Project ID: NTC2014-SU-R-06
STATE OF THE ART IN FREIGHT RELIABILITY ANALYSIS AND MODELING

by

George F. List
gflist@ncsu.edu
919-515-8053
North Carolina State University

Elizabeth Williams
North Carolina State University

Bo Yang
North Carolina State University

Lu Yu
Harbin Institute of Technology

for

National Transportation Center at Maryland (NTC@Maryland)
1124 Glenn Martin Hall
University of Maryland
College Park, MD 20742

June, 2015
ACKNOWLEDGEMENTS

This project was funded by the National Transportation Center @ Maryland (NTC@Maryland), one of the five National Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT)

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program and Institute for Transportation Research and Education in the interest of information exchange. The U.S. Government and Institute for Transportation Research and Education assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government and Institute for Transportation Research and Education. This report does not constitute a standard, specification, or regulation.
# TABLE OF CONTENTS

ABSTRACT ................................................................................................................................. 1

1.0 INTRODUCTION ................................................................................................................... ERROR! BOOKMARK NOT DEFINED.

2.0 HISTORY OF THE RESEARCH AREA ...................................................................................... 4

3.0 THEMATIC REVIEW ................................................................................................................. 7

3.1 REVIEWS AND SYNTHESSES ................................................................................................. 8

3.2 BASIC THOUGHTS .................................................................................................................... 9

3.3 LINK, NODE, SEGMENT, AND ROUTE RELIABILITY ............................................................. 10

3.4 SYSTEM-LEVEL RELIABILITY ASSESSMENT ...................................................................... 18

3.5 VEHICLE ROUTING AND SCHEDULING PROBLEM (VRP) .................................................... 19

3.6 CLASSICAL ROUTING AND SCHEDULING .......................................................................... 19

3.7 ADVANCED ROUTING AND SCHEDULING ANALYSIS PROCEDURES ............................... 23

3.8 SYSTEM AND NETWORK RELIABILITY .................................................................................. 23

3.9 VALUE OF RELIABILITY ........................................................................................................ 24

3.10 MODE/PATH/CARRIER CHOICE ......................................................................................... 25

3.11 LOCATION DECISIONS ........................................................................................................ 28

3.12 LOGISTICS AND SUPPLY CHAIN ...................................................................................... 29

3.13 PUBLIC SECTOR PLANNING ................................................................................................ 32

4.0 SUMMARY AND CONCLUSIONS .......................................................................................... 34

REFERENCES

LIST OF FIGURES

Figure 1: Concepts of desired and actual times of arrival .......................................................... 9

Figure 2: Disutility function to characterize desired and actual times of arrival ...................... 10

Figure 3: A hypothetical network and possible monitoring locations ...................................... 11

Figure 4: A 3-segment route ....................................................................................................... 12

Figure 5: Segment and route cumulative distributions (CDFs) for two 3-segment routes .......... 13

Figure 6: Probability of making a connection (PMAKE) for rail cars passing through railroad yard ....................................................................................................................................................... 14
ABSTRACT

This state-of-the-art review paper focuses on freight reliability. It presents a description of prior work based on a taxonomy which helps to categorize prior work and identify where there will be opportunities in the future. Extensive research has been conducted insofar as vehicle routing is concerned. Also significantly examined is the assessment of the value of travel time reliability using empirical data. Less explored are the areas of how public agencies should make decisions insofar as freight is concerned and the use of optimization techniques in conjunction with simulation to find the best solutions to freight reliability problems.
1.0 INTRODUCTION

Reliability is very important in logistics activities. Just-in-time deliveries are a premiere example. More generally, early or late arrivals can create problems for shippers, carriers, and receivers alike. The absence of reliability makes inventory costs variable. Just-in-time manufacturing schedules become hard to assure. Asset investments have to be made to buffer the process and ensure that delivery schedules are met. Moreover, MAP-21 places an onus on states to report information about the reliability of their transportation systems and describe the steps being taken to improve it.

The earliest research efforts in freight reliability focused on the impacts of stochasticity on routing, logistics management, and travel time reliability within specific modes. More recent efforts aim to find solutions for multiple-vehicle routing problems, multi-modal logistics networks, and optimal mode choice and path selection. The literature is extensive, but there are few state-of-the-art reviews. More than five hundred articles were found in preparing this paper.

A useful categorization of the literature comprises eight areas of research:

• **Link, node, segment and route reliability.** Examination of the variation in travel times within a specific mode or among a collection of nodes for network routes and segments.

• **System-level reliability.** Studies of reliability in terms of system-level assessments. An example is a trucking company where multiple terminal handlings are involved, plus pick-up and delivery from the shipper to the receiver. The absence of reliability creates missed connections and shipment delays; and these impacts can be mitigated, albeit at a cost, by building slack into the operating schedule.

• **Vehicle routing and scheduling.** Examination of decisions about how to assign loads to vehicles, how many vehicles to employ, how to route and schedule them, and how to make all of these decisions effective in light of the reliability of the network.

• **Facility location.** Decisions about where to locate fixed assets such as warehouses, transportation terminals and manufacturing plants.

• **Mode/path/carrier choice.** Examinations of travel time reliability for routes that involve two or more modes, such as truck-rail-truck, where the mode-specific segments of the trip have an impact on overall reliability.

• **Supply chain logistics.** Decisions associated with mode and path choice for single and multiple shipments, and the impacts of unreliability on economic order quantities, ordering intervals, inventory levels, and other aspects of supply chain management.

• **Public sector planning.** Investigations of the role public agencies should play in facilitating improved freight reliability. A simple example is network capacity investment to alleviate bottlenecks and congestion-caused delays. Another is pricing strategies that help ensure reliable travel times for those who choose to pay. A third is multi-modal investments to facilitate coordination between modes.

• **Reviews and syntheses.** Papers that review the literature related to freight reliability from a variety of perspectives.

The remainder of the paper is organized as follows. Section 2 describes the chronology of the research that has been performed to date. Section 3 reflects on these efforts and identifies
common threads and unanswered questions. Section 4 provides a summary of the existing literature and points to opportunities for future research.

2.0 HISTORY OF THE RESEARCH AREA

Research on freight reliability as we know it today traces its roots back to the advent of the digital age: specifically, the availability of general purpose digital computers and barcode readers. This is because technological innovation forever changed the way in which this research could be conducted. The advent of general purpose computers made it possible to study reliability using numerical methods and simulation. The introduction of barcodes made it possible to track vehicles and see how a system was performing.

Insofar as simulation is concerned, Miller and Little (1967) might have been the first to use it to examine the performance of a network. They compared signal timing strategies when subjected to stochastic vehicle arrival patterns. Bechofer et al. (1969) conducted similar analyses for signalized networks. Burt and Garman (1971) used Monte Carlo simulation to study the stochastic performance of small highway networks. Wilson (1971) used simulation to examine the reliability of transit networks. Van Vliet (1976) examined route choice for congested networks where the segment travel times are stochastic. While these efforts were not directly related to freight reliability research, the dates establish the timeframe in which such research became possible.

More related to reliability, in the 1970’s railroads began using simulation to study their operations. Railroads had large computer systems to manage their finances and operations and they took advantage of these systems to study network performance. The Southern Railway created such a system as described by Sauder (1976). Input data were automatically obtained from the field, such as car and locomotive movements and locations. Operating plan guidance was automatically generated by the simulation model. The operation of a network comprised of 125 terminals and up to 16 traffic destinations per terminal could be modeled. Individual cars could be traced. The Canadian National (1976) created a single track dispatching simulation model designed to study the changes in over-the-road performance of trains in response to operating policies and track and signal configurations. The program could handle stretches of railroad 1000 miles long and time spans up to 10 days. A train performance calculator estimated the running time of trains from point to point and dispatching logic determined which trains would operate when. Outputs included delay summaries by train, class of train and direction as well as by track location.

Not only did the railroads become involved in simulation, but they also began conducting reliability assessments based on AVI-like data for freight cars and locomotives. In the late 1960’s, railroads, nationwide, invested in a technology called KarTrak that made it possible for them to trace every car moved in interchange service. Cars could be tracked from shipper to consignee. Per Diem charges could be assessed. Cars could be checked to ensure that they were correctly routed through classification yards. Lang and Reid (1970) conducted a study of railroad car movement reliability based on data for 1,065 trains operating over a single main-line division during a two-month period. Delays were classified by type (brake, coupler, and other types of failures—not including engine failures) and possible causal factors (train length, trailing tonnage,
and track profile). The objective was to determine what caused delays and what actions could be taken to eliminate them. This line of inquiry led to the freight car utilization project conducted by MIT, reported first by Lang (1970).

Driven by national concerns about the economic health of the railroad industry, FRA decided to sponsor creation of a freight car scheduling system (see Shamberger, 1975). The system was first put in service on the Missouri Pacific (see Sines, 1972). Sierleja, Pipas, and List (1981) examined the benefits that the system had produced. List and Bongaardt (1981) estimated the benefits that such a system would produce for Conrail. And List, Buchan, Bongaardt, and Pipas (1981) assessed the benefits that would arise for railroads in New England.

The St. Louis – San Francisco Railroad Company created a similar tracking program for individual cars. It produced daily reports of elapsed time by pairs of activities (e.g., arrival to industry placement), including up to 48 stations or terminals (see Clinkenbeard, 1976). At each location, activity pairs were grouped by cars outbound, cars inbound, and cars placed at industries. Activities of trailers and containers were identically reported. Terminal performance was assessed on a tri-monthly basis by examining average elapsed hours for cars in the yard.

These explorations of reliability impacts arose in other freight arenas as well. Whybark (1974) developed a methodology for jointly determining the reorder points, order quantities and transportation alternatives which provide minimum total transportation and inventory costs for a receiving facility given the impacts of variations in travel times. An effective heuristic procedure was developed and evaluated over a broad range of conditions. Bevilacqua (1978) examined the relationship between energy conservation and different modes of freight for delivery services. Alexander (1978) reviewed actions taken by ports to better coordinate rail and steamship operations given the such aspects of the rail travel as length of haul, speed and reliability of the journey, terrain, gradients and radius of line curvature.

Also related to reliability, the Plotkin (1969) examined technologies that could be used for tracking vehicles. The main emphasis was on schedule adherence for transit buses. The Chicago Transit Authority was used as the testbed. That same year, MIT explored the development of a computer-based program for routing and scheduling dial-a-ride services. Howson and Heathington (1970) presented similar ideas for demand-responsive transportation systems. Simulation was used to perform the assessment.

In a separate domain, Cassidy and Bennett (1972) developed a computer program that was capable of scheduling multiple vehicles that had vehicle size restrictions and pick-up and delivery windows. Agin and Cullen (1975) described a computer-based algorithm for routing vehicles in response to service requirements that involve multiple stops, commodities, modes of transportation, and time periods. Rubin (1975) created a program for routing rapid transit cars from points of release to points of demand. Nussbaum, Rebibo, and Wilhelm (1975) developed RUCUS, a software system that became very popular for the routing and scheduling of transit buses. Fielding (1977) described a computer system capable of scheduling and routing shared-ride taxi services. Bodin (1978) presented computer-assisted methods for routing and scheduling street sweepers. These investigations culminated in a landmark document prepared by Bodin, Golden, Assad, and Ball (1981). It presented a state-of-the-art assessment of the methods available for routing and scheduling vehicles and crews. More than 500 citations were included.
The report not only described the methods and their underlying characteristics. It also discussed implementation issues and the prospects for using routing and scheduling models within real-world settings. Possible future areas for research were presented. An additional review of similar microcomputer applications in the transit industry was prepared by Reilly and D'Ignazio (1983). After this flurry of activity in the 1970’s, the research on reliability died down. Vehicle routing and scheduling became a major focus. Domains beyond transit became included in the research. Sexton and Bodin (1985) examined strategies for optimizing single vehicle many-to-many operations where there were desired delivery times. Levy and Bodin (1998) examined the challenge of scheduling postal carriers (people on foot) for the United States Postal Service. Pape (1988) examined the distribution of automobiles by truck. Assad (1988) prepared a review of modeling and implementation issues involved in solving general vehicle routing problems. Ballou et al. (1990) conducted a performance comparison of various algorithms for this domain. Kikuchi (1987) examined routing and scheduling schemes for specialized transportation vehicles used to meet the needs of the elderly and handicapped. With an eye toward reducing transit travel times and improving their reliability, Abkowitz et al. (1987) conducted an assessment of hub-and-spoke based, timed transfer systems. Powell (1988) provided a comparative review of algorithms for addressing dynamic vehicle allocation problems.

The emphasis on vehicle routing and scheduling continued into the 1990’s. The problems being addressed became more complicated. Langevin and Soumis (1989) studied the design of multiple vehicle delivery tours where time constraints needed to be satisfied. Teodorovic, Kikuchi, and Hohlacov (1991) considered routing and scheduling methods in light of the differences between the objectives of users and operators. An explicit address of service quality was addressed in Gendreau, Laporte, and Solomon (1991) where routing and scheduling issues were addressed in the context of missed deadlines for deliveries. List, Mirchandani, and Turnquist (1990) examined scheduling and routing issues in the context of transporting hazardous materials. Koskosidis, Powell, and Solomon (1992) studied optimization-based heuristics for vehicle routing and scheduling with soft time window constraints. Advanced search techniques became a focus of research activity, as illustrated by Garcia (1993) in the context of Tabu. Backhauls in conjunction with time windows was the focus of Duhamel et al. (1994). A comprehensive bibliography search related to routing problems was prepared by Laporte and Osman (1995). Stochastic vehicle routing was the focus of the research by Bertsimas et al. (1995). These efforts expanded into the domain of city logistics as illustrated by Kokubugata, Itoyama, and Kawashima (1997). And probabilistic vehicle routing and scheduling in the context of variable travel times was the focus of an effort by Taniguchi, Yamada, and Tamagawa (1999).

The turn of the century saw a significant expansion of freight reliability-related research. AVI (automatic vehicle identification) and AVL (automatic vehicle location) technologies began to be deployed and real-time routing and scheduling under stochastic conditions became possible. Moreover, the notion of reliability began to surface as a major consideration for freight logistics. Kwon, Martland, Sussman and Little (1995) examined reliability in the context of samples of railroad freight car movements collected by the Association of American Railroad’s Car Cycle Analysis System. Clear differences were found in reliability between general merchandise, unit train, and double-stack container services. Fu and Rilett (2000) explored the estimation of time-dependent, stochastic route travel times by using artificial neural networks. Ichoua, Gendreau, and Potvin (2000) examined dynamic vehicle routing and scheduling options in exploitation of


In 2001, the first international symposium on transportation network reliability (INSTR) was held. Taniguchi and Yamada (2003) presented a paper focused on probabilistic vehicle routing and scheduling for urban pickup/delivery trucks that made use of dynamic traffic simulation to assess the results. Mitrovic-Minic and Laporte (2004) examined waiting strategies to compensate for variation in travel times for dynamic pickup and delivery problems that involve time windows. Taniguchi and Okamoto (2005) studied dynamic vehicle routing and scheduling strategies that capitalize on traffic information to maximize the likelihood of making on-time deliveries. Taniguchi and Ando (2008) examined vehicle routing and scheduling with time window constraints from the perspective of reliability as a performance objective.

The most recent ten years have continued the focus on reliability in the context of system performance under stochastic operating conditions, multi-modal network operations, and consideration of the monetary value of reliability insofar as logistics decision making is concerned. Bone, Wallis, O’Fallon, and Nicholson (2013) provide an overview of both literature and practice insofar as freight reliability is concerned. There has been an intense focus on studying travel time reliability, especially for the highway environment as exemplified by List et al. (2014), Zhao et al. (2013), Kittelson (2013) and Figliozzi (2014). Other modes have been studied as well, including water (Cambridge Systematics, 2013), air (Anderson, 2014 and Hsu et al., 2013), and rail (Zheng and Hensher, 2012). The value of reliability has been studied intensely so that benefit/cost assessments can capture the benefits of improved reliability. Examples include McLeod (2012) and Gong et al. (2012). The tradeoffs between reliability and environmental impacts have also been studied (Li et al., 2013)

3.0 THEMATIC REVIEW

While it is interesting to see how the field has developed, a thematically-based review can also provide insights. As described in the introduction, this review employs eight categories to assist with this review:

* Link, node, segment and route reliability.
* Vehicle routing and scheduling.
• System-level reliability assessment.
• Value of reliability.
• Mode/path/carrier choice.
• Location decisions.
• Logistics.
• Public sector planning.
• Reviews and syntheses.

The research work in each of these areas is reviewed in the subsections that follow. Of course, the reviews and syntheses are not thematically specific, but broad brush in their examination of topical areas. But they are a good place to start in conducting this thematic review.

3.1 Reviews and Syntheses

A few reviews of the literature in freight reliability have been published. The first major one appears to have been Bodin, Golden, Assad, and Ball (1981) who provided a thorough review of papers and/or books focused the routing and scheduling of vehicles and crews. While the spectrum of articles they review is broad, many of the documents focus on reliability and the analytical actions that can be taken to address it. More than 500 citations are included. The report not only describes the methods and their underlying characteristics. It also discusses implementation issues and the prospects for using routing and scheduling models within real-world settings. Possible future areas for research were presented. In the same year, Assad (1981) prepared a review of analytical models that are employed in studying the performance of freight railroad systems. While no substantial integrative narrative is provided, nine papers and/or books are identified which focus on reliability. Two years later, Reilly and D’Ignazio (1983) wrote a similar review for the transit industry. Emphasis was placed on tools that made use of the emerging microcomputer. The next review appears to be Macharis and Bontekoning (2004) who review opportunities for the use of operations research-based techniques in freight transportation. While “reliability” as a word appears only once in the article, “delay” is mentioned numerous times, especially in the context of tools and techniques that can be used to simulate its occurrence and mitigate its impacts. Feo-Valero, Garcia-Menendez, and Garrido-Hidalgo (2011) appear to provide the next review. They focus on valuing freight transport time in demand modeling and the role played by reliability. The emphasis is on the effects of reliability on logistics decisions including mode choice. Their article reviews the ways in which reliability can be defined and categorized, the manner in which it can appear in mode choice models, values that can be attributed to reliability, and the implications this has for decision-making. The two newest reviews appear to be by Bone et al. (2013) and Coelho et al. (2014). Bone et al. (2013) provide a review of four themes related to reliability research: 1) journey time reliability for work commuting, tourism, other travel purposes and freight transport, 2) the impact of time delays and journey time unreliability for freight, and the effectiveness and efficiency of measures to reduce these impacts including operational congestion at inter-modal transfer locations, rail yards, inland and sea ports and freight hubs, 3) the value of journey time and journey time reliability for freight for application in the economic evaluation of projects including component values for vehicle time based on vehicle holding costs and the utilization of drivers and staff, and freight inventory and stock holding costs, and 4) demand elasticity values for freight, in particular demand elasticities for application in freight mode choice, and within-mode demand
elasticities against cost and other level of service factors. The report addresses a review of international and New Zealand literature and practice (the review was sponsored by the New Zealand Transport Agency), definition of terminology and development of a suitable market segmentation to form the basis of further market research, development of market research plans, and a suggested program for future research. Coelho et al. (2014) provide a review of the inventory-routing problem (IRP) which they say dates back 30 years. The IRP can be described as the combination of vehicle routing and inventory management, in which a supplier has to deliver products to a number of geographically dispersed customers, subject to side constraints. The article provides a comprehensive review of this literature, categorizing IRPs with respect to their definition of the structure of the problem and the availability of information on customer demand.

3.2  Basic Thoughts

Before beginning the detailed discussions about research focus areas, it seems useful to present a basic, conceptual description of what reliability analysis is about. Since this paper is concerned with freight reliability, it makes sense to focus on shipments (packages) and how they move. It is important to recognize that packages cannot move by themselves; they have to be handled and transported. Hence, the packages do not take actions that directly affect the travel times they experience, unlike people. Also, freight shipments are typically tracked from their origin to the destination, so the path taken is known as are the timestamps at intermediate locations (at least for the carrier, shipper, and receiver).

![Figure 1: Concepts of desired and actual times of arrival](image)

Ideally, freight reliability analyses are focused on assessing whether packages have actual times of arrival (ATA) that match their desired times of arrival (DTA), as shown in Figure 1. If the ATA is inside the DTA window, a reliable trip has occurred. Otherwise, it has not. Hence, a freight transportation system’s reliability should be based on the percentage of trips that have ATAs within their DTA windows.
If it is possible to observe all the trips, as a carrier or shipper can do, with the DTA windows being known, then reliability can be assessed in the complete manner described above. The percent of ATAs that fall within their DTA windows can be computed. This is a useful metric both for the entities making the trips as well as the organizations providing the service.

![Disutility function to characterize desired and actual times of arrival](image)

**Figure 2: Disutility function to characterize desired and actual times of arrival**

It is also useful to extend this assessment into a utility theory context, as described for example by Hansson (1994). Each trip has a disutility. In terms of reliability, that disutility is minimized if the ATA is inside the DTA window. Moreover, the disutility increases as the ATA moves away from the DTA window, either earlier or later. And the disutility of being late may be different from that associated with being early, as shown in Figure 2. This is shown by the slopes of the disutility function. The steeper the slope, the more important it is to be on-time. In the aggregate, the disutility of all freight trips relates to the “societal cost” of the unreliability of the system.

Public agencies typically do not have access to the ATA or DTA information. They have to assess the reliability of their systems (highways, airways, waterways, etc.) using consistency instead. Focusing on consistency is a good idea, but care has to be exercised in using the metric.

### 3.3 Link, Node, Segment, and Route Reliability

Link, node, and route reliability is concerned with the consistency in travel (transit) times experienced by shipments (or in some cases vehicles) as they traverse the network. Considering the hypothetical network shown in Figure 3, link reliability is focused on the travel times between the nodes, as in the directional travel times between the two boxes on link ED. Node transit times (handling times) are associated with the times between and among the boxes surrounding a node (as in D, for example). Route travel times relate to the travel times from one box to another, as from the box next to E on link ED to the box next to H on link CH. In
principle, the travel time distribution for the trip from E to H can be assembled based on the distribution of travel and transit times between E and H, although doing this is sometimes complicated because of correlation.

From a monitoring perspective, the biggest issue is where to collect travel time information, specifically timestamps for packages and vehicles, and to a lesser degree, how. These issues are discussed in List et al. (2014). Using the network shown in Figure 3 as an example, timestamps for a shipment from E to H should be collected when it is picked up, when it arrives and leaves E, D, C, and H (the boxes adjacent to the nodes), and when it is delivered.

![Figure 3: A hypothetical network and possible monitoring locations](image)

If timestamps for the shipments are not available, as would be the case for a public agency, then just travel times through the network can be observed (say from Bluetooth devices). The question is where should the vehicles/shipments be observed? Should the monitoring locations be at the nodes, at the mid-link locations (triangles) on the approaches to the nodes (the boxes)?

A mistake that is sometimes made is to collect time stamps at the nodes. It is easy to make the mistake. Transport analysts think in terms of flows between nodes. Packages arrive and depart at the nodes. Inbound flows arrive and outbound flows depart.

But the consistency analysis will be confounded by collecting timestamps at the nodes. The amount of time spent at the node depends on the handling that takes place. For example, at D, a shipment from E to C may be handled very differently than one going from E to B or E to F. If travel times on link ED include some or all of these handling times, the analysis of reliability on link ED will be completely confounded. Much of the variance in what appears to be the link travel time will actually be due to the differences in handling times at the nodes. It is the same mistake that traffic engineers make when they collect timestamps in the middle of intersections. The variations in delays for left turns, right turns, and through moves confound the analysis.

A good strategy for carriers and shippers is to collect timestamps immediately upstream and downstream of the nodes, as shown by the boxes in Figure 3. Then, for shipments from E to H, data can be collected for the pick-up time, the handling time at E, the transit from E to D, the
handling time at D, the transit from D to C, the handling time at C, the transit from C to H, the handling time at H and the delivery time. With this arrangement, it is easy to examine and understand the sources of variability.

Another option is to collect timestamps at mid-link locations. List et al. (2014) indicate this may be the best least-cost choice for highway networks. No time is spent at these locations. No handling occurs. No significant variability in the travel times is introduced. Rather, the variability occurs on the half-links upstream and downstream of the monitoring points and at the nodes. By establishing these mid-link monitoring locations, virtual links (segments) are created between the monitoring locations and travel time distributions are associated with these segments. Routes are formed from these segments. The segment travel times are likely to be similar because they reflect similar handling. Variations in the segment travel times will be due to differences in the durations of the times spent traveling and being handled, not the sequence of handling events that occur. Bluetooth readers can be placed alongside the facility at the timestamps and vehicle IDs fed back to a database.

Building on the ideas of segments and routes, assume, as in Figure 4 below, that a sequence of three segments, A, B and C forms a route. Also assume that each segment has a distribution of travel times; and that the route, overall, has a distribution of travel times. The questions are: 1) what are the distributions of travel times for the three segments 2) what is the distribution of travel times for the route as a whole; 3) how can the travel time for the route be predicted from the travel times for the segments; and 4) what can be done to improve the distribution of travel times on the segments and the route?

![Figure 4: A 3-segment route](image)

Figure 5 presents travel time distributions for two hypothetical 3-segment routes. In each case, the cumulative distribution function (CDF) is shown for each segment separately and for the route as a whole. In the case of route #1, the CDFs appear to be positively correlated and it might be possible to estimate the CDF for the route by adding together percentile-by-percentile values of the travel times (comonotonicity). Isukapati et al. (2013) demonstrate that this can be done for freeway networks under specific conditions. For route #2, however, a simple percentile-by-percentile addition is not likely to succeed. The long tail associated with segment C is not evident in the CDF for the route. It is likely that negative correlation exists between the segments.
A very early article focused on monitoring node-to-note reliability is Lang and Reid (1970). They examined road train delays due to various types of mechanical failures and derailments (delays that occur between the boxes on the links shown in Figure 3). The causes of delay were studied for more than a thousand trains operating over a single main-line division during a two-month period. The causes were classified by major type (brake, coupler, and other types of failures-not including engine failures) and by possible causal factors (train length, trailing tonnage, and track profile) operative at the time of the failure.

The article by Lang and Reid (1970) also appears to be the first produced by the Freight Car Utilization Program, an effort sponsored by the Federal Railroad Administration (FRA), and conducted by MIT, that focused on improving railroad reliability. It investigated the relationship between the reliability of freight car trip times and the quality of railroad operations. More than two dozen reports and papers resulted from this program including Folk (1972), Belovarac et al. (1972), Martland (1975), Martland et al. (1990), and Kraft (1995).

A significant contribution from the FRA/MIT effort is the “PMAKE” function (see Martland, 1982). As illustrated in Figure 6, the PMAKE function statistically describes the cumulative probability that an arriving car at a classification yard will connect to outbound trains within a specific amount of time. This is a “nodal”, cumulative transit time distribution, between boxes at a node, as illustrated in Figure 3. With service frequencies often being one train per day, high PMAKE values for low transit times are extremely important because missing a connection can introduce a delay of a day or more. Small improvements in train and yard operations can translate into significant decreases in delay, which results in better travel time reliability. To make use of the PMAKE distributions and develop route or trip-level reliability information, the PMAKE distributions have to be combined with distributions for the over-the-road, yard-to-yard travel times.

Link, node, and route travel times have been studied for other modes as well. Dai and Schonfeld (1991) studied the reliability of barge trips on a section of the Ohio River using microscopic simulation to evaluate the impacts travel time reliability. Their objective was to analyze the economic effects of waterway congestion and service reliability. Underpinning the model is a representation of the delays that accrue as barges pass through locks and interact with other river traffic. More recently, Wang (2007) examined the reliability of air cargo services in China and Johnson and Dupin (2012) studied the reliability of oceanic trips. Woo and Pettit (2010) examined reliability at ports. Zhao and Goodchild (2011) examine the same issues for a drayage network at a port. They propose a simple method to predict the 95 percent confidence interval of travel time between any OD pair. The method is validated using global positioning system (GPS) data.
The sector of transportation that has recently seen the most emphasis on the assessment of link and route reliability is the highway mode. This work commences with Lomax et al. (2001) who began monitoring highway congestion for the major US metropolitan areas. The ideas of buffer time, planning time, travel time index, and other reliability performance metrics were identified in this effort (see Lomax et al., 2003). The group has continued to publish nationwide performance assessment reports annually, as in the case of the 2012 Urban Mobility Report (Schrank et al., 2012).

Van Lint, van Zuylen, and Tu (2008) also examined the issue of how to measure and assess travel time reliability in a highway context. They reviewed a number of measures reported in literature. Their most important conclusions were twofold from comparing the various measures on a large empirical dataset. First, the measures were inconsistent. This was true even when comparing existing commonly used travel time reliability indicators. For example, the results of the misery index differ largely from the results of the buffer time index. Second, a compound measure was suggested, which the authors felt worked well. Following after Lam and Small (2001), they suggested monitoring both the difference between the 90\textsuperscript{th} and 50\textsuperscript{th} percentile as a robust indicator and the ratio of the difference between the 90\textsuperscript{th} and 50\textsuperscript{th} percentile and the difference between the 50\textsuperscript{th} and 10\textsuperscript{th} percentile as a measure of skew. They interpret this new measure as the likeliness of incurring a very bad travel time (relative to the median). This new compound measure, in contrast to classical statistical metrics for width and skew, allows a partial reconstruction cumulative distribution function which is useful from a reliability perspective.

More from a freight perspective, Jones and Sedor (2006) describe a study of freight reliability for trucking operations. They examined the reliability of freight travel times in significant US highway freight corridors. The research team used satellite data from trucks traveling on five freight-significant corridors to calculate travel rates and to derive measures of travel time and reliability.

A study of truck-related reliability was also done by Elefteriadou and Ciu (2007). Their primary objective was to create a model for estimating travel time reliability on freeway facilities. They made the interesting observation that the commonly held notion of reliability among the highway analysts is very different from the one originally articulated by Ebeling (1997). Ebeling said reliability should be “the probability that a component or system will perform a required function for a given period of time when used under stated operating conditions. It is the probability of a non-failure over time.” Highway analysts have focused instead on the idea of consistency, which has to do with the absence of variability.

Chu et al. (2010) examine various reliability measures such as the planning time index, the buffer time index, and the reliability index in the context urban freight corridors that provide access to a seaport. The on-board global positioning system (GPS) installed on heavy-duty commercial vehicles was utilized to collect travel time and speed data. Also examined is the validity of using parametric distributions such as Gamma, log-logistic, log-normal, and Weibull to fit the data. Their goodness-of-fit tests indicate that the log-logistic is the best statistical function for freight travel time during the mid-day period. In addition, their travel time prediction models can identify the relationships between travel time, speeds, and variance-related factors that affect travel time reliability such as incidents, work zones, and traffic signal breakdowns.
Czuch et al. (2011) examined the travel time reliability of freight shipments by truck. Use of the buffer index as a reliability measure found to be valuable. Bluetooth units were used to measure travel time and reliability. The study concludes that the use of Bluetooth readers in combination with a simple metric like the buffer index provides a cost-effective way for municipalities to measure travel time without major infrastructure changes.

Most recently, a series of SHRP2 projects have focused on the topic. For example, List et al. (2014) outline the building blocks needed for a system to monitor highway-related travel time reliability and techniques which can be employed to study it from a personal, freight, and transit perspective. Several use cases focused on reliability from a freight perspective. One of them identified the most reliable delivery time given the variability in travel times occurring on a specific route. Another focused on how to maximize the probability of an on-time delivery given path and departure time choices. A third determined what delivery window should be promised based on travel times by operating condition for a specific route. Other use cases examined decisions about vehicle routing, supply sourcing, and warehouse location.

### 3.4 System-Level Reliability Assessment

Links and nodes aggregate into systems. So do segments and routes. Of great interest is the reliability of these systems and actions that can be taken to improve their reliability. The systems can be uni-modal, as in the example of a rail or truck, or they can be multi-modal as in the context of air-cargo or oceanic container networks.

Detmold (1972) may be the first to examine reliability in a system-wide context. He considered how to set the standards of rail service at levels that would offer the best compromise between shipper needs and rail costs. From the railway’s perspective, he perceived that this meant creating service schedules and timetables that ensure there is sufficient slack to make up time after delays. From the shippers’ perspective, he believed it involved holding inventories that guard against stock-out situations. The monograph describes some routines Detmold developed for assessing the optimal combinations of service to offer.

Williams (1972) also asserted that poor rail service reliability was a probable cause for the then extant decline in the railroad share of intercity ton-miles and revenues. He suggested that four basic operating requirements had to be met if reliable rail service was to be provided: 1) an operating, 2) adequate resources, 3) a disciplined organization, and 4) service controls. The paper describes the value of using a real-time operations data base, a terminal management system, and innovative train scheduling to achieve these objectives.

Martland and Sussman (1972) examined the impacts on the reliability of railroad services that could be created by improved operations practices. Their paper can be viewed as ‘scoping’ railroad performance with respect to service time reliability. They analyzed railroad data in order to understand the level of service then being provided by the railroads as well as the reasons for the observed differences in the service provided to individual O-D pairs.
In 1972, Sines (1972) described the freight car scheduling (FCS) system being developed by the Missouri Pacific Railroad (MoPac). The intent of this system was to allow the railroad to understand when cars would arrive at their destination, estimate travel times, and reorganize train schedules to improve reliability. Ten years later, Sierleja et al. (1981) reviewed this system under the sponsorship of the Federal Railroad Administration; and subsequently, List et al. (1981) explored the potential benefits of its implementation on ConRail and Buchan et al. (1981) examined its value for the railroads in New England. The findings from these studies were that: 1) FCS improved freight train management but not necessarily freight car transit times, 2) it would have significant value on ConRail for managing freight car movements, and 3) its value to the New England railroads (specifically the Boston & Maine and the Maine Central) would be limited because the number of freight cars being handled was too small.

In 1974, Sussman and Martland (1974) described their case study of the Southern Railway. It verified conclusions presented in Martland (1982) about the ways in which reliability could be ensured. They also suggested strategies for a test program to improve reliability. Results of the test show that both reliability and mean trip times can be improved without increasing costs. They suggest that other railroads can develop similar programs to improve reliability.

Williamson (1977) examined the issues of car reliability and utilization once again. A main conclusion in this case is that the rail terminals create the reliability problems. Maximizing the number of trains that bypass yards helps significantly to improve reliability. Unit trains or solid origin-to-destination blocks are very valuable transportation devices. However, these service objectives are difficult to use on a widespread basis because of the dispersed traffic origination and termination points.

A recognition that rail trip time reliability and profitability was intensely coupled to operations led to a new body of literature focused on blocking plans. A blocking plan indicates how railcars are to be handled at the classification yards (nodes, sorting locations) as the cars transit the network. Turnquist and Daskin (1982) examined the use of queuing models to represent and study freight car delays. Daganzo et al. (1983) devised a mixed integer linear programming model that could identify “optimal” blocking plans for a given objective and pattern of flows. Yagar et al. (1983) devised a similar model as did Crainic et al. (1986) and Daganzo (1986). A practical solution to the problem was described by Van Dyke (1986) and that model has been employed by many railroads. Kraay et al. (1991) explored the optimal spacing of trains to avoid meet/pass delays caused by poor dispatching. Martland (1992) portrayed the role that control systems play in managing operations and reliability. Kraft (2002a, 2002b) presented models for scheduling the classification activities at rail yards and for scheduling railway operations.

Improving Shipment connections at terminals has been the focus of Chen (2010) and Chen and Schonfeld (2011). Chen (2010) examines transfer coordination in intermodal and intra-modal logistic networks. One model is developed for coordinating vehicle schedules and cargo transfers at freight terminals. A mixed integer nonlinear programming problem of a multi-mode, multi-hub, and multi-commodity network is formulated and solved using sequential quadratic programming (SQP), genetic algorithms (GA) and a hybrid GA-SQP heuristic algorithm. This is done primarily by optimizing service frequencies and slack times for system coordination, while also considering loading and unloading, storage and cargo processing operations at the transfer
terminals. A second model focuses on counteraction strategies for schedule disruptions. The dispatching control method proposed determines whether each ready outbound vehicle should be dispatched immediately or held waiting for some late incoming vehicles with connecting freight. An additional sub-model deals with the freight left over due to missed transfers.

The ability for a system to overcome adverse conditions in the system was also investigated. In 2002, Armacost, Barnhart and Ware discussed using composite variable formulations in the logistical planning for companies such as UPS. Using composite variables, the network design becomes robust and exercises greater flexibility. It was also found that the optimal solution of the network allows for the number of packages to be less than the capacity of the aircraft, allowing an opportunity for recourse, if needed. Armacost, Barnhart, Ware, and Wilson further investigated UPS in 2004. Utilizing a planning system called Volume, Location, and Aircraft Network Optimizer (VOLCANO), UPS is able to plan for anticipated changes in the schedule and allow flexibility in the schedule. VOLCANO makes use of composite variables that will leaves room on the system if an adverse system arises.

System-wide reliability has been studied for other modes as well. The annual urban mobility assessments (see for example Schrank et al., 2012), while not explicitly focused on freight, portray the impacts of congestion on travel times in major metropolitan areas. More from a freight standpoint, Washburn (2007) and Dowling et al. (2014) examined reliability in developing trucking level-of-service models. Xu et al. (2008) explored trip and network reliability by incorporating truck trip assignment into a dynamic traffic assignment model. Czuch and McDaniel (2011) propose a methodology for measuring reliability on the highway network. One of the preferred measures is the buffer index. This is the extra time that a driver must add to the average travel time to ensure an on-time arrival. Ogawa et al. (2012) study a set of managed motorways around Birmingham, United Kingdom, with an aim of finding ways to address congestion and improve journey time reliability. The tool is advanced intelligent transport system applications. They present the high-level results obtained from the traffic data analysis undertaken for the three schemes. The results demonstrate that traffic conditions within the scheme extents can be improved by the introduction of managed motorways and further identified findings which can be applied to assist the development of future schemes.

McCormack et al. (2010), and subsequently Ma et al. (2011) collected and analyzed global positioning system (GPS) data for trucks operating in the central Puget Sound region. The researchers examined truck freight performance measures that could be extracted from travel times and speeds. The utility of spot speeds and the GPS data in general was evaluated in the context of a three-week construction project on Interstate-90.

Figliozzi et al. (2011) developed a programming logic that uses GPS data to a) identify natural segments or regions in a corridor between urban centers, Interstate junctions, or rural areas and b) estimate corridor-wide reliability. While the study focused on an I-5 corridor in Oregon, the methodology is applicable on a network-wide basis. The researchers applied statistical techniques to compute vehicle travel time and reliability for freight movements within each segment. The methodology successfully identified distinct segments and characteristics of travel time reliability in freight corridors. Travel time information was used to compute cost effects of delays within rural and urban areas along the I-5 corridor.
3.5 Vehicle Routing and Scheduling Problem (VRP)

This area is concerned with how to operate a freight transportation system. Specifically, it addresses the use of vehicles to carry loads. In some instances, the loads are full truckloads, so the vehicles are assigned to carry one load and then reassigned to carry another. In others, they are to pick-up loads and bring them back to a depot or deliver loads to customers. In yet a third, they may be simultaneously involved in picking-up and dropping off loads, like a courier, in some sequence. The objectives are always to 1) minimize the number of vehicles needed to cover the demands and 2) use each vehicle as efficiently as possible. Tools and techniques in this area maximize fleet utilization, minimize operating costs, ensure on-time deliveries, and address other objectives. The body of literature is vast. When Bodin et al. (1981) did their review more than two decades ago, over 500 papers were identified. This review, which is broader in scope and almost 25 years later, will not be able to review all of this literature in detail.

The literature on VRP can be broken down into two subgroups. The first treats the travel times between locations and the servicing times at each location as being fixed or nearly so. This representation of the problem is the classical one and has been studied for a long time. Mathematical programming techniques, like combinatorial optimization, are used to find problem solutions. So are heuristics. The second subgroup assumes the travel times and servicing times are stochastic. Techniques like stochastic optimization and simulation in combination with optimization (search routines) are used to find solutions. This work is more recent, spawned by the advent of computers that can simulate the movement of large fleets of trucks in reasonable time.

3.6 Classical Routing and Scheduling

The first treatment of the topic appears to be Dantzig and Ramser (1959). They present a formulation of the truck dispatching problem that assigns loads to multiple trucks based on truck capacity. The motivation was refinery trucks delivering gasoline to filling stations. No direct treatment is given to the distances traveled by the trucks. The loads are sorted into a specific order and then assigned to trucks sequentially given the truck capacities.

A subsequent paper by Clarke and Wright (1964), which describes their savings heuristic, explicitly considers the distances traveled. They state the problem in a somewhat informal manner (by today’s standards) by indicating that tours are to be established for $K$ trucks such that 1) all loads are carried, 2) the total distance traveled by the trucks is minimized and 3) the capacities of the trucks are not exceeded.

The problem can be stated as follows. Assume there are $N$ loads to be carried to destinations and assume that each has a size given by $dem_i$. Also let $K$ vehicles be available and assume each one has a capacity of $cap_k$. Choosing to use vehicle $k$ is reflected by $z_{0k} = 1$ and the assigning load $i$ to vehicle $k$ is designated by $z_{ik}$. The sequence for visits to the destinations is captured by $x_{ijk}$ which indicates that load $i$ is to be delivered before load $j$ by vehicle $k$. If the distance between $i$ and $j$ for vehicle $k$ is given by $c_{ijks}$, then the problem is:
Minimize:
\[ \sum_{k} \sum_{i} \sum_{j} c_{ij} x_{ijk} \] (1)

Subject to:
\[ \sum_{k} \text{dem}_{i} \cdot z_{ik} \leq \text{cap}_{i} \cdot z_{0i} \quad \forall \ k \] (2)
\[ \sum_{k} z_{0k} \leq K \] (3)
\[ \sum_{k} z_{ik} = 1 \quad \forall \ i \] (4)
\[ \sum_{j} x_{jk} = z_{j} \quad \forall \ j, k \] (5)
\[ \sum_{j} x_{ik} = 1 \quad \forall \ i, k \] (6)
\[ \sum_{i \in S} \sum_{j \in S} x_{ijk} \leq |S| - 1 \quad \forall \ k \quad \text{where} \quad S \subseteq N(z_{ik}) \text{ is the set of all deliveries made by } k \] (7)

The objective function specifies that the total vehicle miles should be minimized. Equation (2) ensures that the capacity of each truck used is not exceeded. Equation (3) ensures the selected fleet size is not greater than the fleet available. Equation (4) ensures that the loads get assigned, Equations (5) and (6) establish the load assignment sequences and equation (7) ensures that the number of arcs traversed by each truck is less than or equal to the number of deliveries made.

The next major focus for VRP is the dispatching of special purpose vehicles to accommodate the needs of elderly and handicapped individuals. “Dial a Ride” is how it was described. The problem was as follows. A set of requests are made for trips to and from specific locations with specific departure and arrival times, like doctor’s appointments and shopping trips. The task was to determine how to assign these trips to the dial-a-ride vehicles, and how many vehicles to use. The solution became the pick-up and delivery schedule. Unlike delivering loads, multiple people could be on-board the vehicle at any given point in time.

Bruck (1969) was one of the first to present a formulation. He described a tool called CARS (Computer Aided Routing and Scheduling) that was intended to be a decision support system for solving the dial-a-ride problem. The work was sponsored by the Urban Mass Transit Authority. Papers on the same topic were prepared by Howson (1970), Deleuw, Cather (1971) Arthur D. Little (1971), Roos (1971), Roos and Porter (1971), and Roos and Wilson (1971).

As stated by Cordeau (2006), the problem is as follows. Paraphrased slightly, let \( n \) denote the number of users (or requests) to be served. The problem can be defined on a complete directed graph \( G = (N,A) \) where \( N = P \cup D \cup \{0, 2n + 1\} \), \( P = \{1, \ldots, n\} \) and \( D = \{n + 1, \ldots, 2n\} \). Subsets \( P \) and \( D \) contain pick-up and drop-off nodes, respectively, while nodes 0 and \( 2n + 1 \) represent the origin and destination depots. Thus, for each user \( i \) there is an origin node \( i \) and a destination node \( n + i \). Each vehicle \( k \in K \) has a capacity \( Q_k \) and the total duration of its route cannot exceed \( T_k \). With each node \( i \in N \) are associated a load \( q_i \) and a non-negative service duration \( d_i \) such that \( q_0 = q_{2n+1} = 0 \), \( q_i = -q_{n+i} \quad (i = 1, \ldots, n) \) and \( d_0 = d_{2n+1} = 0 \). A time
window \([e_i, l_i]\) is also associated with node \(i \in N\) where \(e_i\) and \(l_i\) represent the earliest and latest time, respectively, at which service may begin at node \(i\). For each arc \((i, j) \in A\) there is a routing cost \(c_{i j}\) and a travel time \(t_{i j}\). Finally, \(L\) represents the maximum ride time allowed by policy for a user. For each arc \((i, j) \in A\) and each vehicle \(k \in K\), \(x_{i j}^k = 1\) if vehicle \(k\) travels from node \(i\) to node \(j\). For each node \(i \in N\) and each vehicle \(k \in K\), let \(B_i^k\) be the time at which vehicle \(k\) begins service at node \(i\), and \(Q_i^k\) be the load (number of people) on vehicle \(k\) after visiting node \(i\). Finally, for each user \(i\), let \(L_i^k\) be the ride time of user \(i\) on vehicle \(k\). The formulation is as follows:

\[
\text{Minimize:} \quad \sum_{i \in N} \sum_{j \in N} \sum_{k \in K} c_{i j}^k x_{i j}^k \\
\text{Subject to:} \quad \sum_{j \in N} \sum_{k \in K} x_{i j}^k = 1 \quad \forall i \tag{1}
\]

\[
\sum_{j \in N} x_{i j}^k - \sum_{j \in N} x_{n+i, j}^k = 0 \quad \forall i \in i, k \tag{2}
\]

\[
\sum_{j \in N} \sum_{k \in K} x_{0 j}^k = 1 \quad \forall k \tag{3}
\]

\[
\sum_{j \in N} x_{j i}^k - \sum_{j \in N} x_{i j}^k = 0 \quad \forall i \in i, k \tag{4}
\]

\[
\sum_{j \in N} x_{i,2n+1}^k = 1 \quad \forall k \tag{5}
\]

\[
B_i^k \geq (B_i^k + d_i + t_{i j}) x_{i j}^k \quad \forall i, j, k \tag{6}
\]

\[
Q_i^k \geq (Q_i^k + q_j) x_{i j}^k \quad \forall i, j, k \tag{7}
\]

\[
L_i^k = B_i^k + (B_i^k + d_i) \quad \forall i, k \tag{8}
\]

\[
B_{2n+1}^k - B_0^k \leq T_k \quad \forall k \tag{9}
\]

\[
e_i \leq B_i^k \leq l_i \quad \forall i, k \tag{10}
\]

\[
t_{i,n+i} \leq L_i^k \leq L_i \quad \forall i, k \tag{11}
\]

\[
\max \{0, q_i\} \leq Q_i^k \leq \min \{Q_i^k, Q_i^k + q_i\} \quad \forall i, k \tag{12}
\]

\[
x_{i j}^k \in \{0,1\} \quad \forall i, j, k \tag{13}
\]

The objective function (1) minimizes the total routing cost. Constraints (2) and (3) ensure that each request is served exactly once and that the origin and destination nodes are visited by the same vehicle. Constraints (4)-(6) guarantee that the route of each vehicle \(k\) starts at the origin depot and ends at the destination depot. Consistence of the time and load variables is ensured by constraints (7) and (8). Equalities (9) define the ride time of each user which is bounded by constraints (12). It is worth mentioning that the latter also act as precedence constraints because the non-negativity of the \(L_i^k\) variables ensures that node \(i\) will be visited before node \(n + i\) for
every user $i$. Finally, the inequalities (10) bound the duration of each route while (11) and (13) impose time windows and capacity constraints, respectively. This formulation is non-linear because of constraints (7) and (8) but there are ways to convert it to a mixed integer LP. Those techniques are discussed by Cordeau (2006) but need not be reviewed here.

The research on VRP saw a diversity of applications at the same time that the dial-a-ride problem was being addressed. Nussbaum (1975), Field (1976), and Hinds (1979) describe a software package for routing and scheduling transit buses (The program is called RUCUS for Run Cutting and Scheduling.) Fielding (1977) examines shared-ride taxis (similar to dial-a-ride). Bodin et al. (1978) study the routing and scheduling of street sweepers. Ghoseiri, Ghannadpour, and Seifi (2010) examine the problem of dispatching railroad locomotives.


Attention has also been given to finding procedures that can solve VRP problems. Agin (1975) examines a variety of algorithms. Buxey (1979) explores the possibility of using Monte Carlo simulation to find solutions. Baker and Rushinek (1982) examine large-scale implementation issues.

Since 1980, much has been done to advance the area of vehicle routing and scheduling. The literature search for this paper identified over 80 articles in this timeframe. Undoubtedly, there are more. Much of the research focuses on finding efficient algorithms to find solutions for specific problem formulations. Other investigators focus on specific types of problems. For example, Powell (1988) describes algorithms that can be used to solve the dynamic (time-based) routing of vehicles in response to known and anticipated loads.

Treatment of the problem from a stochastic standpoint starts about 1990. Laporte et al. (1992) address the problem of finding solutions to the vehicle routing and scheduling problem when stochastic travel times are present. A chance-constrained programming formulation is presented along with two stochastic optimization formulations and a branch-and-cut algorithm for solving all three formulations. The chance constrained formulation performs well as should be expected since it is a variant on the mixed integer LP formulation. Of the two stochastic optimization formulations, the one that more explicitly represents the problem formulation does much better. The authors conclude that such problems can be solved for significant size problems in reasonable time.

Many papers focused on solving stochastic vehicle routing problems followed Laporte et al. (1992). There is the notable paper by Bertsimas et al. (1995) and the proceedings paper by Taniguchi et al. (1999). Campbell (2004) describes heuristics for considering a variety of complicating constraints not typically included in the traditional formulations. Yamada, Yoshimura, and Mori (2004) is an interesting paper because it endeavors to use VRP procedures to study and assess road network reliability.

3.7 Advanced Routing and Scheduling Analysis Procedures

The advent of optimization schemes such as genetic algorithms, simulated annealing, and tabu search has motivated explorations of ways to use these techniques to solve VRP problems. The earliest investigation appears to be Garcia and Arunapuram (1993) who explored the use of tabu search. Potvin (2007) provides a survey of evolutionary algorithms that have been applied to VRP. Included in the review are genetic algorithms, evolutionary strategies, and swarm optimization. Weise, Podlich, and Gorldt (2009) provide a similar, newer review.


3.8 System and Network Reliability

System and network-level reliability are integrative manifestations of the reliability that arises on the links and at the nodes. They reflect the effects of operational and strategic decisions made to mitigate the impacts of travel time variability. For example, just-in-time delivery systems have to buffer travel time variability by planning on early arrivals and wait times. Systems that have one or more transshipment points can buffer the variability in link travel times by carefully setting the arrival and departure times of vehicles. The challenge is to time them so that the connection times for transferring the freight will be adequate. For rail systems, it is setting the schedules for inbound and outbound trains, as with the trucks, plus the sequencing of train classifications for inbound trains and the assignment of yard tracks to blocks. Of course, short connection times produce missed connections and delays. But long connection times, while they improve the connection probability, add to the time shipments spend in the system. And in that regard, for systems that are well instrumented, dynamic decisions can adapt the operating plan to the evolving conditions.
One of the earliest papers on system reliability is Detmold (1972). He discussed the issue of specifying timetables for train operations. His assertion was that looking at the timetable alone was insufficient. It was important to employ an integrated approach that combined shipper and carrier costs when determining the operational strategy.

Contemporaneous with Detmold (1972) other authors were providing perspectives on the connection between carrier operations and reliability. Lang (1972), Martland (1972), Williams (1972), Sussman (1974), and Shamberger (1975) all suggested ideas about how the reliability of railroad operations could be improved. All of them related to the reliability of train-to-train connections in yards and to equipment reliability for over-the-road trains.

Studies of blocking plans and yard management arose. Bodin et al. (1980) presented an optimization model for creating a blocking plan for an overall railroad. Turnquist and Daskin (1982) used queuing models to study the performance of classification yards. Yagar, Saccomanno, and Shi (1983) proposed an algorithm that they asserted would yield efficient train classification sequences. Van Dyke (1986) presented the automated blocking model (ABM), which is the genesis of many of the blocking procedures currently used by the railroad industry.

The spectrum of these studies has since increased to include other aspects of the service development process and other modes and multi-modal environments as well. Beginning in 1978, Bevilacqua discussed the relationships between using alternative modes of transport service, economic efficiency and energy consumption. Crainic and Rousseau (1986) presented a multi-commodity, multi-modal service design model that emphasized reliability as well as cost and operational efficiency. Kraay, Harker, and Chen (1991) studied the optimal pacing of trains to eliminate delays from meet and pass conflicts. Kraft (2002) examined the scheduling of freight deliveries using a bid price approach. Barnhart, Belobaba, and Odoni (2003) studied issues of reliable service design in the context of the air transportation industry. Chen and Schonfeld (2011) examined scheduling issues for single-hub intermodal freight systems.

In a highway context, Taniguchi et al. (2001) examined the issue from a city logistics perspective, including the potential impacts from capitalizing on the information sharing from intelligent transportation systems. Jones and Sedor (2006) described a study of reliability from an interstate trucking perspective. Lomax et al. (2003) began the process of doing national-level assessments of congestion and reliability in urban areas. Washburn and Ko (2007) focused on understanding the impacts of travel time reliability on the perceptions of highway level-of-service held by trucking companies. List et al. (2006) examined the repercussions of considering reliability in truck fleet sizing decisions for national-level services.

3.9 Value of Reliability

It is very clear from all of the foregoing review that reliability is perceived as being important: to carriers, shippers, the government, and other entities. Finding ways to improve it is a high priority.

But this sentiment begs the question about the monetary value perceived in reliability. How does that value compare with the valuation of time itself? If reliability is valued, then it ought to factor into mode choice decisions, service selection decisions, carrier selection decisions, etc., but does it? With these thoughts in mind, several studies have been performed to assess the value of reliability and see how it plays a role in various decision-making settings.

Van Der Mede, Palm, and Flikkema (1996) might have been the first to assert that travel time variability should be a “new” service quality indicator. They measured variations in travel time and subjective reaction to travel time variability. This was done for trips by cars and trucks from door-to-door. The data collection techniques included trip diaries for drivers, black-box data from trucks, and questionnaires. Their finding was that reliability did, indeed, have value.

Other studies followed that assessed the value decision-makers placed on this service attribute. Wigan et al. (2000) reported the values of travel time and reliability for long-haul and metropolitan freight services. Lam and Small (2001) reported the values of time and reliability that they obtained from a value pricing experiment. These studies continue to the present as prices change and the economy continues to evolve. Recent examples include Fowkes and Whiteing (2006), Zamparini and Reggiani (2007, 2010), Nunez et al. (2008), de Jong et al. (2009), Fosgerau and Karlstrom (2010), and Halse et al. (2010).

Weigman, Hekkert, and Langstraat (2007) assert that their research suggests reliability and costs are the most important aspects of service quality in the intermodal market. However, there appear to be a number of differences between how terminal operators see these attributes compared with customers. For terminal operators, reliability and flexibility appear to be more important than they are for customers. This suggested to them that terminal operators could reduce their focus on these aspects without reducing total perceived quality by customers. Moreover, less focus by the terminal operators on flexibility and reliability would offer opportunities for increased focus on other quality aspects (e.g. costs). For the customers, costs and total quality are more important. Other quality aspects also matter, but are relatively less important. Moreover, the differences among these less important quality aspects are small.

Researchers have also explored the value of reliability in the context of the role it plays in various kinds of freight movement decision-making. Poole (2007) explored its role in the context of truck-only toll lanes. Figliozzi and Zhang (2010) examined its impacts on cost. Ozkaya et al. (2010) studied it from the perspective of freight rates in the less-than-truckload sector. McLeod (2012) considered the role it should play as a performance measure for evaluating freeway systems.
3.10 Mode/Path/Carrier Choice

Inasmuch as reliability has been shown to have value from a decision-making standpoint, it should be possible to demonstrate that it influences mode, path, and carrier choice decisions. If shippers see a value in reliability, it should be possible to demonstrate that they would select mode or carrier B instead of mode or carrier A, ceteris paribus, if the reliability of B is better than A. In fact, some of the experiments upon which the value of reliability is based in the aforementioned studies use stated preference surveys as the data collection mechanism. Inherently, those instruments compare alternatives, as is done in mode, path, and carrier choice. Meixell and Norbis (2008) provide an excellent, recent review of the literature in this area. In their review, they repeatedly observe that reliability is an important factor in the choice of modes and even carriers within modes (especially for trucks).

Spurred by the advent of deregulation in the trucking industry, Bardi, Bagchi, and Raghunathan (1989) conducted a survey of 1,000 transportation shippers randomly selected from the Council of Logistics Management membership directory. Twenty-nine percent of those surveyed responded. Reliability was ranked the first out of 18 criteria by which a carrier could be selected. The next five criteria, in rank order, were door-to-door rates or costs, door-to-door transit time, rate negotiation flexibility, financial stability of the carrier, and equipment availability. It is clear from their research that reliability was important.

Crum and Allen (1997) also examined reliability as a factor in selecting one carrier over another. They reported the results of two surveys, one conducted in 1990 data and another in 1996. Based on the 1990 results, pick-up and delivery reliability was the top ranked criterion and transit time reliability was the second. In 1996, the order was reversed, but the two top measures were still the same.

Kent and Parker (1999) conducted a similar study similar to Bardi, Bagchi, and Raghunathan focused on the shippers of international containers. They surveyed export shippers, import shippers, and containerized transportation companies and asked for rank order evaluations of the same 18 criteria used by Bardi, Bagchi, and Raghunathan. The most important service attribute again proved to be transit time reliability/consistency. The next five attributes, in descending rank order, were equipment availability, service frequency, rate changes, and operating personnel. Transit time was sixth.

Swan and Tyworth (2001) looked at the issue the other way around, from the carrier’s perspective. They focused on customer retention, asserting that the US railroads were losing the most profitable share of their business by providing unreliable service. They argued that by choosing to focus on reducing costs, rather than providing better service, they were forcing their customers to shift to other modes, notably truck. They asserted that railroads should provide better service and recapture the costs by charging higher rates.

Bontekoning and Priemus (2004) made a similar assertion for intermodal services. They said that the main growth potential for intermodal was in markets for flows that demand speed, reliability and flexibility. They further said that innovations in service offerings will produce a breakthrough in modal split and allow the use of the mode to expand.
Shinghal and Fowkes (2002) presented the results of an empirical study of mode choice for mode choice in the Delhi to Bombay corridor. Travel time, reliability, and service frequency are all found to be important. Service frequency is the most important attribute. The importance of reliability was generally lower than the authors expected. The reliability of transit times was found to be very important for exporters and the auto parts sector due to the effect it can have on the production process.

Danielis, Marcucci, and Rotaris (2005) conducted a formal study of freight mode preferences among logistics managers in two regions of Italy. Four attributes were employed to characterize each hypothetical option: cost, time, reliability and damage/safety. Two estimates were obtained of each attribute were obtained: (1) the utility associated with each level of the same attribute, and (2) the attribute utility revealed by an ordered probit model. Both estimates indicated, on average, a strong preference for attributes of quality (time, reliability and safety) over cost. They felt this indicated that modal shift policies needed to focus more on the quality aspects of the modes rather than just their costs.

Fowkes (2007) considered the concepts of freight value of time and reliability in the context of shipments in the UK. He presents findings for nine commodity groups as well as the group overall. Care was taken in developing the results since the estimated valuation of one attribute can vary depending on the presence of a related variable. The main empirical finding was that, when respondents ignored driver and vehicle costs, for many commodities the valuations of improvements in journey time and its variability were negligible. However, shippers of some commodities did exhibit a willingness to pay for improvements, and occasionally a lot.

Fries (2008) reported the results of an effort to develop a freight demand model that could be a comprehensive tool for freight demand forecasting in Switzerland. Fries presents the methodology and results of the project focusing on the development of modal split functions that represent the shippers' demand elasticities. The core part of this project consisted of preparing and executing a survey among shippers and freight forwarders. Stated preference experiments based on revealed preference data were conducted within the framework of the survey to collect the data necessary for the estimation of modal split functions for different commodity groups. Interestingly in several commodity groups reliability was ranked equal to or even higher than transport cost. Moreover, travel time was generally less important than reliability.

Grosso and Montiero (2008) did a similar study in Italy. They were interested in seeing what factors influenced the decision about choosing a port. A questionnaire was sent to about 30 companies, including shipping companies, freight forwarders and shippers, currently operating in the port of Genoa. They found that port service reliability was among the criteria used.

Train and Wilson (2008) did a study of grain shippers in the upper Mississippi River valley. They sent survey forms to 2,000 shippers and received responses from 480. The survey presented changes in rates, transit times, and reliability, and the respondents were asked to state how their annual volumes shipped by barge as opposed to truck or rail would change given that all other factors remained the same. The basic finding was that, as might be expected, larger declines in reliability increased the likelihood that firms would adjust the volume shipped by barge. For
example, if the percentage change in reliability was less than 10%, the elasticity for those shippers that actually made a change was 2.417. That is, for them, a 1% decrease in reliability would result in a 2.417% decrease in shipment volume. For all of the survey respondents, including those that did not make a change, the elasticity was 0.619. That is, a 1% decrease in service reliability would produce a 0.619% decrease in the use of barge. In comparison, these same elasticities for a change in rates were -1.407 and -0.075, and for a change in transit time, the elasticities were -1.841 and -0.310 respectively. In these latter two cases the elasticities are negative because increases in either rates or travel times would result in a decrease in the use of barge. The absolute values are what are important for comparison purposes, and those values show that the sensitivity to reliability was the highest among these top attributes.

Brooks et al. (2012) examined the Australian domestic freight transport market with a focus on the decision-making process by which cargo interests and their agents make mode choice decisions between land-based transport and coastal shipping. While their ultimate interest lay in seeing if short-sea shipping could provide a reduced carbon footprint to truck and rail, they nonetheless looked at shipper sensitivities to various service attributes including reliability. The attributes examined were: service frequency, cost (price), transit time, freight distance, direction (headhaul/backhaul), and reliability, measured both by arrival within the delivery window and delay. The authors concluded that shippers would be willing to pay significant amounts for improvements in the on-time reliability of rail, road, and short-sea shipping.

### 3.11 Location Decisions

Building a distribution center, a terminal, or any other type of large facility is a major investment decision. Ceteris paribus, it makes sense to cite these facilities at locations where the travel times will be reliable to all major trip origins or destinations. On the manufacturing side, it makes sense to locate plants where the inbound and outbound travel times are reliable. Of course, the travel times are only part of the overall stochastic process. So solutions that focus on the travel times are only myopically optimal, but it still an interesting area on which to focus research.

Mirchandani and Odoni (1979) may have been the first to consider the location of facilities on networks where the travel times were stochastic. They examined problems where the travel times on the network links were random variables with discrete probability distributions. They demonstrated that solution algorithms for such problems can be developed and reasonable size problems can be solved as long as the number of states of the system (considering the stochastic travel times) is small.

More from the standpoint of siting emergency response services than logistics facilities, Mirchandani (1980) again considered the problem of locating facilities when the travel times are stochastic. He shows that realistic and reasonable size problems can be formulated and solved using a variety of solution techniques. Daskin (1985) reviewed the location decision-making literature and indicated that “both demands and link travel times should, in principle, be treated as random variables” as should the demands. But his review does not identify location models where the network travel times are stochastic.
Owen and Daskin (1998) provide a second review in which stochastic location problems are considered. They indicate that “any number of system parameters might be taken as uncertain, including travel times, construction costs, demand locations, and demand quantities. The objective is to determine robust facility locations which will perform well (according to the defined criteria) under a number of possible parameter realizations.”

In spite of these assertions that it is important to treat the travel times as stochastic, it appears that only limited work has been done to advance this frontier. Wang and Ma (2008) appear to be the next authors to explore ways to solve this problem. They use a mixed genetic algorithm to solve problems of various sizes. The results are compared with two greedy heuristic algorithms which have been shown to be good at solving set covering location problems. The computational experiments show good performance for the mixed genetic algorithm.

More recently, Fazel-Zarandi, Berman, and Beck (2013) have considered stochastic facility location / fleet management problems where the travel times are random variables. They use stochastic programming to solve the problem. Two-level and three-level logic-based Benders’ decomposition models are employed. The computational experiments show that the models are able to substantially outperform an integer programming model they also present in terms of finding optimal solutions.

3.12 Logistics and Supply Chain

When supply chains become the focus of a reliability analysis, the whole end-to-end process is brought into the picture. The travel times are one source of stochasticity, but there are others: manufacturing disruptions, component delivery delays, and demand variability. In cases like shipping coal where one can assume an infinite supply of material at the source, managing the variability in the travel times becomes the major focus. In other instances, some of the other sources of stochasticity may be more important. Arvis, Raballand, and Marteau (2007) provide a thoughtful commentary on the impacts of supply chain reliability on logistics costs, especially for enterprises located in landlocked, third-world countries. Vernimmen, Dullaert, and Engelen (2007) provide a similar review of the impacts of schedule reliability for shippers worldwide that rely on containerships in their logistics networks.

Whybark (1974) appears to have been the first to focus on reliability in the context of a supply chain. He asserts that while most organizations set inventory policies and choose transportation alternatives separately, there is an interaction between these decisions when the transportation alternatives have different speed, reliability and cost characteristics. Whybark presents a heuristic procedure for jointly determining the reorder points, order quantities and transportation alternatives for minimum total transportation and inventory costs for a receiving facility.

Almost ten years later, Allen, Mahmoud, and McNeil (1985) present a model that shows how a cost-minimizing shipper should adjust its economic order quantity (EOQ) as reliability and/or time in transit changes. A matrix shows the minimum cost attainable with each combination of mean and variance values for the transit time distribution. In addition, by comparing one cell with another, the matrix shows how costs are affected by changes in the mean transit time and the variance. It shows how reductions in cost can in some cases be achieved by improving
reliability while increasing average transit time. The paper also shows how reorder points can be adjusted in response to changes in the travel time mean and variance.

Muthuraman, Seshadri, and Wu (1991) present an analysis similar to that of Allen, Mahmoud, and McNiel. They combine a continuous time back-ordered inventory system with stochastic demand and stochastic delivery lags for placed orders. By modeling demand as a diffusion process, they reformulate the inventory control problem as an impulse control problem which they then simplify into a Quasi-Variation Inequality (QVI). This allows them to obtain the optimal ordering policy, the limiting distribution of the inventory level, and the long run average cost. Computational experiments show that significant losses can be incurred in approximating a stochastic lead time system with a fixed lead time system, highlighting the need to consider the stochasticity in the lead time.

Later articles consider variations on these ideas. Ouyang and Chang (2001) create a formulation that uses fuzzy set theory to determine the backorder rate. Zhao and Simchi-Levi (2006) consider an assemble-to-order problem where the lead times are stochastic. Hnaien, Delorme, and Dolgui (2008) examine supply planning for two-level assembly systems with stochastic component delivery times. They study the trade-off between holding cost and service level. A basic multi-objective meta-heuristic is used to explore the trade-off between holding cost and stockout probability. Louly, Dolgui, and Hnaien (2008) explore ways to address problems involving supply planning for single-level assembly systems with stochastic component delivery times and a high service level constraint.

A recent study of the impacts of reliability on logistics managers and freight operators is reported by McKinnon et al. (2008). The project examined the impacts of congestion for nine sectors of the economy and inquired about company reactions to the significant deterioration in road traffic conditions in the United Kingdom (UK). Thirty-seven managers were interviewed in 28 companies or divisions of companies. Five visits were made to distribution centers. Detailed inquiries were made about the impact of congestion on the logistics operations, the relationship between congestion and other sources of unreliability and any measures companies were taking to mitigate the effects of congestion. Very few of the companies were able to provide statistical information about the operational and financial impacts. The interview data suggested that there were wide variations in the impact of congestion. There was little evidence of congestion causing companies to restructure their logistics systems. Nor was it causing companies to run more vehicles, increase tractor-trailer ratios, carry more inventory or modify internal warehouse design and capacity. Companies learn to compensate by altering schedules, building in extra slack, making internal processes more flexible and, in some cases, upgrading their dispatching systems. There were companies, however, for which congestion did have a significant impact. And for them, congestion was clearly having serious impacts on cost and quality of service.

Jane (2011) presents a technique for estimating the reliability of a distribution network. The measure is identified as $R_{b,d}$. It is the probability ($R$) of successfully delivering a particular demand $d$ to a specific destination given a budget constraint of $S_d$. Jane presents a hybrid approach for computing this metric and demonstrates that it is effective and efficient. It is interesting for several reasons: 1) it defines reliability in the classical sense of on-time deliveries, 2) it combines the reliability of the source location with that of the transportation system, and 3)
it applies a budget constraint to achieve the assessment. It represents a holistic perspective on the reliability of the supply chain.

Lai (2011) presents a similar analysis based on the semiconductor industry. Multiple manufacturing sites are involved, stochastic travel times are assumed, the vehicles have capacity constraints, and time windows are imposed for both pick-up and delivery. A chance constrained programming model is employed. It uses two categories of chance constraints, one for the time windows and another for the duration of driver service. The objective is to find a set of routes and schedules for the vehicles that minimizes travel distance without violating time windows, product/vehicle compatibility, pickup and delivery, driver duration, and vehicle capacity constraints.

Lee and Song (2011) provide a commentary on the importance of reliability in the context of the maritime industry. They say that within the context of the maritime industry, maritime logistics value represents the quality with which shipping and port operators meet reliability. An exploratory case study within the Korean maritime transport industry is used to conclude that the most valuable knowledge is acquired through having maritime transport operators embedded in cooperative/coopetitive networks, which then improves the maritime logistics value which they offer.

Kim and Simchi-Levi (2012) consider what could be regarded as a “realistic, real-world problem” in the logistics world of today. A company is viewed as operating multiple delivery modes such as standard freight shipping and air. The lead times (travel times) are viewed as being stochastic (subject to delays). A model is presented that shows how to make the best use of these multiple delivery modes to minimize the impacts of lead time variability. As might be expected, the model depends on an order tracking system so that expedited handling becomes possible. The goods move stochastically among the installations and the system faces a stochastic demand. Kim and Simchi-Levi identify systems that result in simple and tractable optimal policies, in which both regular ordering and expedited handling follow a variant of the base stock policy. They show that optimal expediting results in a significant reduction in the total logistics cost and the reduction increases as variability in delivery lead time increases. They also show that expediting allows the system to be operated in a leaner way due to the reduced regular order amount and provide various managerial insights linking expediting, base stock levels, and expediting costs. This study exactly illustrates the holistic perspective on supply chain management that is needed to mitigate the impacts of stochasticity in the transportation system.

Hayya et al. (2013) present an interesting study focused on stochastic lead times for JIT delivery systems. They point out that FIFO (first-in-first-out) may not be preserved when stochasticity is present. They find that with stochastic lead times there is a possibility of order crossover, and what order crossover does is to transform the original lead times into effective lead times. The mean value of arrivals at the destination is the same as that at the origin, but the variance can be smaller. The implication is that when order crossovers are considered, the cost can be less than it would be without the crossovers. They demonstrate some important properties of the effective delivery time.
Li (2013) studies an integrated logistics network design problem and endeavors to optimize the assignment of supplier locations to terminal facilities. It allows for expedited shipments. Li first formulates elementary models for certain special case problems and discusses their properties and solution methods. Li then proposes a mathematical programming model that minimizes the sum of supplier set-up costs, expected shipment costs for both regular and expedited services, and expected inventory holding cost under stochastic demand rates and transportation lead times. This represents a general network logistics system working against a planning horizon. After studying the properties of the problem, Li develops a solution approach that makes use of Lagrangian relaxation. This approach is tested on three logistics networks of different scales. From these analyses, Li notes that under the optimal design, utilizing expedited shipment services actually does not produce much extra cost while it guarantees service reliability. Li also finds that with the integrated design, all planning and operational components complement each other in an optimal, holistic manner.

### 3.13 Public Sector Planning

Public sector planning is a recent addition to the freight world and a public focus on freight reliability is even newer. The public planning process for at least the past 40-50 years has focused on facilitating journey-to-work trips and transit services. It is only very recently that goods mobility has surfaced as an objective.

A conference was held at Sussex University in 1986 to discuss the implications of freight insofar as infrastructure policy is concerned (see Wright, 1986) but this appears to be a singularity.

There were some studies of hazardous materials transport that focused on the ability of existing infrastructure to accommodate the shipments. Examples are Beattie (1989), Tuler, Kasperson, and Ratick (1989), and List, Mirchandani, and Turnquist (1990). The latter article presents models that can be used to select the safest routes for hazardous material shipments and the best places to locate emergency response teams. This often-cited paper laid the groundwork for the routing and siting studies that followed.

A landmark conference focused on the freight data needs of public agencies was held in 2001 (Meyburg and Mbwana, 2001). Perspectives were presented by industry, consultants, the Federal government, and academics. New York State DOT was highlighted as an example of where the needs and capabilities stood for a state agency. This conference heralded the beginning of the current focus on public sector participation in freight planning. The most recent in this series of conferences is the one held in 2013 (Transportation Research Board, 2013).

Also in 2001, Barber and Grobar (2001) described a statewide goods movement strategy and performance measurement process for California. Their findings are as true today as they were then. Performance indicators from passenger travel assessment like delay can be redefined and adapted to the freight context. A key performance indicator at ports and other terminals is the truck waiting/turnaround time. This can be used as a measure of the performance of the truck-terminal gate interface in the goods movement supply chain. For the ports of Los Angeles and Long Beach, the aggregate impact was estimated at $3.8 million in terms of waiting delays. The study further found that capacity is the limiting factor. It impedes the goods movement supply
chain and makes it vulnerable to stress during peak or surge conditions. In the absence of major infrastructure improvements, voluntary demand management measures should be implemented to curb the impacts.

The following year, 2002, Caldwell and Sedor (2002) announced the debut of the Freight Analysis Framework (FAF). Based on three years of work, its objective was to improve the Federal government’s understanding of the nature of freight movement, identifying challenges to improving freight productivity and security, and developing strategies to increase freight productivity. The Federal Highway Administration identified the following key challenges confronting freight transportation: congestion and capacity, operations, planning, financing, safety, national security, environmental impacts, and professional capacity building. Work on the FAF is still underway today as evidenced by Wurfel et al. (2008).

Also in 2002, ICF Consulting and HLB (2002) presented the freight story insofar as the economic effects of transportation are concerned. Their bottom line comment was that “transportation policy and planning is not as robust as it should be in relation to the freight sector. For instance, project analysis tools do not appropriately recognize how and why infrastructure design and capacity problems drive down the productivity of freight transportation and drive up the cost of industrial production. Likewise, transportation planners and decision-makers cannot anticipate readily how infrastructure improvements would make freight carriers, their industrial customers, and the economy at large better off. With a significant portion of the focus of transportation policy and planning shifting to freight-related matters, filling the planning gap is essential. Clearly, highway investments that increase capacity and/or speed and reduce accidents will improve the performance of trucks, as will improvements in operations planning. Improvements in intermodal connections will also have an effect. Furthermore, Intelligent Transportation Systems (ITS) can be particularly important, especially when they reduce incident-based congestion. It is clear that transportation agencies at all levels of government can bring about improvement in highway freight-carrriage.” Beneficial actions can include “targeted capacity expansion projects that alleviate high-frequency bottlenecks in the freight system can improve transit time variability. Freight planning can help to make sure that freight movement needs are appropriately considered by decision-makers by providing state and local transportation planners with the necessary tools to better account for the impacts of alternative investments on the efficiency of the freight system. Programs that strive to improve operations planning (or the interaction of planning and operations functions within a transportation agency) can improve system performance. ITS deployment can enhance the efficiency of the highway system through operational improvements, better user information, and incident management (which is particularly problematic from the perspective of system reliability). [And] Federal grant programs [can] provide financing mechanisms for freight transportation improvements can help to generate the types of investments needed to improve the productivity of the freight system.” That these issues are still as relevant today as they were a decade ago can be seen in Strocko and Schoener (2012) and Herr (2013).

Metrics have been proposed that can be used to assess the performance of the transportation network insofar as goods movement is concerned. Lomax et al. (2003) set the stage with their monograph on travel time reliability measures. Articles that have suggested additional measures and/or frameworks include Cambridge Systematics (2004), Cook and Kent (2007), Pickrell (2007), Lyman and Bertini (2008), Varma (2008), Eisele and Schrank (2010), Li and Hensher
(2012), Young (2012), Casavant (2013), Li et al. (2013), Zhao et al. (2013), and Eisele et al. (2014).


A special emphasis on city logistics has been focus of additional articles. Among these are Taniguchi and Thompson (2002), Raicu et al. (2005), Yamada and Taniguchi (2005), Barcelo et al. (2008), D’Acierno et al. (2010), Figliozzi (2011), Walsh and Somers (2011), Dominicis et al. (2012), Motraghi and Marinov (2012), and Saiyed et al. (2012).

More general examinations of transport planning in the context of logistics have been forthcoming as well. Among these articles are Kreutzberger (2008), Winder (2009), Iikkanen et al., (2012), Tavasszy et al. (2012), Kittelson and Associates (2013), Li and Savachkin (2013), and Kittelson (2014).

Much has been done to advance the understanding of what the public sector should do insofar as freight transportation is concerned, but much still remains to be done. Theory and ideas need to be brought into practice, as was evident from the most recent freight planning workshop (Transportation Research Board, 2013).

4.0 SUMMARY AND CONCLUSIONS

This state-of-the-art review paper has focused on freight reliability. It presents a description of prior work based on a taxonomy which helps to categorize prior work and identify where there will be opportunities in the future.

The earliest research efforts in freight reliability focused on the impacts of stochasticity on routing, logistics management, and travel time reliability within specific modes. More recent efforts aim to find solutions for multiple-vehicle routing problems, multi-modal logistics networks, and optimal mode choice and path selection. More than five hundred articles were found in preparing this paper.

To help facilitate the presentation of the literature, eight areas of research have been employed:

- **Link, node, segment and route reliability.** Examination of the variation in travel times within a specific mode or among a collection of nodes for network routes and segments.
- **System-level reliability.** Studies of reliability in terms of system-level assessments. An example is a trucking company where multiple terminal handlings are involved, plus pick-up and delivery from the shipper to the receiver. The absence of reliability creates missed connections and shipment delays; and these impacts can be mitigated, albeit at a cost, by building slack into the operating schedule.
• **Vehicle routing and scheduling.** Examination of decisions about how to assign loads to vehicles, how many vehicles to employ, how to route and schedule them, and how to make all of these decisions effective in light of the reliability of the network.

• **Facility location.** Decisions about where to locate fixed assets such as warehouses, transportation terminals and manufacturing plants.

• **Mode/path/carrier choice.** Examinations of travel time reliability for routes that involve two or more modes, such as truck-rail-truck, where the mode-specific segments of the trip have an impact on overall reliability.

• **Supply chain logistics.** Decisions associated with mode and path choice for single and multiple shipments, and the impacts of unreliability on economic order quantities, ordering intervals, inventory levels, and other aspects of supply chain management.

• **Public sector planning.** Investigations of the role public agencies should play in facilitating improved freight reliability. A simple example is network capacity investment to alleviate bottlenecks and congestion-caused delays. Another is pricing strategies that help ensure reliable travel times for those who choose to pay. A third is multi-modal investments to facilitate coordination between modes.

• **Reviews and syntheses.** Papers that review the literature related to freight reliability from a variety of perspectives.

The bottom line is that extensive research has been conducted insofar as vehicle routing is concerned. Also significantly examined is the assessment of the value of travel time reliability using empirical data. Less explored are the areas of how public agencies should make decisions insofar as freight is concerned and how to use of optimization techniques in conjunction with simulation to find the best solutions to freight reliability problems.

See below for general format.
REFERENCES


Barber, D., & Grobar, L. (2001). IMPLEMENTING A STATEWIDE GOODS MOVEMENT STRATEGY AND PERFORMANCE MEASUREMENT OF GOODS MOVEMENT IN CALIFORNIA (pp. 48 p.-48 p.).


Belovarac, K., & Kneafsey, J. T. (1972). STUDIES IN RAILROAD OPERATIONS AND ECONOMICS. VOLUME 3. DETERMINANTS OF LINE HAUL RELIABILITY (pp. 97 p.-97 p.).


Caldwell, H., & Sedor, J. (2002). THE FREIGHT STORY: A NATIONAL PERSPECTIVE ON ENHANCING FREIGHT TRANSPORTATION (pp. 38 p.-38 p.).


Fielding, G. J. (1977). SHARED-RIDE TAXI COMPUTER CONTROL SYSTEM REQUIREMENTS STUDY (pp. 53 p.-53 p.).


Folk, J. F. (1972). STUDIES IN RAILROAD OPERATIONS AND ECONOMICS. VOLUME 5. MODELS FOR INVESTIGATING RAIL TRIP TIME RELIABILITY (pp. 213 p.-213 p.).


Herr, P. R. (2013). Department of Transportation: Key Issues and Management Challenges, 2013 (pp. 29p-29p).


List, G., Mirchandani, P., & Turnquist, M. (1990). Logistics for hazardous materials transportation: scheduling, routing and siting (pp. 1 vol (various pagings)-1 vol (various pagings)).


Little Inc, A. D. (1971). ECONOMIC IMPACT OF FREIGHT CAR SHORTAGES. EXECUTIVE SUMMARY (pp. 8 p. -8 p.).


Nussbaum, E., Rebibo, K. K., & Wilhelm, E. (1975). RUCUS (RUN CUTTING AND SCHEDULING) IMPLEMENTATION MANUAL (pp. 198 p.-198 p.).


Transportation Research Board. (2013) Innovations in Freight Demand Modeling and Data. (pp. 4p): Transportation Research Board.


Weise, T., Podlich, A., & Gorldt, C. (2009). *Solving real-world vehicle routing problems with evolutionary algorithms* *Natural Intelligence for Scheduling, Planning and Packing Problems* (pp. 29-53): Springer.


Young, R. (2012). Use of Travel Time, Travel Time Reliability, and Winter Condition Index Information for Improved Operation of Rural Interstates.


