes full utilization of cars and locomotives and that any excess equipment could be used to move more cargo. If equipment is not fully utilized, then lading delay cost should be reduced or omitted from delay costs.

TABLE 7.7 Total Train Delay Cost per Train-Hour

	Intermodal	Bulk
Freight Car Ownership Cost	$84.90	$57.54
Locomotive Ownership Cost	$87.52	$52.51
Crew Cost	$66.64	$66.64
Shipment Delay Cost	$1,153.38	$409.67
TOTAL TRAIN DELAY PER HOUR	**$1,392.43**	**$586.35**

7.4.1.2 Infrastructure Expansion Cost

Using information obtained from multiple sources, estimates for the cost to build a siding or a second track were calculated. The calculation includes the cost of the civil work, track, signals, design and additional fees. It is assumed that no additional right-of-way acquisition is required, although in some circumstances this may be necessary The construction cost for a new 8,750 ft

siding was estimated to be $6,500,000, and for a new main track $2,750,000 per mile.

Additional infrastructure will require increased maintenance costs as well. Zarembski et. al.

(2004) provide estimates on the maintenance costs for various tonnages and track classes. Using

this information the maintenance costs were estimated at $50,000 per mile of siding and $70,000

per mile of mainline track.

7.4.1.3 Equipment Cost

As discussed above the purchase cost of a new locomotive varies by type and extra features

included. For this analysis the additional cost of a new locomotive was assumed to be

$1,750,000 *(Murray 2008, Railway Age 2008)*. Fuel consumption calculations from RTC and a

cost of $3.13 per gallon of fuel were used *(AAR 2008a)*. Locomotive maintenance costs are not

considered due to lack of data.

7.4.2 Analysis of Net Present Value of Infrastructure and Operational Changes

The net present value (NPV) of each alternative was calculated for each volume and amount of

heterogeneity per 100 miles of track. A study period of 10 years was used with a discount rate

of 7%.

7.4.2.1 Additional Locomotive for Each Bulk Train

Adding locomotives to each bulk train was not a cost-effective solution to improve operations

(Figure 7.8). Although the reduction in delay increases with the higher number of bulk trains,

the cost of additional locomotives outweighs the benefits of the reduction in delay. This results in

a zero or negative NPV for all traffic volumes and mixes, with losses increasing as the percentage of bulk trains increases.
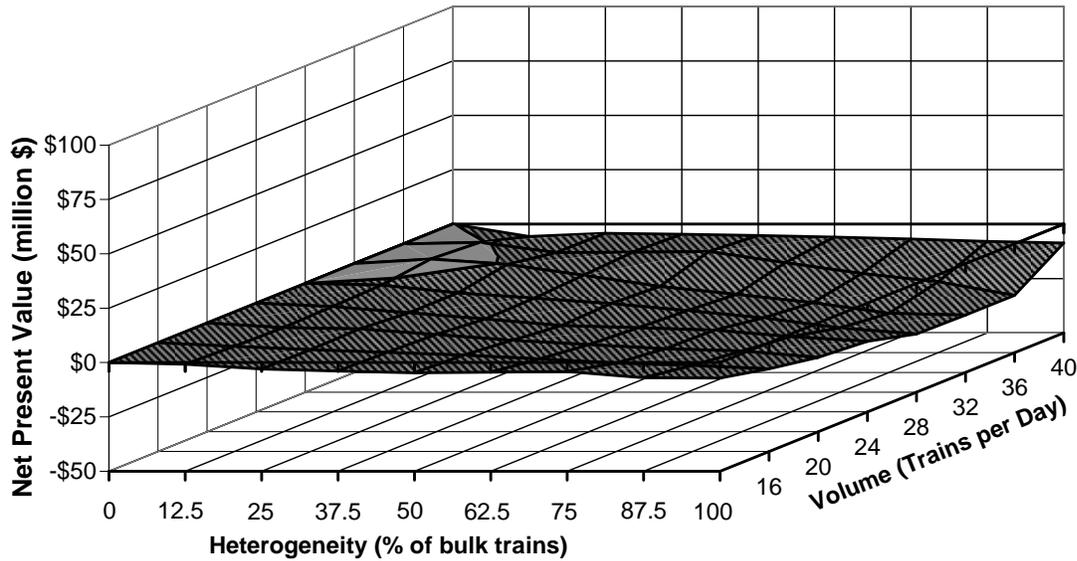
FIGURE 7.8: NPV of Additional Locomotives for Each Bulk Train

7.4.2.2 Additional Locomotive for Each Bulk Train and 10 mph Speed Increase

Adding additional locomotives in order to increase the speed of the bulk trains has a negative NPV for almost all volumes and traffic compositions (Figure 7.9). As the percentage of bulk trains increases the NPV decreases. As more bulk trains are added the costs of locomotives and fuel increase and the resultant reduction in delay does not offset these additional costs.

7.4.2.3 Equalization of Priorities

Removal of heterogeneity in priorities is cost-effective only when intermodal trains are a majority at the highest traffic volumes (Figure 7.10). When traffic is homogenous the delays are

100

unaffected and therefore there are no additional costs or benefits of removing priorities.

Removing the priorities reduces the delays to bulk trains but increases the delays to intermodal

trains. This trade-off is most beneficial when the traffic volumes are high with mostly

intermodal trains. However, if the traffic is only intermodal trains then equalizing priorities has

no effect.
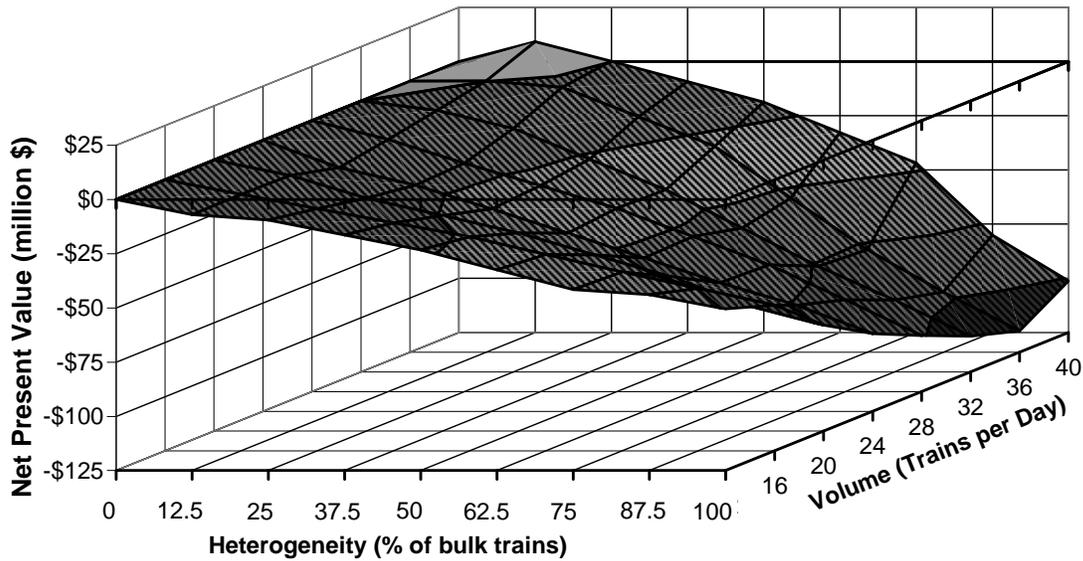
FIGURE 7.9: NPV of Additional Locomotives for Each Bulk Train and a 10 mph Speed Increase

7.4.2.4 Additional Sidings

The cost effectiveness of adding sidings is directly related to traffic volume (Figure 7.11).

Below 32 trains per day the NPV for all traffic mixes is negative but it becomes positive when

traffic volume increases above 36 trains per day. However, even at the highest volumes the NPV

is negative when the traffic is all intermodal since homogenous intermodal traffic has a much

higher capacity than heterogeneous or homogenous bulk train traffic.
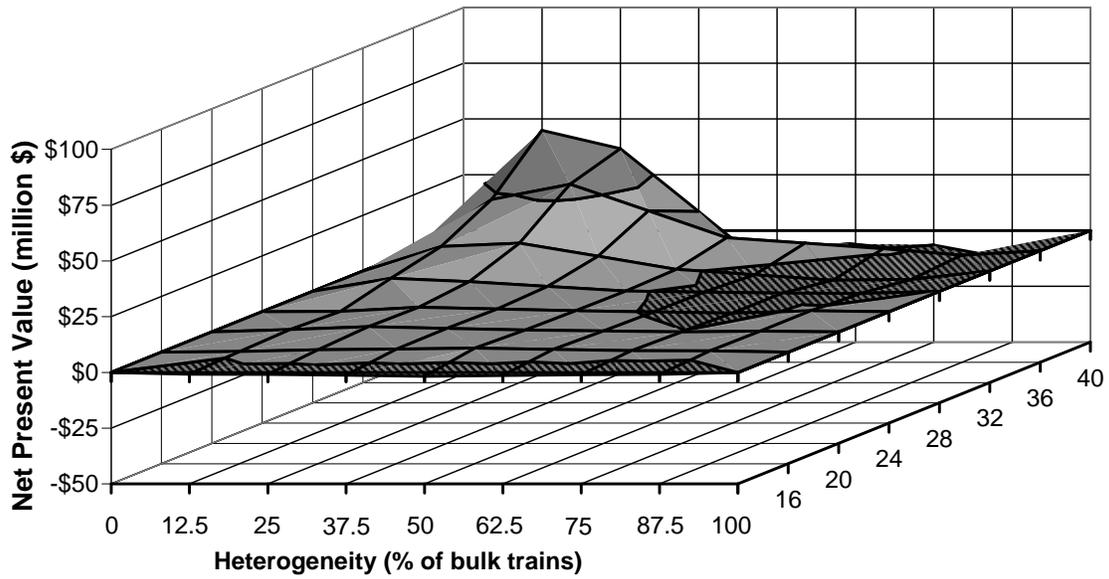
FIGURE 7.10: NPV when the Priority are Equalized for all Traffic
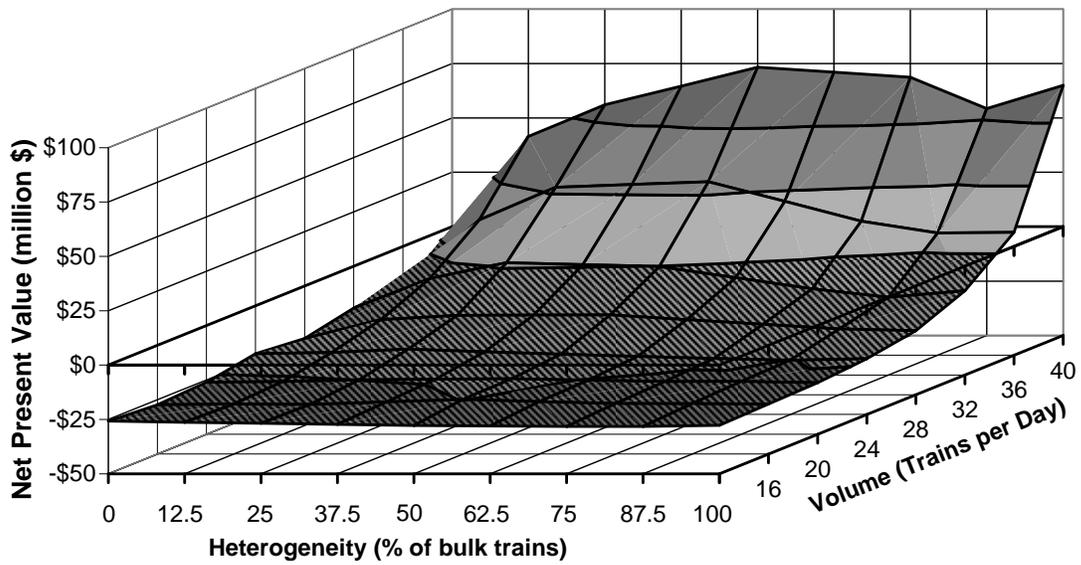
FIGURE 7.11: NPV when Sidings are Added

7.4.2.5 Adding Second Track

At the traffic levels considered, adding a second track is not a cost effective way to reduce delay (Figure 7.12). While adding a second track nearly eliminates delay, the cost to build and maintain a second track is substantial. The NPV increases with volume; however, even at the highest volumes the reduction in delay does not justify the expenditure of a second track.
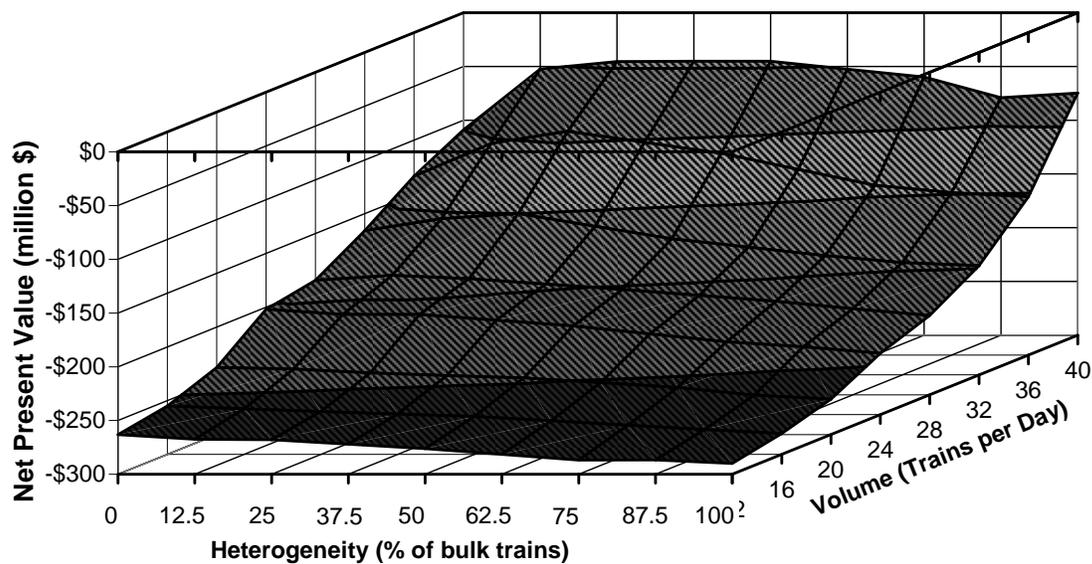
FIGURE 7.12: NPV when a Second Track is Constructed

7.4.2.6 Best Delay Reduction Strategy

Each volume and level of heterogeneity has a specific operational or infrastructure change that provides the best economic return. NPV for each scenario was considered at various volumes and traffic mixes. The alternative with the best NPV in each condition is listed in Table 7.8. This study found that for all traffic mix and volume combinations one of three alternatives was best: no change, equalizing priorities, or adding sidings (Table 7.8). When the traffic was all intermodal there was sufficient capacity and therefore no changes were cost justified. At

moderate to high volumes, equalizing priorities was beneficial with heterogeneous traffic. At the highest volumes, and when bulk was a majority of the traffic, sidings were the alternative with the best NPV.

TABLE 7.8: Best Alternative to Reduce Train Delay

		Volume							
		12	16	20	24	28	32	36	40
% of Bulk Trains	0	-	-	-	-	-	-	-	-
	12.5	-	=P	=P	=P	=P	=P	=P	=P
	25	-	=P	=P	=P	=P	=P	=P	+SD
	50	-	=P	=P	=P	=P	=P	+SD	+SD
	75	-	=P	=P	-	-	-	+SD	+SD
	87.5	-	=P	=P	=P	-	-	+SD	+SD
	100	-	-	-	-	-	-	+SD	+SD

TABLE 7.9: Alternative Nomenclature

=P	Equalizing Priorities
+SD	Adding Sidings

7.5 Discussion

With increasing demand for freight rail services, railroads must evaluate the most economical methods to reduce train delay and increase capacity. Depending on the volume and specific traffic mix the best alternative may be infrastructure expansion, operational changes or some combination.

Operational changes are advantageous because they can be implemented more rapidly, are more flexible than infrastructure changes, and may be less capital intensive. Such changes enable a railroad to respond to changing traffic levels and patterns, provide relief during short periods of high traffic volumes, or serve as an interim measure while additional infrastructure is built.

One of the most effective operational change is equalizing priorities. The only costs of equalizing priorities are the additional delays to higher valued traffic. On capacity constrained routes equalizing priorities is a rapid and flexible method that can be used to improve operations. Depending on the traffic level, a dispatcher can choose whether to utilize equal priorities as a operational strategy.

7.6 Conclusions

The projected, long-term demand for rail freight transportation and expanded rail passenger service on North American railroads will require considerable capital investment in new infrastructure. However, some additional capacity may be achieved though altering operations. This solution is often less expensive and faster to implement than building physical infrastructure. I performed analyses using dispatch simulation software to determine the benefits of various operational and infrastructure changes through the reduction of train delay. For each scenario a cost-benefit analysis was performed to determine the most cost-effective ways to improve railroad line capacity. Analysis showed that for moderate volumes and heterogeneous traffic, equalizing dispatching priorities is a cost-effective method of improving capacity. At higher volumes more cost intensive infrastructure expansion becomes a better investment option since it reduces the delay by a greater amount than operational changes.

CHAPTER 8: IMPACT OF CBTC AND ECP BRAKES ON CAPACITY

8.1 Introduction

Beginning in the early 2000s major North American railroads were increasingly experiencing capacity constraints, and long term projections indicate substantial further growth in freight traffic *(AASHTO 2007, Cambridge Systematics 2007)*. Furthermore, new initiatives to expand intercity passenger rail operations on freight railroads will have a disproportionate impact on capacity due to the differences in operational characteristics between freight and passenger trains *(AREA 1921, Mostafa 1951, Harrod 2009)*. Consequently understanding factors that affect rail capacity and the options available to cost-effectively improve it are important.

Infrastructure expansion will undoubtedly play an important role in accommodating new traffic demand; however, two new technologies are being introduced that will also affect rail capacity; communications based train control (CBTC) (often referred to as positive train control or "PTC" in the U.S.) and electronically controlled pneumatic (ECP) brakes. Both offer safety benefits and both have been touted as offering capacity benefits as well, but in actuality the situation is more complicated. These technologies can enhance capacity under some circumstances, have little or no effect under others, and in some cases may actually reduce capacity. Consequently, understanding their net effect on a particular rail line or network requires understanding the status quo of the system they are being introduced into, and in what manner they are being introduced. In this chapter I attempt to identify each critical aspect of these technologies that has the potential to affect capacity and consider what this affect will be under which implementation conditions. Since both of these systems require significant investment from the railroads (estimates range up to $10 billion for PTC *(FRA 2009a)* and over

$6.5 billion for full ECP brake implementation *(FRA 2006)* if the capacity impacts of these two technologies can be better understood, railroads can make more informed decisions about their implementation.

CBTC is a system in which train monitoring and train control are integrated into a single system via data links between vehicles, central office computers and wayside computers *(IEEE 2003)*. ECP brakes use an electronic signal instead of the train-line air pressure to transmit braking signals. CBTC has been under development since the mid-1980s *(RAC & AAR 1984, Detmold 1985, FRA 1999a)* and freight railroad ECP brake technology since the early 1990s *(FRA 1999b)*; however, wide-scale adoption has not occurred due to technical, practical, economic and institutional barriers *(Moore Ede et. al. 2009)*. Recent regulations and legislation have altered the situation. The Federal Railroad Administration (FRA) is encouraging implementation of ECP brakes by offering relief from certain requirements pertaining to conventional pneumatic brake operation *(Rail Safety Improvement Act 2008, Blank et. al. 2009)*. With regard to PTC, the Rail Safety Improvement Act of 2008 and the subsequent regulations issued by the FRA *(FRA 2010a)* have mandated its implementation on a large portion of the Class 1 railroads' mainlines by 2015.

A number of previous studies have investigated the impact of CBTC on capacity. Lee et. al. *(Lee et. al. 2000)* determined that moving blocks could increase the capacity of the Korean high speed railway. Another study quantified the capacity benefits of the European Train Control System (ETCS), Europe's version of CBTC *(Wendler 2009)*. In the United States, Smith, Resor and Patel *(Smith & Resor 1989, Smith et. al. 1990, Smith et. al. 1997, Resor et. al. 2005)* studied the potential benefits of the Burlington Northern's Advanced Railroad Electronics System (ARES) and other possible CBTC systems. They calculated how the more efficient meet/pass

planning and the increased dispatching effectiveness possible with CBTC will affect capacity. Martland and Smith *(1990)* calculated the potential terminal efficiency improvements resulting from the estimated increases in reliability offered by CBTC. While many authors have claimed that a CBTC system with moving blocks will increase capacity *(Detmold 1985, Martland & Smith 1990, Dick 2000, Moore Ede 2001, Resor et. al. 2005, Drapa et. al. 2007, Kull 2009, FRA 2009b)*, there has been some debate about whether this will in fact be the case *(Twombly 1991, Moore Ede et. al. 2009).*

There has been less work addressing the capacity effects of ECP brakes. Most agree that they will reduce stopping distances and when fully implemented this will allow closer spacing of trains; however, the incremental effect of this reduction will be affected by what other technologies are already in use. Furthermore, taking advantage of this will often require changes in the signal system.

As discussed above, the effect of CBTC and ECP brakes will be context specific, that is, in some circumstances one or both technologies have the potential to increase capacity, either alone or in combination, in other cases they will have little or no effect, and in some they may reduce capacity. Consequently, the net effect of these technologies on capacity will be determined by the magnitude of these context-specific impacts and the relative frequency that they occur over a particular route or network.

8.2 Elements of a CBTC System that will Affect Capacity

In North America most of the potential CBTC systems are still under development. While specific technical details remain unclear, in general each will have similar features and

capabilities. These systems are characterized by the data links that provide better information to dispatchers and train crews. This has the potential to increase efficiency though better train management and control *(Ditmeyer 2006)*. However, in order to comply with the legislative requirements for PTC they must also "prevent train to-train collisions, over-speed derailments, incursions into established work zone limits, and the movement of a train through a switch left in the wrong position *(FRA 2010a)*." The legislation is a performance standard and does not specify the technology that must be used to meet the requirements. In principle, CBTC can be implemented without enforcement braking; however, this has been envisioned as an element of CBTC since the earliest concepts of its development *(RAC & AAR 1984)*. It is also technically possible to meet the PTC requirements without use of a pure CBTC system *(Hoelschcer and Light 2001)*; however, most PTC systems in the U.S. will likely be some form of pure or hybrid CBTC system with enforcement. Since they are not part of the PTC regulation the additional elements available with a CBTC system will not necessarily be part of a PTC-compliant system and therefore the potential benefits or costs of PTC and CBTC are different. For this work I consider the potential elements of a CBTC system that may affect capacity including those required to meet the PTC requirements.

8.2.1 Current Traffic-Control Systems

Most current automatic traffic control systems use wayside signals to manage train speed and headway. Signal spacing is typically set based on the distance it takes for the worst-case train that normally operates on a line, using normal service braking, to stop from the maximum permitted train speed at a location. Since the signals are designed for this worst-case train, many trains may have stopping distances shorter than the line's signal system was engineered for. Furthermore, although individual railroads' rules vary on the exact language, normally an engineer is required to begin reducing speed when their train passes a signal displaying a restrictive signal. This means that in order for a train to continuously maintain normal track speed it must not encounter signals less favorable than "clear." Consequently trains must generally be separated by at least two blocks in a three-aspect system and three blocks in a four-aspect system. Due to these operating rules and use of worst-case braking distances, trains are separated by a distance several times longer than their braking distance.

There are a variety of traffic control systems currently in use on North American railroads but they can be broadly categorized into two types: those in which a manual system of spoken or written messages convey movement authority to trains, and those in which the dispatcher conveys this authority directly via the wayside signals. Lower density lines tend to use a manual system such as track warrants control, or something similar. Capacity on these can be increased by overlaying them with automatic block signals (ABS) but the authority is still conveyed manually. If more capacity is needed it can be upgraded to centralized traffic control (CTC) in which the signals themselves convey movement authority. On some track warrant and all CTC systems the dispatcher is able to remotely control switches allowing for more efficient planning and management of meets and passes of multiple trains on a line.

There are technologies that offer further improvement in operational efficiency, some of which provide more information to train crews and others that help dispatchers. The oldest of these is cab signals that takes advantage of the coded track circuits in the rails that communicate the aspect information to the wayside signals. Specialized equipment on the locomotive enables the current signal block aspect to be displayed in the cab. With wayside signals a signal ahead may change to a more favorable indication but the locomotive engineer does not know this until the next signal comes into view. Cab signals allow the engineer to know immediately if a more favorable indication applies and can immediately take advantage of it. Another technology that assists the dispatcher in managing all the traffic on a line is computer-aided dispatching (CAD). In these systems the computer accounts for the operational characteristics of trains and the features of a route to help the dispatcher better plan meets and passes.

8.2.2 Elements of a CBTC System

A PTC-compliant CBTC system has several components and features that have the potential to affect capacity, either positively or negatively. These are:

- Enforcement braking

- Real-time train operating and location data

- In-cab display

- Moving blocks

Enforcement braking is necessary in order to comply with the PTC requirements. Real-time train operating and location data gives the dispatcher additional information. This information can also be provided to the locomotive on an in-cab display. CBTC also potentially permits the

use of flexible moving blocks. Each of these components will impact railroad operations and capacity and will be considered separately.

8.2.2.1 Enforcement Braking

The element of a PTC system mandated by regulation is the enforced braking in order to prevent unsafe situations. The intent is that the system will stop the train automatically if the engineer fails to take appropriate action to prevent the train from violating its authority limits or speed restrictions. In order to provide continuous enforcement, an on-board computer must determine when a train must begin braking. This computed braking curve is composed of the distances traveled during *(Thurston 2004)*:

- Equipment reaction time

- Propulsion removal

- Brake build-up

- Full service brake application

These distances are highly dependent on factors including initial train speed, train length, car weights, braking efficiency, operative brakes, brake propagation rate, adhesion and rail condition. These factors are not accurately known when a train leaves the terminal resulting in considerable uncertainty in the exact braking distance required *(Anderson 1995, Moore Ede et. al. 2009)* (Figure 1). For safe operations a train must have close to zero probability of overshoot (FRA has targeted 0.000005, or 5 chances in a million *(Moore Ede et. al. 2009, FRA 2009)*). This necessitates a conservative braking algorithm that considers the worst case condition for each of the unknown variables. This causes the enforced braking distance to be greater than the average braking distance *(Thurston 2004)*.

112
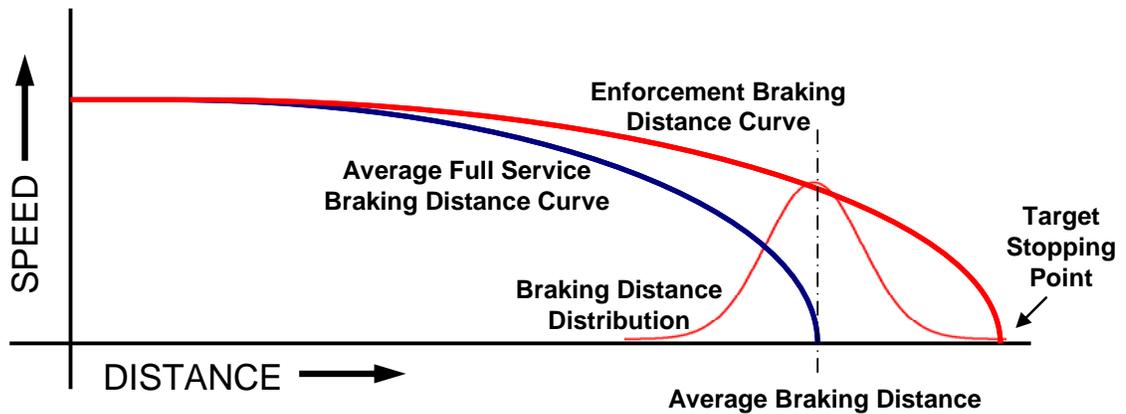
FIGURE 8.1: Speed, Uncertainty in Average Braking Distance, and Resultant Safe Braking Distance

Consequently, the brake application with a PTC system will begin earlier than required for a typical full service brake application. With or without braking enforcement a train will brake in the same distance, consequently an earlier application will cause the train to stop sooner than the engineer intends *(FRA 2009b)*. Simulations have shown that the difference between the average stopping distance and enforced target can be greater than 1,700 ft *(FRA 2009b)*. Braking enforcement can have several negative effects on capacity including:

- An unacceptably large number of trains are forced to start slowing much earlier than normal service braking to prevent enforcement from taking over, slowing the overall operation;

- Train crews are not able to prevent enforcement, thus stopping well short of the target;

- Train crews experience difficulty closely approaching a target stopping point, such as when pulling into a siding potentially causing the back of the train to remain on the main line blocking traffic *(Moore Ede et. al. 2009)*.

Work is underway to create a more accurate and adaptive braking algorithms *(Moore Ede et. al. 2009)*. However, trains may travel long distances after departing a terminal without making enough brake applications to obtain adequate data to develop sufficiently accurate, updated

estimations of braking distance *(FRA 2009b)*, and there will always be some difference between the calculated braking distance and the actual or performance braking distance *(Thurston 2004)*. The magnitude of this difference is dependent on the conservativeness of the braking algorithm used; a more conservative algorithm will increase the difference between the actual and enforcement braking distances. The probability of overshoot used is dependent on the current specifications regarding enforcement braking, consequently, the manner in which those specifications are interpreted will have a direct impact on the effect of enforcement braking on capacity.

It is also possible that enforcement may have little or no impact on operations or capacity. Current wayside, signal spacing is based on the braking distance of the worst-case train plus an additional margin of safety. Signal spacing may be greater than the enforced braking distance; therefore, if signals are still used, trains will begin to slow down in response to them instead of the enforcement. Additionally, enforcement algorithms are based on a full service brake application. In most cases the engineer makes use of dynamic brakes and slows the train at a more gradual rate than with a full service brake application potentially preventing enforcement.

Depending on the railroad's operations and rules, enforcement braking has the potential to either increase travel times for the affected train or have no impact at all. If trains are slowed they may also delay following trains, further reducing capacity. Further discussion and explanation of braking enforcement, adaptive braking and their implications can be found in papers by Thurston *(2004)* and Moore Ede et. al. *(2009)*.

8.2.2.2 Real-time Train and Location Data

Real-time train and location data offer the dispatcher additional information. The dispatcher is able to accurately know a train's location and current speed with more precision than existing train control systems provide. This information will allow train dispatchers to respond more quickly to any disruptions or changes and to more quickly formulate alternative dispatching plans as circumstances change. This information also permits more effective meet pass planning. When combined with a CAD system this can potentially decrease run times by reducing the time trains wait for meets and passes *(Smith et. al. 1990, Smith et. al. 1997)*.

Real-time train and location data are also vital to braking enforcement and moving blocks. A technical challenge that has been encountered with real-time data is communications delay in the data links. In a CBTC system a train's movement depends on receiving periodic authority updates as the track ahead clears. Any limitations in the data link throughput and message reliability could limit train capacity. If the data link delivers a movement authority too late the train may have to reduce speed. Unreliability in the system could result in train position information being inaccurate to the extent that the uncertainty buffer distances must be increased, increasing train headways *(FRA 2009b)*. If the communications delay is not excessive, real time train and location data can increase capacity.

8.2.2.3 In-Cab Display

In-cab displays offer additional information to the locomotive engineer permitting him to more efficiently operate the train. An in-cab display will most likely have the following information *(FRA 2004)*:

- Location information

- Authority and speed limits

- Route and route integrity

- Start of warning and enforcement braking

- Location of maintenance-of-way work limits

- Position of other track vehicles

An in-cab display offers the engineer near real-time information on the status of blocks ahead. With wayside signals, this information is only updated at discrete points as the train approaches and passes each block signal. If the signal is anything less favorable than clear, the engineer will need to reduce speed soon or immediately unless already traveling at the speed indicated by the signal. Although the status of the block ahead may improve after the front of the train has passed, the engineer has no way of knowing this and will continue reducing speed until the next signal comes into view and is displaying a more favorable indication. However, if the engineer has access to continuously updated information on the status of the block ahead they may not have to reduce speed as much if the block ahead clears. A CBTC in-cab display can also have benefits in territories where movement authority is given through a manual system because it eliminates the time required for the voice transmission and confirmation *(Moore Ede 2001)*. Cab signal technology provides some of the capacity benefits of a CBTC in-cab display by displaying the aspect of the next block *(Thurston 2004)*; however, most locomotives and routes in North America are not equipped with these technologies so in these cases, CBTC will provide these incremental benefits.

<u>8.2.2.4 Moving Blocks</u>

Moving blocks provide continuous train separation and have the potential for this to be based on each train's individual stopping characteristics, rather than the discrete fixed blocks characteristic of current signal systems. Moving blocks thus have the potential to reduce minimum headways. With a fixed block system trains outside of terminals or interlocking limits traveling at normal track speed are typically separated by at least two blocks, irrespective of their individual stopping characteristics. By contrast, in a moving block system trains can be separated by little more than a single block, and potentially by a distance related to each train's individual stopping distance. This effectively reduces minimum train separation from two or more blocks, as required with a fixed-block system, to a single block (or even less for some trains) of separation.

Moving blocks thus have the potential to reduce minimum headways. With a fixed block system trains outside of terminals or interlocking limits traveling at normal track speed are typically separated by at least two blocks, irrespective of their individual stopping characteristics. By contrast, in a moving block system trains can be separated by little more than a single block, and potentially, with a flexible block system, by a distance related to each train's individual stopping distance. This effectively reduces minimum train separation from two or more blocks, as required with a fixed-block system, to a single block (or even less for some trains) of separation.

Flexible moving blocks can have a significant benefit on routes with trains having similar speeds but heterogeneous stopping distances. With a fixed block system the signals are spaced for the train with the longest braking distance and therefore the headway is longer than needed for much of the traffic. Slower or lighter trains with shorter braking distances, such as passenger or commuter trains, will be able to more closely follow other train traffic. This might help

117

mitigate the disproportionate impact of certain types of heterogeneity due to mixing of passenger and freight traffic *(AREA 1921, Mostafa 1951, Harrod, 2009)*. Flexible moving blocks also offer a benefit when recovering from temporary track outages or delays. Successive trains will be able to follow each other more closely because of their shorter braking distance at slower speeds. With a single track, in order to get operations back to normal as quickly as possible, moving blocks will allow trains to be fleeted through the work area with much closer spacing than with conventional signal systems. This fleeting may also be of value when a double-track section has to be single-tracked during maintenance *(Moore Ede 2001)*.

Moving block capability can also reduce delays due to passes on single track lines. Shorter headways reduce the time the overtaken train waits in the siding *(FRA 2009b)*. Also when leaving the siding new movement authority can be issued to a train immediately after an overtaking train has passed the exit switch and the switch has been lined. It is not necessary to wait until the first block has been cleared, as may sometimes be required with conventional traffic control systems *(Moore Ede 2001)*.

8.3 Elements of an ECP Brake System that will Affect Capacity

ECP brakes change how the brake signal is transmitted. The signal will be transmitted using an electronic signal instead of a reduction in train line air pressure. Currently each car is connected with an air line that is used to charge the brakes and transmit the braking signal. With ECP brakes each car will also be connected by an electrical connection.

8.3.1 Current Systems

The current pneumatic brake system uses air pressure both to transmit the braking signal and to charge the brake reservoirs of the cars in the train. A reduction in air pressure along the brake line causes the control valve to admit air into the brake cylinder applying the brakes. Two important limitations in this system in typical North American freight train applications are that, it does not permit the reservoirs to be recharged while the brakes are being applied, nor does it permit graduated release. Repeated application and release of the brakes can deplete the air pressure in the reservoirs and substantially reduce the braking force available. Avoiding this poses several operational limitations that affect capacity, and potential safety problems if the brake system is not handled properly. The other limitation is that the air pressure signal is transmitted along the length of the train at approximately two-thirds the speed of sound *(FRA 2006)*. With longer trains there is a time lag between application and release at the rear of the train compared to the front, causing significant in-train forces. Consequently, this means there is a direct relationship between propagation time and braking distances. This problem is reduced when using distributed power (DP) because it permits the braking signal to be initiated at more locations in the train, thereby reducing brake signal propagation time and thus braking distance *(Barrington & Peltz 2009, Petlz 2009)*. Railroads are increasingly using DP; one major railroad estimates that 50% of its operations are now using distributed power.

8.3.2 Elements of an ECP Brake System

ECP brakes have several characteristics that have the potential to affect capacity. These are:

- Instantaneous transmission of the brake signal
- Steady brake line pressure

- Self-monitoring capabilities

Using an electronic signal instead of air pressure to transmit the brake signal allows for virtually instantaneous transmission enabling nearly simultaneous application or release of the brakes along the entire length of the train. ECP brakes have a steady brake pipe pressure allowing for continuous charging of the brake reservoirs even while brakes are being applied. The use of a train line cable also allows real-time, self-diagnostic 'health check' functions to be incorporated into the brake system that inform the train crew when maintenance is needed *(FRA 2006)*.

Each of these characteristics will be considered for their impact on capacity. There are several proposed elements of an ECP brake system, including tri-couplers and the ability to remotely uncouple cars, that have the potential to impact capacity. These have not been included in any of the developed systems and therefore they are not considered in this analysis.

8.3.2.1 Instantaneous Transmission of Brake Signal

With current brake systems there is a delay during the propagation of the brake signal whereas with ECP brakes this is eliminated. It is estimated that this will reduce braking distance by about 40 to 60 percent compared to conventional braking distance *(FRA 2006)*. Since headway between trains is limited by safe braking distance, if ECP brakes are installed on all trains such a reduction will permit closer train spacing if the traffic control system can accommodate it. The alternative to shorter headways is the ability to travel at higher speeds with the same signal spacing *(Carlson 1994)*. Another benefit to having all the brakes on a train apply simultaneously is that it reduces in-train forces, permitting longer trains. Fewer, longer trains free up train slots, thereby allowing additional traffic. However distributed power can provide some of the same benefits in reduced braking distances and longer train lengths but not the reduction in signal

spacing that ECP brakes provide. Consequently, in some instances, railroads are already

deriving some of the benefit that this aspect of ECP brakes offers.

8.3.2.2 Steady Brake Line Pressure

A steady brake line pressure allows for the continuous charging of the brake reservoirs. This

facilitates greater use of the braking system and reduces the time lost waiting to recharge brake

line and reservoir pressure after an application. With conventional freight train brakes, once the

engineer has selected a brake level, the braking force cannot be reduced without completely

releasing and reapplying the brakes. Trains must sometimes travel with more braking force

applied than necessary resulting in slower operations *(FRA 2006)*. Continuous charging of brake

reservoirs enables graduated release of brakes offering greater braking flexibility. This will

potentially allow a train to conform more closely to appropriate track speed limits and increase

average speeds. Another benefit is the shorter restarting time after stops. With current brake

technology, in areas of descending grades, the auxiliary reservoirs on each car of the train must

be recharged before restarting from a stop *(FRA 2006, Blank et. al. 2009)*. With ECP brakes this

is not necessary, reducing dwell time on routes with large grades.

8.3.2.3 Self-Monitoring Capabilities

An electrical signal to control the brakes has the added benefit of potentially enabling

transmission of brake condition data to the locomotive. The engineer could monitor brake

condition and be informed of any failure in any car on the train. In response to these capabilities

the FRA issued a new regulation that requires brake inspections to be performed every 3,500

miles instead of 1,000 miles as is required with conventional brakes *(Class 1A brake tests-*

1,000-mile inspection 2004). This potentially allows an ECP-brake-equipped intermodal train originating from the ports of Los Angeles-Long Beach to travel all the way to Chicago without stopping for routine brake tests. Similarly, ECP brake-equipped coal trains will be able to make quicker deliveries from western coal fields to power plants in the eastern and southern states (*FRA 2010b).* This not only decreases cycle times but may also reduce congestion at terminals where these inspections currently take place. To achieve these results reconfiguration of terminal points and the resulting expenditures may be required.

8.4 Impact of CBTC and ECP Brakes on Capacity

The potential impact of these new technologies on capacity will depend on the type of implementation of each system, traffic mix, track configuration, and the topography of the route. For CBTC there are three different possible implementations, a non-vital or vital overlay to an existing control system or as a stand-alone system *(Drapa et. al. 2007).* In a non-vital overlay, the underlying control system provides movement authority, but CBTC provides an additional, automatic backup to prevent unsafe conditions. With a vital overlay, both the underlying system and CBTC verify and convey authority. In a stand-alone system, CBTC plays the sole role in verifying, conveying, and enforcing authority *(Drapa et. al. 2007).* Non-vital and vital overlay systems will still require the use of the current signal system, while a stand-alone system will permit moving blocks. Whether or not a route has single or multiple tracks will also affect the impact of these systems. A single track route is constrained due to the need for meets and passes, whereas with a multiple-track route, headway may be a more important constraint. The topography of the route also affects train handling and consequently capacity.

8.4.1 CBTC Non-Vital Overlay System

A CBTC overlay provides enforcement per the PTC requirements in addition to the current signal and traffic control systems. This type of implementation makes use of the current signal and traffic control system and therefore closer train spacing is not possible in wayside signal territory. However, in unsignaled (aka,"dark") territory an overlay system provides a more effective means of train separation. Much like a signal system, installation of CBTC would allow closer spacing of trains thereby increasing capacity. Conversely, enforcement braking will result in trains slowing down sooner than they might otherwise, thereby reducing capacity. With or without a signal system, a CBTC overlay does not provide movement authority and therefore the current methods for this will remain in place, limiting some of the benefits of the in-cab display. In Europe the overlay version of ETCS has been found to reduce network capacity *(SRA 2005)*. In North America, the potential capacity impact will be greatest on signalized, single track lines where enforcement has a greater effect due to the more frequent stops from meets and passes.

8.4.2 CBTC Vital Overlay System

A CBTC vital overlay system will have similar capacity constraints as an overlay system due to the inability to take advantage of moving blocks. However with a vital system the signal, traffic control, and CBTC system are interconnected and authorities can be issued immediately via the in-cab display of the locomotive. Capacity under a vital overlay system will generally be the same or slightly higher compared to a non-vital system.

8.4.3 CBTC Stand-alone System

A stand-alone CBTC system permits the use of real-time train and location data, in-cab displays, moving blocks and the benefits they provide. However, the potential capacity losses of braking enforcement still apply. The greatest potential benefit will be on multiple-track routes where reduced headways offer the greatest advantage. If moving blocks are used, this is likely to more than offset any potential capacity losses due to enforcement braking with a resultant benefit in capacity.

8.4.4 Impact of ECP Brakes on Capacity

In an ECP brake system the brake signal is transmitted instantaneously, the brake reservoirs are continuously charged, and the frequency of brake inspections is reduced. ECP brakes provide the greatest benefit relative to current systems for trains on severe grades *(FRA 2006).* Grades can be bottlenecks on a railroad network and ECP brakes provide improved train handling and reduced dwell while traveling on these grades. On single-track lines capacity can be improved because less time is lost during stops and on multiple-track lines because shorter headways are possible. Shorter cycles and increased terminal capacity can be achieved as well due to a reduction in the number of intermediate brake inspections.

8.4.5 Impact of the Combination of CBTC and ECP Brakes

The combination of CBTC and ECP brakes may allow better exploitation of the benefits that each offers. It has been suggested that the data from ECP brakes will increase the accuracy of the braking algorithms thereby reducing the impact of enforcement braking. Both of these systems increase the information available, and in combination the additional train data from

ECP brakes can be transmitted to the dispatcher or other relevant groups via the CBTC data network. Effective use of this information will permit a railroad to more efficiently plan and manage its operations. A stand-alone CBTC system will take greatest advantage of ECP brakes because moving blocks will permit railroads to reduce headways that ECP brakes permit without the need to modify signal spacing. Since it will take time for all trains to be equipped with ECP brakes, a stand-alone system will permit those trains equipped with ECP brakes to follow more closely behind trains ahead, thereby providing incremental capacity benefits before the entire rail car fleet has been equipped with ECP brakes. A related benefit of CBTC with moving block is that it will offer flexibility in train spacing if the train mix changes on a line, or as further improvements in brake system effectiveness lead to shorter stopping distance and potentially closer train spacing.

8.5 Discussion

CBTC and ECP brakes make the train, signal and traffic control systems more "intelligent" *(Ditmeyer 2006)*. This allows the railroad to better plan and control train movements increasing railroad efficiency and capacity. However, braking enforcement will not increase capacity and may reduce it *(Moore Ede et. al. 2009, FRA 2009b)*. As the implementation of these technologies is considered, there remain unanswered questions on their net effect on capacity.

Although railroads are planning to implement overlay CBTC systems and are testing ECP–brake-equipped unit trains, there remain technical challenges. Conservative braking algorithms and excessive communications delays within CBTC may reduce capacity. Also moving blocks have not yet been proven to be technically feasible in the North American operating environment. CBTC may permit removal of existing signal systems; however, to date

there is no practical alternative to track circuits for detection of broken rails. If track circuit

systems cannot be eliminated it may not be possible or economically justifiable to invest in a

stand-alone CBTC system. Some authors have argued that even if it is possible, it may not be

advisable to implement a completely stand-alone system *(Baugher 2004).*

Even when a reduction in headways is possible this may not translate into additional

network capacity due to other capacity bottlenecks. Headway is just one factor influencing

capacity; other operational and infrastructure factors may continue to constrain a route. Sidings,

interlockings, yards, and junctions are fixed points in the network and reduced headways will not

improve these capacity constraints. Additionally, terminals are considered major bottlenecks in

many railroad networks *(Dirnberger 2007).* Consequently, while there may be reductions in

over-the-road time due to CBTC and ECP brakes, increases in line capacity may not improve

network capacity if the principal constraints are the terminals.

When calculating the impact of these new technologies it is necessary to understand how

their potential capacity benefits compares to what can be obtained from current systems. With

ECP brakes the comparative benefits of DP need to be considered. With CBTC the current train

control technology on a line will affect the potential benefits of the system. In areas where there

is no signal system or signals are widely spaced CBTC will likely increase capacity. However,

many of the areas that are currently facing the greatest capacity constraints are urban areas where

the signals are closely spaced. Lastly the incremental benefit of CBTC is dependent on the

implementation; in some cases there may be no benefit without a stand-alone system.

8.6 Conclusions

Implementation of CBTC and ECP brakes will have a direct effect on capacity. In this analysis I considered each critical characteristic of these technologies with respect to their capacity. All CBTC implementation types with enforcement braking have the potential for a loss of capacity; but, as CBTC systems become more fully integrated, the potential for capacity enhancement improves. ECP brakes will provide benefits in most operational scenarios due to shorter braking distances. Furthermore, CBTC may enable one of the principal benefits of ECP brakes - shorter stopping distances - to be more effectively and efficiently taken advantage of. These results will tend to be route and network specific so individual railroads will need to conduct these analyses to understand the effect on their own systems.

CHAPTER 9: FUTURE RESEARCH AND CONCLUSIONS

9.1 Future Work

In the course of this research, several topics were identified as potential areas for further research. These areas are discussed in the following sections.

9.1.1 Double Track Heterogeneity Study

Chapters 5, 6 and 7 consider operations on single track; however, many of the routes with the greatest traffic volume have two or more tracks. The characteristics of operations with multiple tracks are quite different. Multiple tracks allow directional running thereby, eliminating meets. Consequently headways are a greater capacity constraint than on single track. The different nature of multiple track operations means that the key factors contributing to lost capacity, and their relative impact due to train type heterogeneity, are also different. Future work should thoroughly investigate operational approaches to improve capacity on multiple tracks.

9.1.2 Sources of Delay

The methodology used in Chapter 6 can be expanded to better understand the impact of various operations. Additional work should be completed considering multiple volumes, no priorities, different infrastructure configurations and passenger traffic.

9.1.3 Quantitative Analysis of the Impact of CBTC and ECP brakes on Capacity

Chapter 8 is a comprehensive review of the potential effects of CBTC and ECP brakes. However, further work needs to be done to quantify the impact of each of these technologies on

capacity. Some work was done trying to quantify the impacts of the technologies using RTC but further refinements to this software are needed to adequately account for the complexities of how these two technologies will affect capacity. ECP brakes permit shorter braking distances and better handling on grades, while PTC will cause a train to brake according to an enforcement algorithm and not the experience of an engineer. Therefore, in order to accurately quantify the impacts of these technologies the braking distance of various train types with various speeds and conditions must be accurate. Additional work needs to be done developing accurate braking data for various train types and understanding how enforcement braking will influence train handling.

9.1.4 Risk of Delays with Large Traffic Volumes and Levels of Heterogeneity

At higher traffic volumes and heterogeneity levels the probability of large delays due to an unplanned event or even in the course of normal operations increases. A route may have sufficient capacity for normal operations but is unstable because of its sensitivity to disruptions. Alternatively, greater amounts of capacity reduce the risk of train delays. Risk can be used as another capacity metric and utilized to determine the cost of new traffic and the benefit of expanded capacity.

9.1.5 Impact of Passenger Trains

There are numerous proposals for expanded and higher speed passenger rail operations on North American freight railroads. This new traffic will increase the heterogeneity of a route thereby increasing the delays to the remaining traffic. An investigation into the impact of additional and higher speed passenger traffic on new and existing routes should be completed in order to better understand its disproportionate impact.

9.1.6 Impact of Non-Scheduled Delays

The work done for this thesis only considers scheduled delays. A major challenge for railroads is recovering from non-scheduled delays, which include mechanical and infrastructure failures as well as train and grade crossing accidents. Using simulation it is possible to quantify the consequence of these different delays on other traffic. This information can support further analysis of the benefit of various technologies and methods to reduce the likelihood of these events.

9.2 Conclusions

Freight railroads are increasingly facing capacity constraints *(Cambridge Systematics 2007)*. Coinciding with projected increases in freight traffic are new proposals for expanded and higher speed passenger and commuter rail operations and the development of new technologies that have the potential to affect capacity. If railroads do not prepare in advance they will have insufficient capacity and service quality will deteriorate and operating costs will increase. Railroads must understand their operations in order to effectively use existing capacity and efficiently plan new capacity.

One factor that affects railroad operations and capacity is train type heterogeneity. Research was conducted determining the impact of train type heterogeneity, quantifying the specific operations and conflicts that cause delays and possible ways to mitigate delay. In Chapter 5 simulations were performed to look at the relationship between delay, volume and heterogeneity. These showed that delay increases with greater levels of heterogeneity. Further work was done to identify the contributing factors that cause the increased delays. This provided

insight into how acceleration performance and priority together contribute to increased delays. Additional delays due to heterogeneity primarily accrue to lower priority trains, which often have the poorest operating characteristics and must make additional stops.

Chapter 6 investigates the sensitivity of various categories of delay to heterogeneity. For each conflict the delay was categorized by cause, offering insight into what types of delays are increased due to heterogeneity. The work showed that the conflict that results in the most delays are meets and that most of the delay occurs when a train is stopped in a siding.

Chapter 7 uses the data gained from the previous chapters to propose various methods to reduce train delays. On a single track line with wide spacing between sidings, various operational and infrastructure scenarios were considered. The benefit in terms of reduction in delay was compared to the cost in terms of greater delays for some train types and new infrastructure and maintenance costs.

Two new technologies that have been widely discussed for their potential benefits to railroad capacity are communications based train control (CBTC) and electronically controlled pneumatic (ECP) brakes. A comprehensive literature review of articles, papers, reports and regulations pertaining to each technology was conducted in order to identify the key elements of these technologies that will affect capacity. Using this information the potential impacts of each system and the type of locations that will have the greatest impact due to these technologies were identified.

Finally, it should be emphasized that the simulation results here represent general relationships based on idealized conditions on a hypothetical rail line. As such they are intended to provide insight on the relative importance of different factors thought to affect delay, not as absolute measures of capacity under the conditions described. Specific information about a

particular infrastructure configuration and mixture of traffic would require a detailed study using appropriate data, specific to the conditions being studied. The methods described in this paper could be adapted for such an analysis and this work provides insight regarding what type of information is needed and likely to be important in such a study.

REFERENCES

AAR (Association of American Railroads). (2008a). *Analysis of Class I Railroads*. Association of American Railroads, Washington, D.C.

AAR (Association of American Railroads). (2008b). *Railroad facts, 2008 edition.* Association of American Railroads, Washington, D.C.

AAR (Association of American Railroads). (2010). Weekly Railroad Performance Measures, www.railroadpm.org, Accessed May 21, 2010.

Abril M., Barber F., Ingolotti L., Salido M.A., Tormos P., and Lova A. (2008). An Assessment of Railway Capacity. *Transportation Research Part E*, Vol. 44, No. 5, pp. 774-806.

American Association of State Highway and Transportation Officials (AASHTO). (2007). *Transportation - Invest in Our Future: America's Freight Challenge*, AASHTO, Washington, DC.

Anderson, D. (1995). Study of the Sensitivity of Predicted Stopping Distances to Changes to Input Parameters, AAR Transportation Technology Center, Pueblo, CO.

AREA (American Railway Engineering Association). (1921). Notes on the Determination of the Traffic Capacity of Single and Multiple Track Railways. *Proceedings of the 1921 Convention of the American Railway Engineering Association*, Chicago, IL. pp. 744-759.

AREA (American Railway Engineering Association). (1931). Method of Increasing the Traffic Capacity of a Railway. *Proceedings of the 1931 Convention of the American Railway Engineering Association*, Chicago, IL. pp. 643-692.

Barrington, B., and D. Peltz. (2009). 10,000ft Distributed Power Intermodal Trains, In: *Proceedings of the 9th International Heavy Haul Conference – Heavy Haul Innovation and Development*, Shanghai, China, pp. 673-681.

Baugher, R.W. (2004). PTC: Overlay, or Stand-alone? *Railway Age*, May 2004, pp. 68.

Blank, R.W, B. McLaughlin, and S.K. Punwani. (2009). ECP Heavy Haul Coal Operations in North America with ECP Braking, In: *Proceedings of the 9th International Heavy Haul Conference – Heavy Haul Innovation and Development,* Shanghai, China, pp. 558-564.

BNSF Railway Company. (2008). Class 1 Railroad Annual Report. BNSF Railway Company, Fort Worth, TX.

Bronzini M.S. and D.B. Clarke. (1985). Estimating Rail Line Capacity and Delay by Computer Simulation, Tribune des Transports, Vol. 2, No. 1, pp. 5-11.

Cambridge Systematics. (2007). *National Rail Freight Infrastructure Capacity and Investment Study*, Cambridge Systematics, Cambridge, MA.

Carey, M. and A. Kwiecinski. (1994). Stochastic Approximation to the Effects of Headways on Knock-on Delays of Trains. *Transportation Research Part B*, Vol. 28, No. 4, pp. 251-267.

Carlson, F.G. (1994) An Experimental Electric Assist Brake System, *Technology Digest*, Association of American Railroads, TD94-021, March 1994.

Class IA brake tests -1,000-mile inspection. (2004) Title 49 Code of Federal Regulations Pt.232.207. Washington, D.C.

CSX Transportation, Inc. (2008). Class 1 Railroad Annual Report. CSX Transportation, Inc., Jacksonville, FL.

Congressional Research Service (CRS). (2007). *Rail Transportation of Coal to Power Plants: Reliability Issues*. CRS, Washington D.C.

Detmold, P.J. (1985). New Concepts in the Control of Train Movement . *Transportation Research Record: Journal of the Transportation Research Board, No. 1029*, pp. 43–47.

Dick, C.T. (2000). Impact of Positive Train Control on Railway Capacity. *Unpublished Report*, Department of Civil and Environmental Engineering, University of Illinois, Urbana, IL.

Dingler, M.H., Y-C. Lai, and C.P.L. Barkan (2009a). Impact of train type heterogeneity on single-track railway capacity. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2117*, pp. 41–49.

Dingler, M.H., Y-C. Lai, and C.P.L. Barkan. (2009b). Impact of Operational Practices on Rail Line Capacity: A Simulation Analysis. In: *Proceedings of the 2009 AREMA Annual Conference*, Chicago, IL.

Dingler, M.H., Y-C. Lai, and C.P.L. Barkan. (2009c). Impact of CBTC and ECP Brakes on Capacity. In: *Proceedings of the 2009 AREMA Annual Conference*, Chicago, IL.

Dingler, M.H., A. Koenig, S. Sogin, and C.P.L. Barkan. (2010a). Determining the Causes of Train Delay. In: *Proceedings of the 2010 AREMA Annual Conference*, Orlando, FL.

Dingler, M.H., Y-C. Lai and C.P.L. Barkan (2010b). A review of the effects of CBTC and ECP brakes on railroad capacity. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2159*, pp. 77–84.

Dirnberger, J. (2005). Development and Application of Lean Railroading to Improve Classificiation Terminal Performance, *Masters Thesis*, University of Illinois, Urbana, IL.

Dirnberger, J.R. and C.P.L. Barkan. (2007). Lean Railroading for Improving Railroad Classification Terminal Performance. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1995, pp. 52–61.

Ditmeyer, S. R. (2006). Network-Centric Railroading: Utilizing Intelligent Railroad Systems, In: *Proceedings of AREMA 2006 C&S Technical Conference*, Louisville, KY.

Drapa, J., B. Moore Ede, and A. Polivka. (2007). PTC starts to roll out across the USA, *Railway Gazette International*, May 2007. pp. 285-289.

FRA (Federal Railroad Administration). (1999a). *Implementation of Positive Train Control Systems*. FRA, U.S. Department of Transportation.

FRA (Federal Railroad Administration). (1999b). *Electronically Controlled Pneumatic Brake Revenue Service Tests*, Publication DOT/FRA/ORD-99/07. FRA, U.S. Department of Transportation. Washington, D.C.

FRA (Federal Railroad Administration). (2004). *Benefits and Costs of Positive Train Control*. FRA, U.S. Department of Transportation. Washington, D.C.

FRA (Federal Railroad Administration). (2006). *ECP Brake System for Freight Service Final Report*, FRA, U.S. Department of Transportation. Washington, D.C.

FRA (Federal Railroad Administration). (2009a). *Positive Train Control Systems Economic Analysis*, Docket NO. FRA-2006-0132, Notice No. 1 RIN 2130-AC03, FRA, U.S. Department of Transportation. Washington, D.C.

FRA (Federal Railroad Administration). (2009b). *North American Joint Positive Train Control Project*, Docket NO. DOT/FRA.ORD-09/04, FRA, U.S. Department of Transportation. Washington, D.C.

FRA (Federal Railroad Administration). (2010a). *Positive Train Control Systems* Title 49 Code of Federal Regulations Pt. 229, 234, 235 and 236. January 15, 2010. Washington, D.C.

FRA (Federal Railroad Administration). (2010b) *Electronically Controlled Pneumatic Brakes Fact Sheet*, http://www.fra.dot.gov/downloads/PubAffairs/ECP%20Brakes%20FINAL.pdf . Accessed May 15, 2010.

Galaverna, I.M., and I.G. Sciutto. (1999). Influence of Traffic Composition on the Capacity of Mixed-traffic Railway Lines. *Rail Engineering International*, Vol. 28, No. 1, pp. 14-16.

Gibson, S., G. Cooper and B. Ball. (2002). Developments in Transport Policy: the Evolution of Capacity Charges on the UK rail network. *Journal of Transport Economics and Policy*, Vol. 36, No. 2, pp. 341-354.

Goldratt, E.M. (1990). Theory of Constraints. North River Press, Great Barrington, MA.

Gorman M.F. (2009). Statistical Estimation of Railroad Congestion Delay, *Transportation Research Part E*, Vol. 45, No. 3, pp. 446-456.

Hallowell, S.F. and P.T. Harker. (1998). Predicting On-Time Performance in Scheduled Railroad Operations: Methodology and Application to Train Scheduling. *Transportation Research Part A*, Vol. 32, No. 4, pp. 279-295.

Hamberger, E.R. (2006). Testimony Before the U.S. Senate Committee on Science, Commerce, And Transportation Subcommittee on Surface Transportation and Merchant Marine *Hearing on Economics, Service, and Capacity in The Freight Railroad Industry*, 109th Congress, 2nd Session, June 21, 2006, pp. 14-15.

Harrod, S. (2009). Capacity Factors of a Mixed Speed Railway Network., *Transportation Research Part E*, Vol. 45, No. 5, pp. 830-841.

Hoelshcer, J., and L. Light. (2001). Full PTC Today with Off the Shelf Technology: Amtrak's ACSES Overlay on Expanded ATC, *Proceedings of the 2001 AREMA Conference*, Chicago, IL.

Huisman, T., and R.J. Boucherie. (2001). Running Times on Railway Sections with Heterogenous Train Traffic. *Transportation Research Part B*, Vol. 35, No. 3, pp. 271-292.

Institute of Electrical and Electronics Engineers (IEEE). (2003). *IEEE standard for user interface requirements in communications-based train control (CBTC) systems*. IEEE Std 1474.2-2003, IEEE, New York, NY.

Kraft, E. R. (1982). Jam Capacity of Single Track Rail Lines. *Proceedings of the Transportation Research Forum*, Vol. 23, No. 1, pp. 461-471.

Krueger, H. (1999). Parametric Modeling in Rail Capacity Planning. In: *Proceedings of the 1999 Winter Simulation Conference*, Phoenix, AZ. pp. 1194-1200.

Kull, R.C. (2009). ECP Braking and PTC for Increasing Heavy Haul Railway Capacity, In: *Proceedings of the 9th International Heavy Haul Conference – Heavy Haul and Innovation Development*, Shanghai, China, pp. 551-557.

Kwon, O.K., C.D. Martland, J.M. Sussman, and P. Little. (1995). Origin-to-Destination Trip Times and Reliability of Rail Freight Services in North American Railroads. *Transportation Research Record: Journal of the Transportation Research Board.* No. 1489, pp. 1-8.

Lai, Y.C. (2008). Increasing Railway Efficiency and Capacity Through Improved Operations, Control and Planning, *PhD Dissertation*, University of Illinois, Urbana, IL.

Lai, Y.C. and C.P.L. Barkan. (2009). An Enhanced Parametric Railway Capacity Evaluation Tool (RCET). *Transportation Research Record: Journal of the Transportation Research Board*, No. 2117, pp. 33-40.

Laurits R. Christensen Associates, Inc. (2008). *A Study of Competition in the U.S. Freight Railroad Industry and Analysis of Proposals that Might Enhance Competition*. Final Report, Prepared for the Surface Transportation Board (STB), Madison, WI.

Laurits R. Christensen Associates, Inc. (2009). *Supplemental Report to the U.S. Surface Transportation Board on Capacity and Infrastructure Investment*. Final Report, Prepared for the Surface Transportation Board (STB), Madison, WI.

Lee, J-D, J-H Lee, C-H Cho, P-G Jeong, K-H Kim, and Y-J Kim. (2000). Analysis of Moving and Fixed Autoblock Systems for Korean High Speed Railway, *Computer in Railways VII*, WIT Press, Southampton; Boston, pp. 842-851.

Railway Age. (2008). Locomotive Leasing: What's Power Worth Today? *Railway Age,* June 2008. pp. 40-43.

Martland, C.D. and M.E. Smith. (1990). Estimating the Impact of Advanced Dispatching Systems on Terminal Performance. *Journal of the Transportation Research Forum*, Vol. 30, No. 2, pp. 286-300.

Mattsson, L.G. (2007). Railway Capacity and Train Delay Relationships, *Critical Infrastructure: Reliability and Vulnerability,* Springer Berlin Heidelberg, pp. 129-150.

McClellan, J. (2007) "Railroad Capacity Issues," *Vol. 3: Research to Enhance Rail Network Performance*, Washington, D.C.: Transportation Research Board of the National Academies, pp. 31–37.

Moore Ede, W.J. (2001). How CBTC can Increase Capacity, *Railway Age*, April 2001, pp.49-50.

Moore Ede, W.J., A. Polivka, J. Brosseau, Y. Tse, and A. Reinschmidt. (2009). Improving Enforcement Algorithms for Communications-Based Train Control using Adaptive Methods, In: *Proceedings of the 9th International Heavy Haul Conference – Heavy Haul Innovation and Development*, Shanghai, China, pp. 722-727.

Mostafa K.K. (1951). Actual Track Capacity of a Railroad Subdivision. *Doctoral Thesis*, Department of Civil and Environmental Engineering, University of Illinois, Urbana, IL.

Murray, T. (2008). How much does it cost? *Trains Magazine*, Vol. 67, No. 1. pp. 34-43.

Norfolk Southern Combined Railroad Subsidiaries. (2008). Class 1 Railroad Annual Report Norfolk Southern Combined Railroad Subsidiaries, Norfolk, VA.

ORER (The Official Railway Equipment Register). (2009). Vol. 125, No.1, July 2009, R.E.R Publishing Corporation, East Windsor, New Jersey.

Pachl, J. (2002). Railroad Operating and Control. VTD Rail Publishing, Mountlake Terrace, WA.

Parsons Brinckerhoff Quade & Douglas, Inc. (2002). The Long Term Financial Feasibility of the Northwestern Pacific Railroad, Draft Final. www.northcoastrailroad.org/Acrobat/FinancialFeasibility/NWP%20Appendix%20I.pdf Accessed March 15, 2009.

Peltz, D. (2009). Growing Use of Very Long Distributed Power Trains, In: *Proceedings of the 101st Annual Convention and Technical Conference of the Air Brake Association, Inc*, Chicago, IL. pp. 150-162.

RAC (The Railroad Association of Canada) and the AAR (Association of American Railroads). (1984). *Advanced Train Control Systems Operating Requirements,* United States and Canada.

Rail Safety Improvement Act of 2008. (2008). H. R. 2095, 110th Cong., 2nd session, 2008.

Resor, R.R., M.E. Smith, and P.K. Patel. (2005). Positive Train Control (PTC): Calculating Benefits and Costs of a New Railroad Control Technology, *Journal of the Transportation Research Forum*, Vol. 44, No. 2, pp. 77-98.

Saunders Jr., R. (2003). Main Lines: Rebirth of the North American Railroads 1970-2002. Northeran Illinois University Press, Dekalb, IL.

Schafer D. and C.P.L. Barkan. (2008). A Prediction Model For Broken Rails and an Analysis of their Economic Impact, In: *Proceedings of the 2008 AREMA Annual Conference*, Salt Lake City, UT.

Smith, M.E. and R.R Resor. (1989). The Use of Train Simulation as a Tool to Evaluate the Benefits of the Advanced Railroad Electronics System. *Journal of the Transportation Research Forum*, Vol. 29, No. 1, pp. 163-168.

Smith, M.E., P.K. Patel, R.R Resor and S. Kondapalli. (1990). Quantification of Expected Benefits: Meet/Pass Planning and Energy Management Subsystems of the Advanced Railroad Electronics System (ARES), *Journal of the Transportation Research Forum*, Vol. 30, No. 2, pp. 301-309.

Smith, M.E., R.R Resor, and P.K. Patel. (1997). Train Dispatching Effectiveness With Respect to Advanced Train Control Systems: Quantification of the Relationship. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1584, pp. 22–30.

Strategic Rail Authority (SRA). (2005). National ERTMS Programme Team 2004/05 Progress Report, National ERTMS Programme, London.

Stok, R. (2008). Estimation of Railway Capacity Consumption Using Stochastic Differential Equations. *Doctoral Thesis*, University of Trieste, Italy.

STB (Surface Transportation Board). (2008). Wage Statistics Of Class I Railroads In The United States, Statement No. A-300.

Thurston, D.F. (2004). Signaling and Capacity Through Computer Modeling, In: *Proceedings of International Conference on Computer Modeling for Rail Operations*, Delray Beach, FL.

TRB (Transportation Research Board). (2002). Proceedings from Transportation Research Board Workshop on Railroad Capacity and Corridor Planning, Washington D.C. January 13, 2002.

Twombly, J. (1991). Migration from Conventional Signaling to Next-Generation Train Control. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1314, pp. 122–125.

UIC (International Union of Railways). (2004). Capacity. Code 406, Paris, France.

Union Pacific Railroad. (2008). Class 1 Railroad Annual Report. Union Pacific Railroad, Omaha, NE.

Vromans, M., R. Dekker, and L. Kroon. (2006). Reliability and Heterogeneity of Railway Services. European Journal of Operational Research, Vol. 172, No. 2., pp. 647–655.

Washington Group International, Inc. (2007). RTC Simulations – LOSSAN North Railroad Capacity and Performance Analysis. LOSSAN Rail Corridor Agency and IBI Group, www.sbcag.org/Meetings/SCSPC/2007/February/Item5LOSSANSR.pdf. Accessed March 15, 2009.

Weatherford, B.A., H.H. Willis and D.S. Ortiz. (2008). *The State of U.S. Railroads: A Review of Capacity and Performance Data*. RAND Corporation Technical Report. Santa Monica, CA, Arlington, VA, Pittsburgh, PA.

Wendler, E. (2009). Influence of ETCS on the capacity of lines, *Compendium on ERTMS*, Eurail Press, Hamburg, Germany, pp. 211-223.

White, T. (2006). Examination of the Use of Delay as a Standard Measurement of Railroad Capacity and Operation, In: *Proceedings o*f *the 85th Annual Meeting of the Transportation Research Board*, Washington D.C.

Wilson, E. (2010) *Rail Traffic Controller (RTC) Brochure,* Berkeley Simulation Software, Berkeley, CA.

Zarembski, A.M., R. Resor, and J. Cikota. (2004). *Technical Monograph: Estimating Maintenance Costs for Mixed High-Speed Passenger and Freight Rail Corridors*, Unpublished Report, FRA. Washington D.C.