ABSTRACT

Title of Thesis: HOLDING DECISIONS FOR CORRELATED VEHICLE ARRIVALS AT INTERMODAL FREIGHT TRANSFER TERMINALS

Yanshuo Sun, Master of Science, 2016

Thesis Directed By: Professor Paul Schonfeld,
Civil and Environmental Engineering

A vehicle holding method is proposed for mitigating the effect of service disruptions on coordinated intermodal freight operations. Existing studies are extended mainly by (1) modeling correlations among vehicle arrivals and (2) considering decision risks with a mean-standard deviation optimization model. It is shown that the expected value of the total cost in the proposed formulation is not affected by the correlations, while the variance can be miscomputed when arrival correlations are neglected. Some implications of delay propagation are also identified when optimizing vehicle holding decisions in real-time. General criteria are provided for determining the boundary of the affected region and length of the numerical search, based on the frequency of information updates. Theoretical analyses are supported by three numerical examples.

HOLDING DECISIONS FOR CORRELATED VEHICLE ARRIVALS AT INTERMODAL FREIGHT TRANSFER TERMINALS

by

Yanshuo Sun

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2016

Advisory Committee:
Professor Paul M. Schonfeld, Chair
Associate Professor Qingbin Cui
Associate Professor Cinzia Cirillo

Dedication

To my beloved parents.

Acknowledgements

Upon the completion of this master's thesis, I would like to express my sincerest gratitude to my advisor Prof. Paul Schonfeld, who is truly a scholar and mentor. It is my great honor to spend time learning from and working with him. Without his encouragement and patience, this thesis would not have been completed. I also want to extend my gratitude to two other committee members, Dr. Cinzia Cirillo and Dr. Qingbin Cui, for their help in improving my work.

I benefited from discussions with Dr. Ilya Ryzhov, Dr. Nikola Markovic and Xuechi Zhang in the mid-stage of this study. Their useful comments are appreciated. I am also grateful for the research funding from the US Department of Transportation through the National Transportation Center @ Maryland.

I want to thank all my friends in the civil engineering department, who made my life in College Park colorful. Special thanks go to my wife Dr. Qianwen Guo for her caring and love.

Table of Contents

List of Tables

List of Figures

1. Introduction

This study is motivated by the insufficient understanding of the impact of correlated arrivals on vehicle holding decisions. In the literature, inbound vehicle arrivals are assumed to independent for the sake of mathematical convenience and simplification. However, this clearly sacrifices realism because correlations are pervasive in the real world. Inclement weather or congestion might affect all vehicle arrivals in a certain region. A group of vehicles may be delayed temporarily due to roadway construction activities. Vehicles may arrive through a “gate” with limited capacity, such as a runway. All these factors could contribute to arrival correlations. When correlations are positive, delayed vehicles tend to arrive together in “platoons” or “bunches”. The effect of arrival correlations on vehicle holding decisions is still unexplored in the literature and thus unclear. The primary contribution of this thesis is to analytically incorporate such effects into vehicle holding studies.

In substantial freight transportation systems, as well as in air transport of passengers and public transit, it is uneconomical to provide direct services for each origin and destination pair. Freight operators design and maintain transfer-dependent networks where cargos must make connections at transfer hubs. In addition, operators provide a range of services, such as local collection trucks and main-haul by rail or road, to cater to specific demand markets on different parts of their networks (Macharis and Bontekoning 2004). For instance, as shown in Figure 1-1, to serve cargos originating from the Washington-Baltimore area to the New York area, freight carriers might not provide a point-to-point long-haul truck service. Instead, shipments are collected by local trucks and transferred to trains at intermodal transfer stations. Then, most cargo shipments are consolidated on the train link between Baltimore and New York City. Significant cost reductions are expected if large volumes of cargos are transported by trains rather than long-distance trucks. Nonetheless, compared to direct services, such transfer-based services incur additional costs (e.g., waiting and handling) at intermodal transfer stations. Those additional costs can be reduced significantly if we can coordinate (i.e. synchronize or nearly synchronize) vehicle arrivals at intermodal transfer stations.

Due to various stochastic elements, such as demands and traffic congestion, coordinated transfers are difficult to achieve, unless safety margins, also called slack times, are built into the coordinated schedules. In addition to pre-planning which seeks to generate coordinated schedules, operational controls at intermodal terminals are required to maintain a reasonable level of successful connections. This thesis focuses on a specific control measure - vehicle holding. The term “vehicles” can represent here aircraft, ships, trucks and trains.

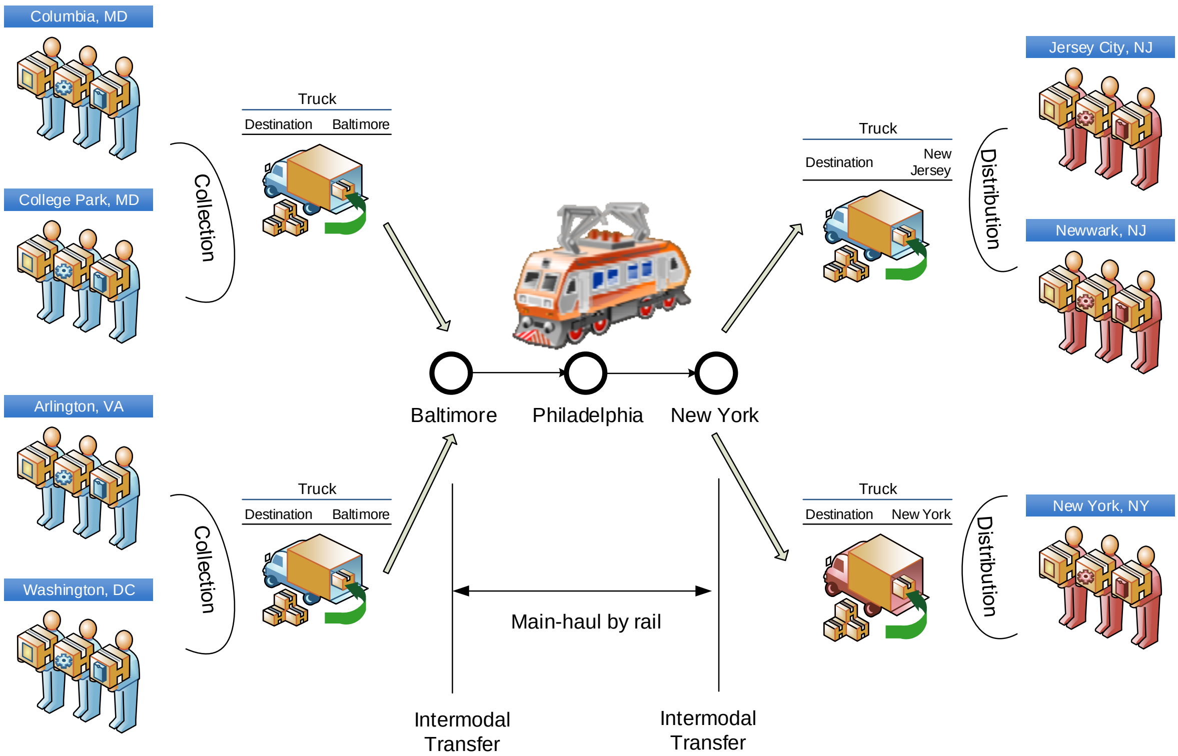


Figure 1-1 Intermodal freight network

Suppose that vehicle arrivals on two feeder routes (one serving southern Maryland and the other serving downtown Washington DC, as shown in Figure 1-1) are coordinated with the departure of a Baltimore-Philadelphia-New York train. When the train is ready to be dispatched, the vehicle on one feeder route is delayed and still on the way to the Baltimore Intermodal Station. In this case, we must decide whether to hold the train and wait for the late vehicle. Holding may be justified to allow cargos on the late vehicle to make the connection; otherwise, a missed-connection cost is incurred and left over cargos have to be carried by the next available train. On the other hand, vehicle holding might increase the waiting by the ready cargos at a downstream station such as Philadelphia. If we hold the train, it will probably run behind its schedule and timed transfers at downstream stations (e.g., New York City) will consequently be disrupted. Therefore, a systematic method is needed to facilitate the vehicle holding decisions by considering additional waiting costs and connection failure costs.

In the literature, although there are essential relations between schedule coordination (also called schedule synchronization or timed transfer) and vehicle holding, relevant studies are usually grouped into two categories, depending on the application purpose, namely pre-planning or real-time control. After briefly reviewing the first stream of studies and examining vehicle holding studies in detail, one would find:

(1) Far fewer studies are devoted to freight transport systems compared to public transit systems, although vehicle holding should have better applicability in goods transportation systems.

(2) There is one critical shortcoming within current vehicle holding models, which assume independent vehicle arrivals. The neglect of arrival correlations can cause inferior vehicle holding decisions.

(3) Significant uncertainties in the objective, which is random, are ignored, with only the expected value being optimized.

(4) Implications of delay propagation effects are not sufficiently addressed.

Therefore, this study presents a more realistic vehicle holding model for managing service disruptions in intermodal freight systems, especially by explicitly modeling vehicle arrival correlations.

The remainder of this thesis is as follows. Chapter 2 reviews relevant studies on both schedule coordination and vehicle holding in air, public transit and freight transportation systems. Chapter 3 presents basic assumptions, formulations of various cost functions, and the complete vehicle holding model. It also briefly discusses solution methods. Chapter 4 presents a simplified example and Chapter 5 provides larger scale case studies to demonstrate the effectiveness of the proposed method.

2. Literature Review

2.1. Schedule Coordination

When a vehicle holding decision is considered, it is implicitly assumed that the held vehicle is coordinated with other vehicle arrivals in a transfer-dependent network. Such practice, which is called "Schedule Coordination", is abundant in public transit (Ting and Schonfeld 2005, Zhao et al. 2006), freight (Jeong et al. 2007, Chen and Schonfeld 2010) and air transportation systems. For example, FedEx Express operates a huge hub-and-spoke network with Memphis International Airport as its "Super Hub", which collects millions of packages from feeder flights and distributes them to hundreds of domestic and international destinations in a few hours every night. Compared to a direct point-to-point service network, such a transfer-based network can greatly reduce the capital and operating costs because origin-destination pairs in the network can be served with significantly fewer routes. However, it is inconvenient to make transfers for some cargos, especially when the transfer time is long and unreliable. By synchronizing vehicle arrivals and departures at hubs or transfer terminals, the following benefits can be obtained: 1) improving users' transfer experiences with significant transfer time savings; 2) reducing storage requirements and inventory costs; and 3) reducing handling costs, e.g. due to more direct vehicle-to-vehicle transfers. Nonetheless, such benefits might be voided when random events force vehicles off the pre-optimized schedule, leading to a connection failure in a probabilistic system.

To maintain the transfer reliability at a reasonable level, safety margins called slack times are built into the schedule in order to absorb the randomness of the system. Different kinds of models (analytical or simulation) for various systems are proposed by Hall (1985), Abkowitz et al. (1987), Lee and Schonfeld (1991), Chowdhury and Chien (2002), Ting and Schonfeld (2005), Zhao et al. (2006) and Kim and Schonfeld (2014).

The pre-optimized schedule, which is obtained with methods developed in the above studies, may become suboptimal or even infeasible when random delays strike the system operation dynamically. Real-time control is needed to mitigate the impact of random incidents and bring the operation back on schedule. The following reviews, under three subheadings (i.e., air, public transit and freight), thus focus on the operational control.

2.2. Aircraft Holding

Major airlines develop and operate their hub-and-spoke networks (Figure 2-1). For example, Delta's hubs within the U.S. include Atlanta, Boston, Detroit, Los Angeles, Minneapolis/St. Paul, New York-JFK, New York-LaGuardia, Seattle and Salt Lake City. These hub airports serve as transfer terminals for passengers from and to various spoke flights. To improve the transfer efficiency (e.g., reducing the layover times), airlines usually schedule aircraft arrivals and departures at their hubs in "banks", which represent waves or batches of flights. For example, passengers from three incoming flights, i.e., Norfolk to Atlanta, Tampa to Atlanta, and Nashville to Atlanta, are scheduled to transfer to outgoing flight "Atlanta - Seattle" at the Atlanta airport. Three hub-bound flights thus form an arriving bank. They land in a narrow time window, limited mainly by the airport's capacity. Similarly, departing flights from the hub to various destinations would also be scheduled soon after the arrival bank exchanges passengers and baggage among its flights.

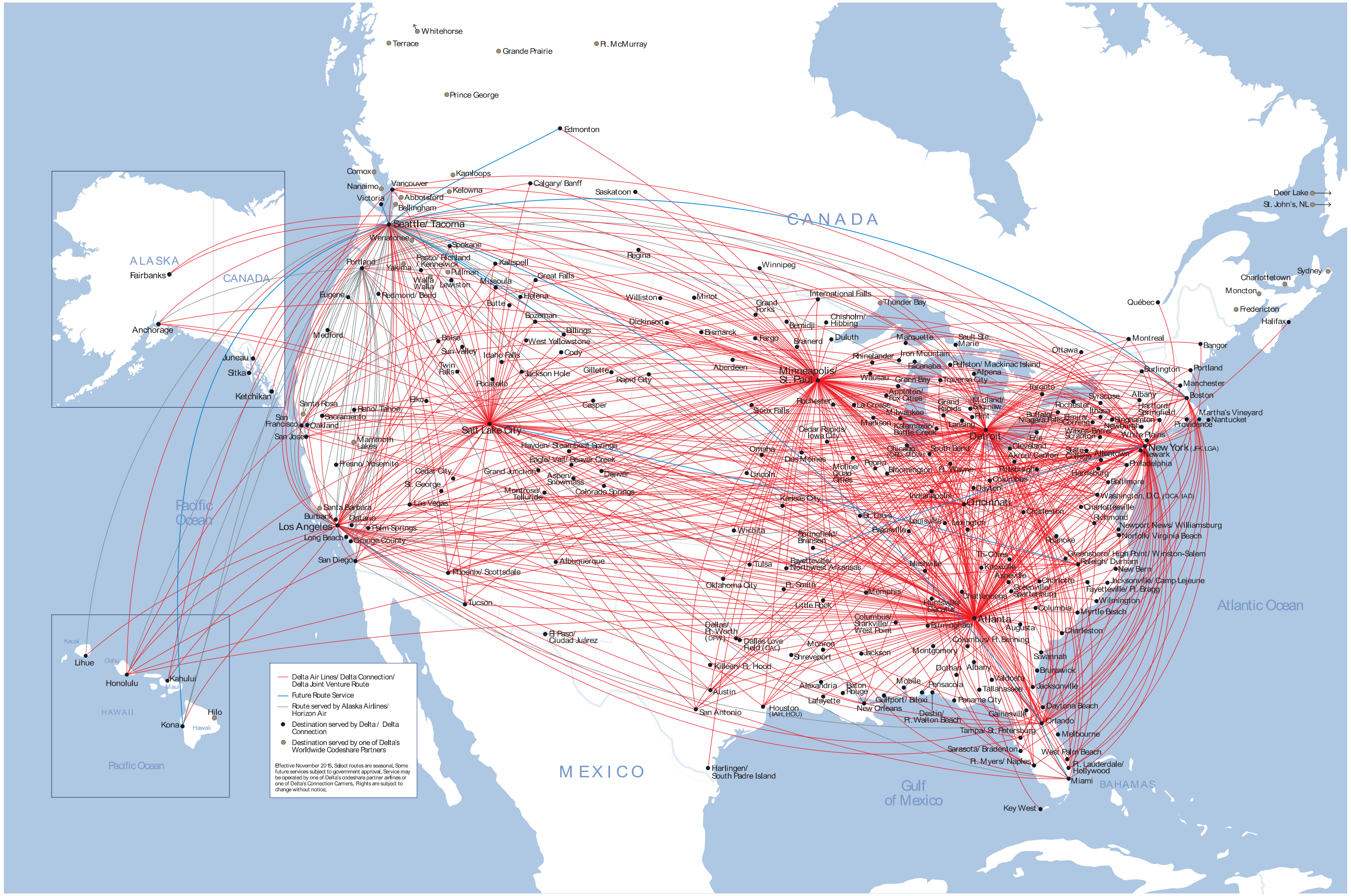


Figure 2-1 Domestic Route map of Delta Air Lines (Modified version)

Source: http://www.delta.com/content/dam/delta-www/pdfs/route-maps/us-route-map.pdf

Due to various factors, such a timed transfer could be disrupted. For instance, if the arrival capacity of the Atlanta airport is reduced by inclement weather, three incoming flights cannot land on schedule. A ground holding control is thus justified to hold aircraft at their departure gates to avoid more costly and dangerous airborne holding. This generally falls into the realm of air traffic management. Ground holding studies have been published by, among others, Odoni (1987), Richetta and Odoni (1993), Vranas et al (1994), Hoffman and Ball (2000). Various models, static vs dynamic, deterministic vs probabilistic, single-airport vs multi-airport have been well documented in the literature. Because en-route travel times are assumed to be predicted with sufficient accuracy, researchers are only interested in optimizing the controlled arrival time of each incoming flight at the adversely affected airport. After the controlled time of arrival is determined, the controlled time of departure can be computed and the ground delay is thus determined (Hoffman and Ball 2000).

The ground holding problem can be illustrated with a basic mathematical program. Let F be the set of incoming flights whose arrival slots have to be assigned. If a binary variable x_{ft} is used to denote whether the flight $f \in F$ is assigned to a discrete time interval t, the objective is formulated as $\min \sum_{f \in F} \sum_{t \in T} c_{ft} x_{ft}$, where c_{ft} is a linear cost function with respect to the decision variable x_{ft}. Usually these resulting problems are large-scale integer programs, which require quite sophisticated solution techniques. However, it is noticeable that costs are additive for each flight, i.e., no correlations are modelled.

This is true in the contexts of air and marine transportation systems whose capacities are usually constrained by point capacities, such as airports and ports, while their link capacities are often unconstrained. For other transportation modes, such as rail, highway or public transit, links could easily become capacity bottlenecks. In such cases, vehicle movements through the shared links tend to be correlated.

Even when no arrival/departure correlations are considered in the air traffic management literature, these integer programs modeling ground holding problems are already difficult to solve. Because the consideration of correlations is likely to introduce nonlinearities, which can significantly complicate existing models, it would be much easier to make decisions for each flight independently without considering their correlations in ground holding problems.

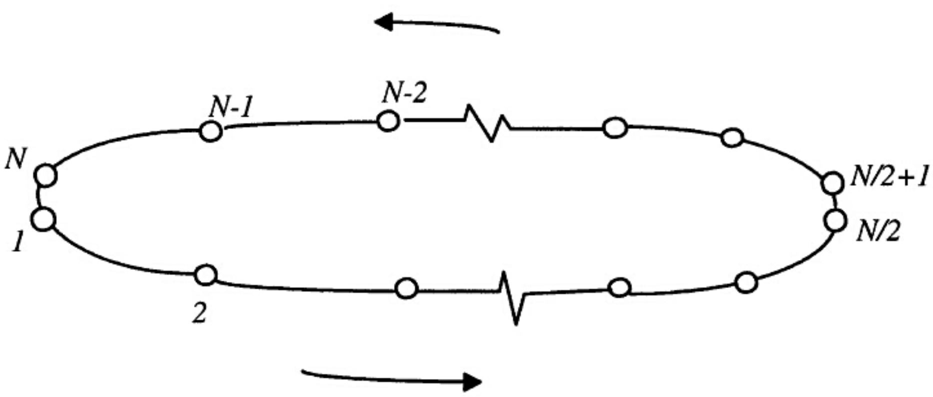
Aircraft holding studies are not reviewed in detail here due to the focus of this study on intermodal freight operations. Interested readers may refer to Bertsimas et al. (2011) for more discussions regarding air traffic flow management.

2.3. Transit Vehicle Holding

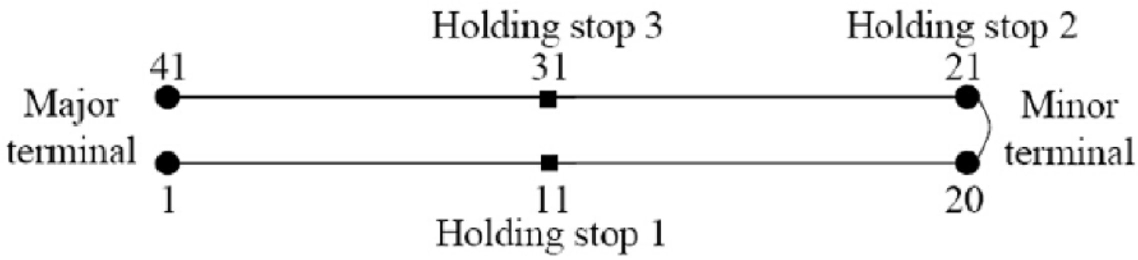
There are two streams of bus holding studies, which are compared in Table 2-1. Figure 2-2 and Figure 2-3 show their clear differences in terms of network structure. The objective of the first group is to reduce the bus bunching effect. In such models, a rolling horizon approach is explicitly considered in Eberlein et al. (2001) to approximate the stochastic problem with a deterministic model. Eberlein et al. (2001) consider the effect of holding on subsequent bus trips and the horizon length is determined to be 3 in their analysis. They also suggest that the number of affected stations depends on scheduled layovers. In a more recent study by Sánchez-Martínez et al. (2016), the rolling horizon approach is also used and the contribution of their work is to account for bus running-time and demand dynamics. Since this study falls into the second group where timed transfers are involved, these relevant studies are reviewed in detail.

Table 2-1 A comparison of two types of bus holding studies

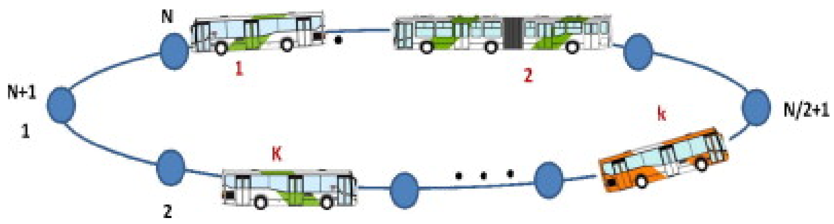
Criteria	Group 1	Group 2
Network Structure	Mostly single route	Multi-route network
Passenger Transfer	Not involved	Timed transfer
Service Frequency	High, e.g., 5 runs/hr, equivalent to a headway of 12 minutes	Low
Optimization Objective	Improve headway regularity; Reduce bus bunching	Improve schedule adherence
Research Methods	Mostly through simulations	Mostly analytical
Sample Studies	Eberlein et al. (2001), Sun and Hickman (2008), Daganzo (2009), Daganzo and Pilachowski (2011), Delgado et al. (2012), Muñoz et al. (2013)	Lee and Schonfeld (1994), Hall et al. (2001), Chowdhury and Chien (2001), Dessouky et al. (2003), Ting and Schonfeld (2007)



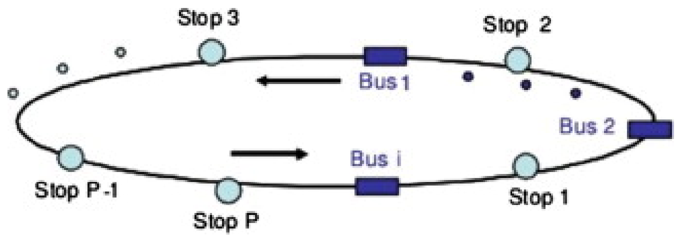
(a) Eberlein et al. (2001)



(b) Sun and Hickman (2008)

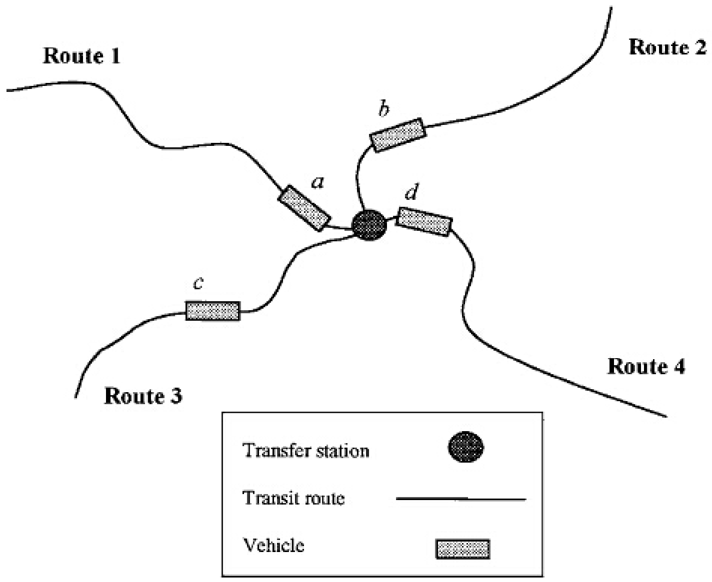


(c) Delgado et al. (2012)

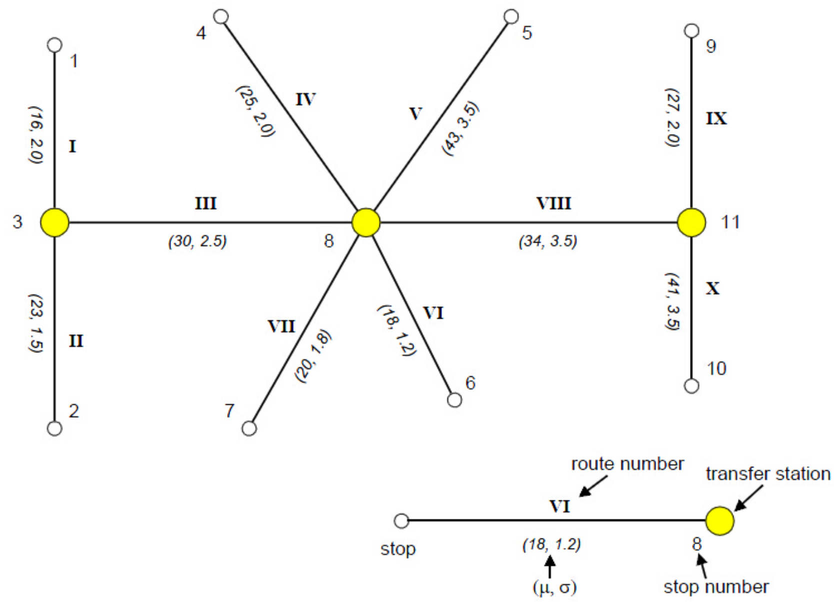


(d) Muñoz et al. (2013)

Figure 2-2 Sample networks used in the first type of transit holding studies



(a) Chowdhury and Chien (2001)



(b) Ting and Schonfeld (2007)

Figure 2-3 Sample networks used in the second group of transit holding studies

In Lee and Schonfeld (1994), the holding time at a transfer terminal is optimized by minimizing the operator's cost of holding the vehicle, onboard passenger waiting and missed connection costs of passengers on delayed incoming vehicles. The implications of holding at downstream transfer stations are not considered.

Dessouky et al. (1999) evaluate the benefits of using bus tracking technologies in making dynamic bus dispatching decisions for timed transfer transit systems. They employ simulation models to compare several holding strategies and find that bus tracking technology can significantly reduce passenger delays when there is a major bus delay and the number of connecting buses is relatively small.

Chowdhury and Chien (2001) study the bus holding problem analytically. A time varying total cost which is a function of the holding time is minimized for a ready vehicle at a transfer station. Only one transfer terminal is considered in a simple network where four routes interconnect at one point.

Hall et al. (2001) study a class of bus dispatching policies to decide whether a bus should be dispatched immediately or held until some criteria are satisfied. They find that at most one non-boundary local minimum of holding time exists when connecting vehicles' arrivals are identically and normally distributed. (Their boundary point is the one where the holding time is zero, i.e., immediate dispatch.) They also recommend that when the forecast lateness is sufficiently large, the vehicle should be dispatched immediately.

Dessouky et al. (2003) further explore the benefit of introducing advanced communication, tracking and passenger counting technologies in making bus dispatching decision for a coordinated system. Through simulation tests, they demonstrate that these technologies can provide considerable benefits, especially when the schedule slack is zero, the service frequency is low and the number of connecting buses is large. It should be noted that in this simulation study, costs at downstream stations due to the vehicle holding at the current station are considered.

Ting and Schonfeld (2007) extend previous analytical bus holding studies (e.g., Lee and Schonfeld 1994, Chowdhury and Chien 2001) by considering a multi-hub transit network where the vehicle holding decision can increase passenger waiting at downstream stations and disrupt coordinated connections at other transfer terminals. Their numerical results show that a hold is justified when the arrival variance of incoming vehicles is small and large transfer volumes are expected on delayed vehicles.

In summary, as an effective measure to improve the transit service reliability, the bus holding problem has been extensively studied, especially to mitigate the bus bunching effect. Another stream of studies which seek to improve successful transfer connections in coordinated transit operations has evolved at least since the 1990s and significant methodological improvements have been achieved. There is no consensus regarding how the “downstream region” which is affected by the holding decision should be determined. In Hall et al. (2001) and Chowdhury and Chien (2001), no downstream stations are considered; Dessouky et al. (2003) seem to consider all downstream stations and Ting and Schonfeld (2007) only consider the next adjacent hub.

2.4. Freight Vehicle Holding

Compared to the extensive studies of vehicle holding for public transit systems, fewer studies are reported for freight and logistics systems. Actually, to some extent, vehicle holding should have better applicability in such systems because there is more complete information about both the user and vehicle. With the adoption of some intelligent transportation system technologies, transit operators might access the real-time location information about vehicles, as well as the number of boarding and alighting passengers at the stop level. However, they are still unlikely to know travelers' destinations unless transit riders register their destinations in advance, as assumed by Ting and Schonfeld (2007). Dessouky et al. (2003) present a method for predicting passenger loads at downstream stops. The forecast of continuing passenger numbers depend on the fraction of passengers remaining on board. However, such a fraction, which could probably be estimated from historical data, is hard to obtain for real-time decisions. When the expected transfer volume from a late vehicle to the held vehicle is hard to estimate or the forecast is unreliable, the effectiveness of holding decision is largely diminished.

In contrast to public passenger transportation systems, goods and freight vehicles are moving in a data-rich environment. Vehicle and shipping tracking is not new to freight and logistic systems. For example, Zhu et al. (2012) provide a comprehensive review of Radio Frequency Identification technology (RFID) for its ability to identify, trace and track information throughout the supply chain. An example is shown in Figure 2-4. Moreover, compared to the passenger transport system, goods are moving in a more passive and centralized manner, which enables the design of effective control measures.

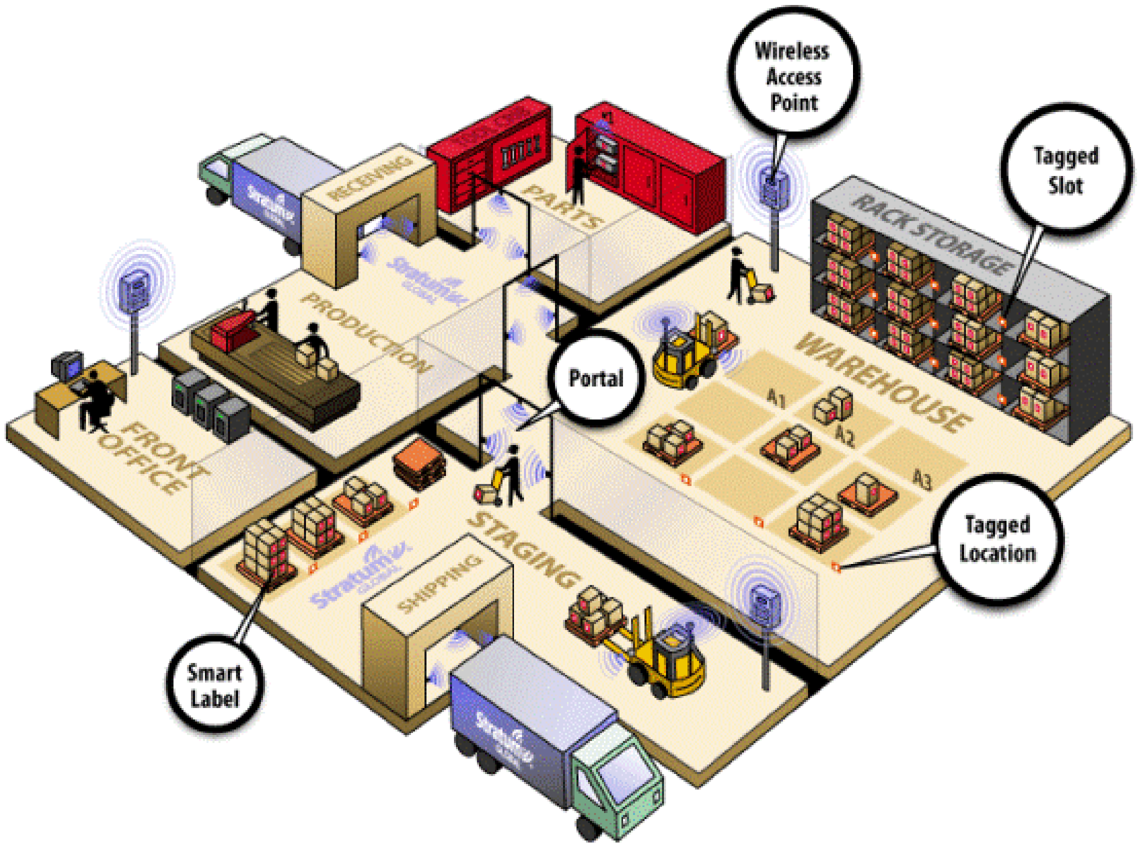


Figure 2-4 An application of the RFID tracking technology

Source: http://www.danbygroup.com/uploads/images/RFID_Factory.gif

The freight vehicle holding problem is largely unexplored in the literature. Since this study is concerned with intermodal transfers, several review articles of intermodal freight transport in the past decade can be cited as evidence. Bontekoning et al. (2004) investigated 92 publications (mainly on the rail-truck mode) in order to identify the characteristics of the intermodal freight research field. Although they mentioned the synchronized schedule between different modes as one major characteristic of the rail-truck freight transport, no relevant studies on either preplanned schedule coordination method or real time vehicle control strategies were found. In

2008, Caris et al. provided an overview of planning models in intermodal freight transport and noted that research interest in this area was growing rapidly. They found that strategic planning problems such as terminal design and infrastructure network configuration had been extensively studied while the number of scientific publications at the operation level remained limited or non-existent. Neither schedule coordination nor vehicle holding was discussed. SteadieSeifi et al. (2014) presented the most recent review of multimodal freight transportation planning from the perspective of operations research. Research efforts from 2005 onward were categorized and reviewed separately at three levels: strategic, tactical and operational. Similarly to Caris et al. (2008), they found that there were remarkably fewer studies on operational planning than on the tactical and strategic planning problems.

Two tactical studies by Andersen et al. (2009) and by Chen and Schonfeld (2010) are noted here. Andersen et al. (2009) explore how the synchronization of multiple collaborating services can reduce the throughput time of the demand in the freight system. By noticing many similarities between public transit and freight system operations, Chen and Schonfeld (2010) present one optimization model for coordinating vehicle movements and cargo transfers at intermodal terminals. They optimize service frequencies and slack times for a multi-hub network and coordinated operations are compared with uncoordinated ones.

For real-time vehicle dispatching control, only Chen and Schonfeld (2011) is found. They develop a vehicle control method for determining whether each ready vehicle should be dispatched immediately or held waiting for delayed connecting vehicles, when the coordinated freight operations are disrupted by random events.

Thus, while there are abundant vehicle holding applications in the public transit area, far fewer studies are devoted to the freight and logistics system even though the holding control might be more applicable and useful in such systems. These methods developed for the public transit operations should be adapted to improve service disruption management in coordinated freight operations.

2.5. Research Needs

(1) Freight Operations

More efforts should be made for improving service disruption management in coordinated intermodal freight operations. Methods developed for public transit systems can be adapted due to the similarities between these two transportation modes. Although there are also numerous aircraft holding studies in the literature, it seems that the aircraft ground holding problem has evolved into a specialized and isolated research topic with very few references to either public transit or freight systems, possibly due to the distinct operating characteristics of air transport.

(2) Correlated Vehicle Arrivals

The impact of correlated vehicle arrivals on the holding decision should be explicitly considered. The modeling of various costs can be greatly simplified when vehicle arrivals are assumed to be independent. However, this assumption is often wrong in reality, especially for public transit and freight systems.

(3) Decision Risks

Existing vehicle holding studies examine only the expected value of the objective, with little attention to the degree of uncertainty in the decision outcomes. In other words, decision makers are unaware of how reliable their decision will be. For risk-neutral operators, the existing methods suffice. However, for a risk-averse operator, an enhanced method considering decision uncertainties is preferred.

(4) Delay Propagation

The implications of delay propagations are not sufficiently addressed. Since the holding decision would affect transfers at downstream hubs, costs at downstream hubs should be included in the analysis. However, it is still unclear how the boundary of the affected region should be defined for making the decision. Is it feasible or justified to consider all downstream hubs even if there are highly efficient optimization methods?

Therefore, this thesis presents a more realistic vehicle holding model for managing service disruptions in intermodal freight systems. The contributions are also in the above four aspects. The foremost one is that the existing assumption of independence among vehicle arrivals is relaxed.

.

3. Model

The proposed model is used to optimize the vehicle holding decision by considering various types of costs, e.g., extra operating cost, additional waiting cost and missed connection cost. Decisions are made whenever a vehicle is ready to be dispatched according to the preset schedules and updated whenever new information, such as a new vehicle arrival prediction, becomes available.

3.1. Assumptions

The following assumptions are provided to facilitate the development of the vehicle holding model.

(1) Vehicle schedules are pre-optimized but vehicle movements might be disrupted, i.e., vehicle arrivals are subject to random delays.

(2) Vehicle arrivals at a certain station can be described with a multivariate probability density function, which can be estimated and thus assumed to be known.

(3) When a scheduled transfer is missed, cargos on the connecting vehicle will be picked up by the next scheduled receiving vehicle on that route. Such an assumption is valid in cases of moderate or routine disruptions. In extreme scenarios, e.g., demand surges, the assumption does not hold since the next receiving vehicle is already full and cannot pick up any left-behind cargos.

(4) The holding decision is conducted independently for each outbound vehicle in the order of dispatch ready time, while neglecting the effect of future decisions. Such an

assumption might be relaxed to consider the interrelation between the current and future decisions.

3.2. Arrival Distribution

When vehicle arrivals are assumed to be independent, each vehicle's arrival is described with a univariate distribution. Although other distributions, theoretical or empirical, are usable, generally two families of distributions are considered: log-normal and normal (Chowdhury and Chien 2001). One cited advantage of log-normal distributions is the absence of long left tails, which allow negative arrival times. Other researchers (Hall et al. 1999, Chowdhury and Chien 2002, Ting and Schonfeld 2007, Chen and Schonfeld 2011) prefer the most widely used general purpose distribution, i.e. the Normal Distribution. Although its left-hand limit is negative infinity, the issue of negative arrival times should be manageable if distribution parameters (i.e., mean and variance) are properly calibrated. In other words, the chance for the arrival time to be negative is negligible in practical analyses. In line with the majority of existing studies, normal distributions are used in this study.

Similarly to the one-dimensional normal distribution used for describing one vehicle's arrival, the multivariate normal distribution is introduced when correlations among multiple vehicle arrivals are considered. An N -dimensional normal distribution is characterized by two set of parameters: (1) a mean vector μ of length N , and (2) a symmetric variance-covariance matrix Σ . The entry in row i and column j of Σ represents the covariance between T_i and T_j . If T is used to denote the N -dimensional vector $\langle T_1, T_2, \cdots, T_N \rangle$, the probability density function of random variable T is:

$$p(T)=\frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp[-\frac{(T-\mu)'\Sigma^{-1}(T-\mu)}{2}] \quad (1)$$

where $|\Sigma|$ is the determinant of Σ. Since we need to find the inverse of Σ, we should ensure Σ is invertible. $(T-\mu)'$ is the transpose of $(T-\mu)$.

3.3. *Cost Classification*

While it is straightforward to compute the operator's cost due to the holding decision, the users' cost is much more complicated to analyze because different types of cargos tend to have various functional forms of costs. As shown in Figure 3-1, for each specific location (or station interchangeably), goods can be categorized into four types: feeder, loading, unloading and receiver. If there are no feeder or receiver routes associated with the location, the station is intermediate; otherwise, the station is also called a transfer terminal or a hub. Feeder cargos are those collected by inbound vehicles and eventually carried by main-haul vehicles to other destinations, while receiver cargos are distributed among outbound vehicles. Loading and unloading cargos are not transferred to either feeder or receiver routes. Loading cargos originate from the station itself and unloading cargos have the station as the final destination. Since loading/unloading cargos are not involved in coordinated operations, they have no missed connection costs. Both feeder and receiver goods suffer from such losses when scheduled connections are lost.

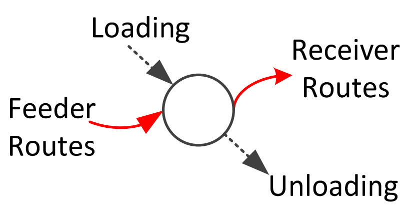


Figure 3-1 Types of cargos associated with a location

For a specific ready vehicle at a location, all stations on the study route can be divided into: upstream, current and downstream ones. All cargos associated with upstream stations, regardless of the type, are already reflected in the on-board cargos. When the study vehicle is ready for dispatch according to the pre-determined schedule, it is implicitly assumed that all loading and unloading operations have been completed and receiver cargos have also been transshipped. Therefore, all cargos except for those from feeder routes are also included in on-board cargos. Goods on these delayed feeder routes can suffer from either extra waiting if finally they are able to make the connection during the vehicle holding or missed connection cost if the ready outbound vehicle has been released by the time they arrive at the current location. If a transfer connection is missed, left-over cargos are assumed to be picked up by the next scheduled vehicle on that route. At downstream hubs, cargos from feeder routes might wait if feeder vehicles arrive before the held vehicle's arrival and miss the held vehicle otherwise. Similarly, at downstream hubs, transshipments from the held vehicle to receiver routes can have either additional waiting time or a missed connection cost. For these cargos which are to be loaded from the downstream station to the held vehicle, there is only waiting cost because the held vehicle will arrive after the ready time of these cargos. For unloading cargos from the held vehicle to a downstream station, no cost is incurred at these locations because extra waiting costs (if any) have been calculated when loaded at the downstream station's preceding stations.

Costs borne by various types of cargos are specified in Table 3-1.

Table 3-1 Costs for each cargo category

	Upstream Stations	Current Location	Downstream Station j
Feeder Routes		Wait/Miss C^f	Wait/Miss D_j^f
Loading		On-board Cargos Wait C^o	Wait D_j^l
Unloading			-
Receiver Routers			Wait/Miss D_j^r

The costs at the current hub and downstream hubs are analyzed separately below.

3.4. Costs at the Current Hub

3.4.1. Extra operating cost E_o

Extra operating cost is defined as the extra operators' cost of holding the ready vehicle a certain amount of time h. If the unit operating cost per vehicle is denoted as b, the extra operating cost is simply written as:

$$E_o = bh \tag{2}$$

3.4.2. On-board cargos C^o

The cargos on the subject vehicle have the following waiting cost:

$$C^o = vhQ \tag{3}$$

where v is time value of cargos and Q is the amount of cargos on board the subject vehicle.

3.4.3. Cargos on delayed feeder vehicles C^f

Costs for cargos on late feeder vehicles are analyzed in following three scenarios.

One-vehicle scenario

To make the model analytically tractable, it is assumed $\int_{-\infty}^{0} f_T(t)dt = 0$ and $\int_{H}^{+\infty} f_T(t)dt = 0$,

meaning that a feeder vehicle can only arrive within the range $(0, H]$, where $f_T(t)$ is the marginal distribution of a vehicle's arrival time T and H is the headway of the held route. For notational brevity, one might omit the T notation and use $f(t)$ instead of $f_T(t)$. The time when the decision must be made is defined as $t = 0$.

The waiting time faced by the feeder vehicle is also a random variable, whose value depends on whether the vehicle can arrive before the departure of the held vehicle, i.e., (1) making the connection or (2) missing it. If the random waiting is denoted as Y, we have:

$$Y = \begin{cases} Y^{(1)} = h - T, 0 < T \le h \\ Y^{(2)} = H - T, h < T \le H \end{cases} \tag{4}$$

Implicitly, the first dispatch policy as described in Hall et al. (2001) is used here. The vehicle is held until the dispatch time even the vehicle arrives prior to the optimal dispatch time. When T falls into the region $(0, h]$, the connection is made; otherwise, the cargos on this late vehicle will be carried by the next vehicle departing at H. The probability density of the arrival time T given $0 < T \le h$ is:

$$f(t \mid 0 < T \le h) = \frac{g^{(1)}(t)}{F(h) - F(0)} \tag{5}$$

where $F(t)$ is the cumulative distribution function of T, $g^{(1)}(t) = f(t)$ for all $0 < t \le h$, and $g^{(1)}(t) = 0$ everywhere else. Note $F(0) = 0$, meaning the vehicle cannot arrive earlier than 0. Such a conditional distribution obtained when a random variable is bounded below or above is also called a truncated distribution. Since the extra waiting time $h-T$ is a linear function of T, its probability density function is given by:

$$f^{(1)}(y)=\frac{g^{(1)}(h-y)}{F(h)} \tag{6}$$

Similarly, when $h<T\leq H$, the missed connection cost $H-T$ has the following probability density:

$$f^{(2)}(y)=\frac{g^{(2)}(H-y)}{1-F(h)} \tag{7}$$

where $g^{(2)}(t)=f(t)$ for all $h<t\leq H$, and $g^{(2)}(t)=0$ everywhere else.

Combing two cases, the probability density of Y is as follows:

$$f(y)=F(h)f^{(1)}(y)+(1-F(h))f^{(2)}(y) \tag{8}$$

which is a mixture distribution, where $F(h)$ and $(1-F(h))$ are mixture weights.

An example of the arrival distribution and a waiting time distribution are shown in Figure 3-2. In this example, the mean of the arrival time is 6 and the standard deviation is 1. The holding time $h=7$ and the headway of the held route $H=20$.

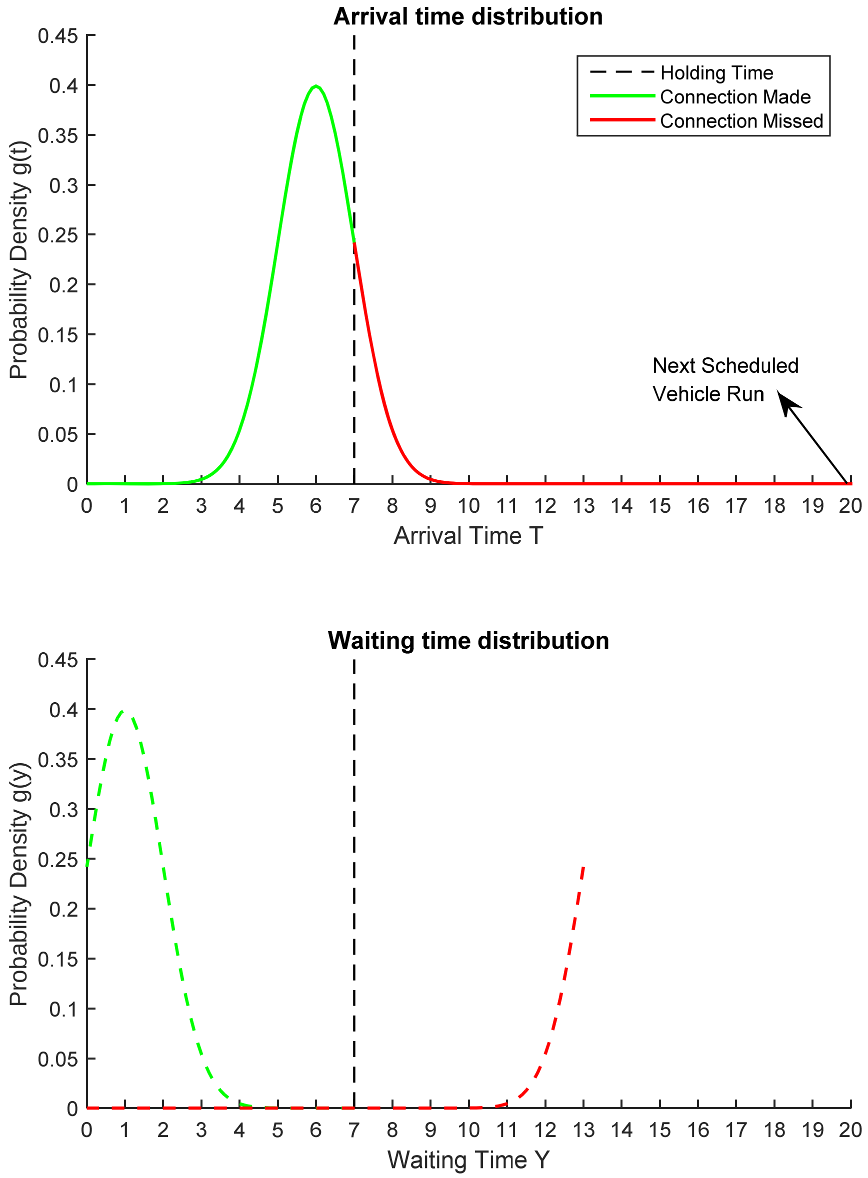


Figure 3-2 Arrival and waiting time distributions – one vehicle case

Proposition 1: The mean and variance of waiting time Y can be calculated as follows:

$$\mu_Y = E(Y) = \Phi(\beta)h + (1-\Phi(\beta))H - \mu \tag{9}$$

$$\sigma_Y^2 = Var(Y) = \sigma^2 + \Phi(\beta)(1-\Phi(\beta))(h-H)^2 + 2\sigma\phi(\beta)(h-H) \tag{10}$$

where μ is the mean of T, σ is the standard deviation of T, $\beta = (h-\mu)/\sigma$, $\phi(\bullet)$ is the probability density function of the standard normal distribution and $\Phi(\bullet)$ is its cumulative density.

Proof:

The mean and variance of the truncated normal distribution (Greene 2012) are as follows:

$$E(T \mid 0 < T \le h) = \mu - \sigma \frac{\phi(\beta)}{\Phi(\beta)} \tag{11}$$

$$Var(T \mid 0 < T \le h) = \sigma^2 (1 - \beta \frac{\phi(\beta)}{\Phi(\beta)} - (\frac{\phi(\beta)}{\Phi(\beta)})^2) \tag{12}$$

$$E(T \mid h < T \le H) = \mu + \sigma \frac{\phi(\beta)}{1-\Phi(\beta)} \tag{13}$$

$$Var(T \mid h < T \le H) = \sigma^2 (1 + \beta \frac{\phi(\beta)}{1-\Phi(\beta)} - (\frac{\phi(\beta)}{1-\Phi(\beta)})^2) \tag{14}$$

Because $Y^{(1)}, Y^{(2)}$ are linear with respect to T, we can obtain their means and variances.

$$\mu_{Y^{(1)}} = E(Y^{(1)}) = h - \mu + \sigma \frac{\phi(\beta)}{\Phi(\beta)} \tag{15}$$

$$\sigma^2_{Y^{(1)}} = Var(Y^{(1)}) = \sigma^2 (1 - \beta \frac{\phi(\beta)}{\Phi(\beta)} - (\frac{\phi(\beta)}{\Phi(\beta)})^2) \tag{16}$$

$$\mu_{Y^{(2)}} = E(Y^{(2)}) = H - \mu - \sigma \frac{\phi(\beta)}{1-\Phi(\beta)} \tag{17}$$

$$\sigma^2_{Y^{(2)}} = Var(Y^{(2)}) = \sigma^2 (1 + \beta \frac{\phi(\beta)}{1-\Phi(\beta)} - (\frac{\phi(\beta)}{1-\Phi(\beta)})^2) \tag{18}$$

Since Y is obtained by mixing $Y^{(1)}$ and $Y^{(2)}$ with weights $F(h)$ and $(1-F(h))$, it is obtained that:

$$\mu_Y = E(Y) = \Phi(\beta)h + (1-\Phi(\beta))H - \mu \tag{19}$$

by noting $\Phi(\beta) = F(h)$.

While the mean of the mixture distribution is simply the weighted average of the conditional mean, the unconditional variance of Y is more complex to calculate.

According to the law of total variance, we have the following equation:

$$Var(Y) = E[Var(Y|T)] + Var[E(Y|T)] \tag{20}$$

In addition to the weighted average of the conditional variances, i.e., $E[Var(Y|T)]$, the second item $Var[E(Y|T)]$, the variance of the conditional mean, also appears as part of the unconditional variance of Y. The uncertainty in the parameter variable T thus has the effect of increasing the unconditional variance of the mixture Y. The variance of Y is then given by:

$$\sigma_Y^2 = \Phi(\beta)(\sigma_{Y(1)}^2 + (\mu_{Y(1)} - \mu_Y)^2) + (1-\Phi(\beta))(\sigma_{Y(2)}^2 + (\mu_{Y(2)} - \mu_Y)^2) \tag{21}$$

Simplifying Equation (21) after substituting the left-hand sides of Equations (15-19) with their respective right-hand sides, we obtain:

$$\sigma_Y^2 = \sigma^2 + \Phi(\beta)(1-\Phi(\beta))(h-H)^2 + 2\sigma\phi(\beta)(h-H) \tag{22}$$

completing the proof for Proposition 1.

Remark 1: The variance of Y is no less than the variance of T, if parameters of the arrival time distribution are properly calibrated.

Figure 3-3 plots the differences of $Var(Y)$ and $Var(T)$, i.e., $\sigma_Y^2 - \sigma^2$, with respect to the holding time h. It shows that $\sigma_Y^2 \geq \sigma^2$, especially when the holding time h is close to the expected arrival time μ, if parameters of the arrival distribution are properly calibrated according to the stated assumptions, e.g., $\int_{-\infty}^{0} f(t)dt = 0$ or $F(0)=0$. When $\sigma = 2.5$, we can observe negative values of $\sigma_Y^2 - \sigma^2$. However, in this case, $\sigma_Y^2 - \sigma^2 \neq 0$ when $h = 0$, violating $F(0)=0$. When $h = 0$, i.e., the subject vehicle is released immediately, the incoming vehicle is unable to make the connection, thus facing only one possibility of $Y = H - T$. Clearly, Y should have the same variance as T , i.e., $\sigma_Y^2 - \sigma^2 = 0$ when $h = 0$. Therefore, $\sigma = 2.5$ is considered improper, because the chance of negative arrival times is no longer negligible in this example.

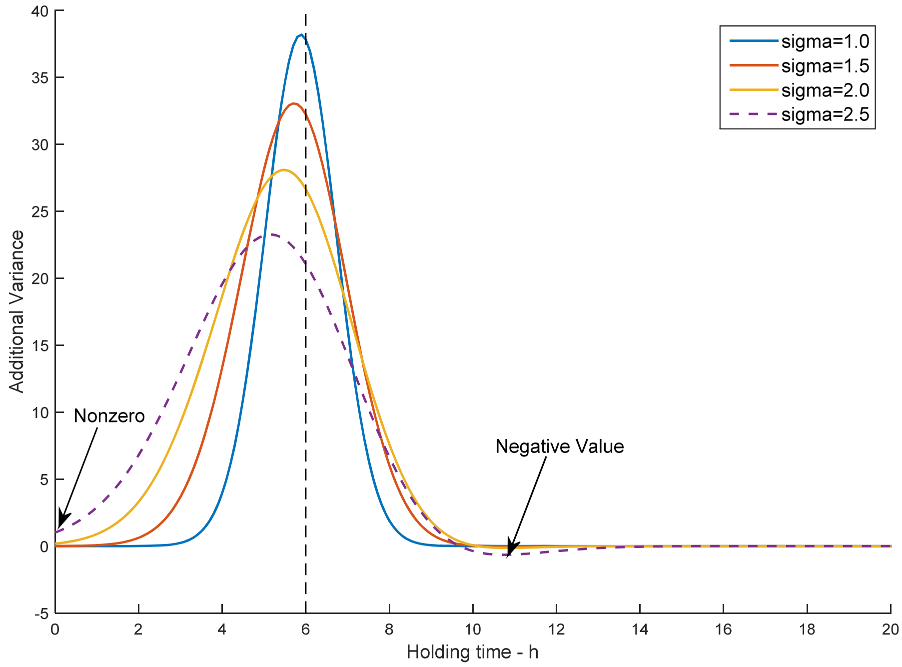


Figure 3-3 Difference of the waiting variance and arrival variance

Because $\sigma_Y^2 \geq \sigma^2$, especially when the holding time is close to the expected arrival time (which is likely to be the optimal holding time), the variance of Y cannot be neglected in the assessment of a holding decision h. It is quite limiting to examine only the expected value of a decision while ignoring its variance, a typical measure of decision risks. For example, in Figure 3-3, $\sigma_Y^2 = 17.03$ when the holding time is 7. Note that the variance of the arrival time σ^2 is only 1.0. The variance of output Y is greatly enlarged due to the mixture of two conditional distributions of input T. Traditional modeling methods focusing only on the expected value are insufficient because (1) the high variance in the output is ignored and (2) the probability of realizing the expectation is quite low, as illustrated in Figure 3-2.

Two-vehicle scenario

Suppose that two inbound vehicles are delayed and their arrivals are correlated. The joint arrival distribution is $f(t_1, t_2)$, where t_1 is the estimated arrival time for vehicle 1 and t_2 is for vehicle 2. Depending on whether the vehicle can make the connection ("Yes" means the connection is successful and "No" represents the opposite side), we have the following four cases in Table 3-2.

Table 3-2 Possible outcomes for a case of two late arrivals

Case	Associated Probability	Vehicle 1	Vehicle 2
		Time (wait/miss)	Time (wait/miss)
(Yes, Yes)	$p^{(1)} = \int_0^h \int_0^h f(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2$	$Y_1^{(1)} = h - t_1, 0 < t_1 \le h$	$Y_2^{(1)} = h - t_2, 0 < t_2 \le h$
(Yes, No)	$p^{(2)} = \int_h^H \int_0^h f(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2$	$Y_1^{(2)} = h - t_1, 0 < t_1 \le h$	$Y_2^{(2)} = H - t_2, h < t_2 \le H$
(No, Yes)	$p^{(3)} = \int_0^h \int_h^H f(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2$	$Y_1^{(3)} = H - t_1, h < t_1 \le H$	$Y_2^{(3)} = h - t_2, 0 < t_2 \le h$

(No, No)	$p^{(4)} = \int_h^H \int_h^H f(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2$	$Y_1^{(4)} = H - t_1,\ h < t_1 \le H$	$Y_2^{(4)} = H - t_2,\ h < t_2 \le H$

The joint probability density of the arrival time given that both vehicles make the connection, i.e., case (Yes, Yes), is:

$$f([t_1, t_2] \mid 0 < t_1 \le h \,\&\, 0 < t_2 \le h\ ) = \frac{g^{(1)}(t_1, t_2)}{\int_0^h \int_0^h f(t_1, t_2)\mathrm{d}t_1\,\mathrm{d}t_2} \tag{23}$$

where $g^{(1)}(t_1, t_2) = f(t_1, t_2)$ for all $0 < t_1 \le h$ and $0 < t_2 \le h$; $g^{(1)}(t_1, t_2) = 0$, otherwise. Then, the joint distribution of $[y_1^{(1)}, y_2^{(1)}]$ is:

$$f^{(1)}([y_1^{(1)}, y_2^{(1)}]) = \frac{g^{(1)}(h - y_1, h - y_2)}{\int_0^h \int_0^h f(t_1, t_2)\mathrm{d}t_1\,\mathrm{d}t_2} \tag{24}$$

In this way, the distributions of $[y_1^{(k)}, y_2^{(k)}]$ in any case $k \in \{1, 2, 3, 4\}$ can be obtained. By mixing the four exclusive cases (noting that $\int_{-\infty}^0 \int_{-\infty}^0 f(t_1, t_2)\mathrm{d}t_1\,\mathrm{d}t_2 = 0$ and $\int_0^{+\infty} \int_0^{+\infty} f(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 = 0$), we obtain the unconditional distribution of the waiting time vector $[y_2, y_2]$ as follows:

$$\begin{aligned} f([y_1, y_2]) = p^{(1)} * f^{(1)}(h - y_1, h - y_2) + p^{(2)} * f^{(2)}(h - y_1, H - y_2) + \\ p^{(3)} * f^{(3)}(H - y_1, h - y_2) + p^{(4)} * f^{(4)}(H - y_1, H - y_2) \end{aligned} \tag{25}$$

The mixture density $f([y_1, y_2])$ is plotted in Figure 3-4. Clearly, there are four mixture components, corresponding to those four cases described in Table 3-2. In Figure 3-2, there are only two components, because one vehicle is considered.

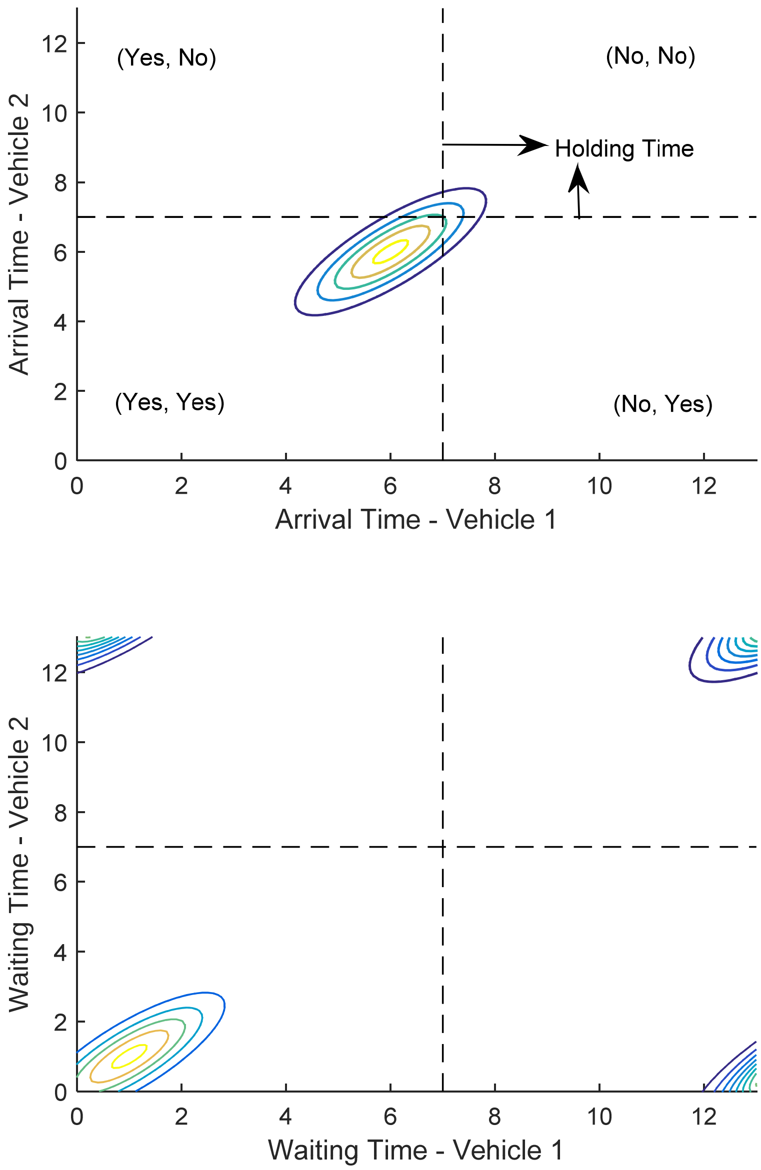


Figure 3-4 Probability densities of the arrival time and waiting time – two vehicles

Proposition 2: The expectation of $Y_i, i \in \{1,2\}$ is given by:

$$\mu_{Y_i} = E(Y_i) = \Phi(\beta_i)h + (1-\Phi(\beta_i))H - \mu_i, i \in \{1,2\} \quad (26)$$

where μ_i is the mean of the marginal distribution $f(t_i)$, σ_i is the standard deviation of $f(t_i)$, $\beta_i = (h - \mu_i)/\sigma_i$. Recall that $\phi(\bullet)$ is the probability density function of the standard normal distribution and $\Phi(\bullet)$ is its cumulative density.

Proof:

Only the case $i=1$ is proved while the logic applies also when $i=2$. According to the definition of the conditional mean, the conditional means of vehicle 1's arrival time are as follows:

$$\begin{aligned}
\mu_1^{(1)} &= \int_0^h\int_0^h t_1 \frac{g^{(1)}(t_1,t_2)}{p^{(1)}} \mathrm{d}t_1\,\mathrm{d}t_2,\ \mu_1^{(2)} = \int_h^H\int_0^h t_1 \frac{g^{(2)}(t_1,t_2)}{p^{(2)}} \mathrm{d}t_1\,\mathrm{d}t_2 \\
\mu_1^{(3)} &= \int_0^h\int_h^H t_1 \frac{g^{(3)}(t_1,t_2)}{p^{(3)}} \mathrm{d}t_1\,\mathrm{d}t_2,\ \mu_1^{(4)} = \int_h^H\int_h^H t_1 \frac{g^{(4)}(t_1,t_2)}{p^{(4)}} \mathrm{d}t_1\,\mathrm{d}t_2
\end{aligned} \tag{27}$$

The expectation of vehicle 1's waiting time $E(Y_1)$ can be written as follows:

$$\begin{aligned}
E(Y_1) = {} & p^{(1)} * (h - \mu_1^{(1)}) + p^{(2)} * (h - \mu_1^{(2)}) + \\
& p^{(3)} * (H - \mu_1^{(3)}) + p^{(4)} * (H - \mu_1^{(4)})
\end{aligned} \tag{28}$$

Simplifying the right-hand side of Equation (28) by noting definitions of $p^{(\cdot)}$ and $g^{(\cdot)}$, one can obtain:

$$\begin{aligned}
E(Y_1) = {} & \left(\int_0^h\int_0^h f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2 + \int_h^H\int_0^h f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2\right)h + \\
& \left(\int_0^h\int_h^H f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2 + \int_h^H\int_h^H f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2\right)H - \\
& \left(\int_0^h\int_0^h t_1 f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2 + \int_h^H\int_0^h t_1 f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2 + \right. \\
& \left. \int_0^h\int_h^H t_1 f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2 + \int_h^H\int_h^H t_1 f(t_1,t_2)\mathrm{d}t_1\,\mathrm{d}t_2\right)
\end{aligned} \tag{29}$$

Then, the variable T_2 is marginalized out.

$$E(Y_1) = h\int_0^h f(t_1)\,\mathrm{d}t_1 + H\int_h^H f(t_1)\,\mathrm{d}t_1 - \int_0^H t_1 f(t_1)\,\mathrm{d}t_1 \tag{30}$$

By definitions of $\Phi(\beta_i)$ and μ_i, $i \in \{1,2\}$, we obtain:

$$\mu_{Y_1} = \Phi(\beta_1)h + (1-\Phi(\beta_1))H - \mu_1 \tag{31}$$

Thus the proof for Proposition 2 is complete.

Remark 2: A vehicle's expected waiting time depends on the marginal distribution of its arrival time rather than the joint arrival distribution.

In other words, the expectation of a vehicle's waiting time does not depend on the covariance of the joint distribution.

The total waiting time $Y = Y_1 + Y_2$ is studied. Although Y_1 and Y_2 are correlated, the mean of Y is given by:

$$\mu_Y = \mu_{Y_1} + \mu_{Y_2} \tag{32}$$

The total waiting time is not affected by the correlations of vehicle arrivals.

In computing the variance of Y, the correlations between Y_1 and Y_2 are included.

Proposition 3: $Var(Y)$ is given by:

$$\begin{aligned}\sigma_Y^2 = Var(Y) = {} & p^{(1)} * (\sigma_{Y^{(1)}}^2 + (h-\mu_1^{(1)} + h - \mu_2^{(1)} - \mu_Y)^2) + \\ & p^{(2)} * (\sigma_{Y^{(2)}}^2 + (h-\mu_1^{(2)} + H - \mu_2^{(2)} - \mu_Y)^2) + \\ & p^{(3)} * (\sigma_{Y^{(3)}}^2 + (H-\mu_1^{(3)} + h - \mu_2^{(3)} - \mu_Y)^2) + \\ & p^{(4)} * (\sigma_{Y^{(4)}}^2 + (H-\mu_1^{(4)} + H - \mu_2^{(4)} - \mu_Y)^2)\end{aligned} \tag{33}$$

where $\sigma_{Y^{(k)}}^2 = Var(Y_1^{(k)}) + Var(Y_2^{(k)}) + 2Cov(Y_1^{(k)}, Y_2^{(k)}),\ k \in \{1,2,3,4\}$.

Note that in any case $k \in \{1,2,3,4\}$, the arrival time $T_i^{(k)}$ and waiting time $Y_i^{(k)}$ of vehicle $i \in \{1,2\}$ have the same variance. Then, the conditional variances $Var(Y_1^{(k)})$, $Var(Y_2^{(k)})$ and covariances $Cov(Y_1^{(k)}, Y_2^{(k)})$ in each case $k \in \{1,2,3,4\}$ are provided in Table 3-3.

Table 3-3 Conditional variances and co-variances – two vehicles

k	$Var(Y_1^{(k)})$	$Var(Y_2^{(k)})$	$Cov(Y_1^{(k)}, Y_2^{(k)})$
1	$\int_0^h \int_0^h (t_1-\mu_1^{(1)})^2 \frac{g^{(1)}(\bullet)}{p^{(1)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_0^h \int_0^h (t_2-\mu_2^{(1)})^2 \frac{g^{(1)}(\bullet)}{p^{(1)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_0^h \int_0^h (t_1-\mu_1^{(1)})(t_2-\mu_2^{(1)}) \frac{g^{(1)}(\bullet)}{p^{(1)}} \mathrm{d}t_1 \mathrm{d}t_2$
2	$\int_h^H \int_0^h (t_1-\mu_1^{(2)})^2 \frac{g^{(2)}(\bullet)}{p^{(2)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_h^H \int_0^h (t_2-\mu_2^{(2)})^2 \frac{g^{(2)}(\bullet)}{p^{(2)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_h^H \int_0^h (t_1-\mu_1^{(2)})(t_2-\mu_2^{(2)}) \frac{g^{(2)}(\bullet)}{p^{(2)}} \mathrm{d}t_1 \mathrm{d}t_2$
3	$\int_0^h \int_h^H (t_1-\mu_1^{(3)})^2 \frac{g^{(3)}(\bullet)}{p^{(3)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_0^h \int_h^H (t_2-\mu_2^{(3)})^2 \frac{g^{(3)}(\bullet)}{p^{(3)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_0^h \int_h^H (t_1-\mu_1^{(3)})(t_2-\mu_2^{(3)}) \frac{g^{(3)}(\bullet)}{p^{(3)}} \mathrm{d}t_1 \mathrm{d}t_2$
4	$\int_h^H \int_h^H (t_1-\mu_1^{(4)})^2 \frac{g^{(4)}(\bullet)}{p^{(4)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_h^H \int_h^H (t_2-\mu_2^{(4)})^2 \frac{g^{(4)}(\bullet)}{p^{(4)}} \mathrm{d}t_1 \mathrm{d}t_2$	$\int_h^H \int_h^H (t_1-\mu_1^{(4)})(t_2-\mu_2^{(4)}) \frac{g^{(4)}(\bullet)}{p^{(4)}} \mathrm{d}t_1 \mathrm{d}t_2$

Multi-vehicle scenario

The analyses can be extended to a multi-vehicle scenario where the set of delayed feeder vehicles is F and the set of all possible cases is K_F. We have $|K_F| = 2^{|F|}$ since each vehicle $i \in F$ faces 2 outcomes, i.e., connection made or missed.

The total waiting time is defined as $Y = \sum_{i \in F} Y_i$. Then its mean and variance are given by:

$$\mu_Y = E(Y) = \sum_{i \in F} (\Phi(\beta_i)h + (1-\Phi(\beta_i))H - \mu_i) \qquad (34)$$

$$\sigma_Y^2 = Var(Y) = \sum_{k \in K_F} (p^{(k)} * (\sigma_{Y^{(k)}}^2 + (\sum_{i \in F} (\pi_i^{(k)} - \mu_i^{(k)}) - \mu_Y)^2)) \qquad (35)$$

where $\sigma^2_{Y^{(k)}} = \sum_{i\in F}\sum_{j\in F} Cov(Y_i^{(k)}, Y_j^{(k)}),\, k \in K_F$ and $\pi_i^{(k)} = h$ if the vehicle i makes the connection in case k; otherwise, $\pi_i^{(k)} = H$. Other symbols have been defined in previous texts.

The total waiting cost of cargos on delayed connecting vehicles is:

$$C^f = \sum_{i\in F} vq_i Y_i \tag{36}$$

where v is the value of time and q_i is the amount of cargos on vehicle i. The expectation and variance of C^f are given by:

$$\mu_{C^f} = E(C^f) = \sum_{i\in F}(vq_i(\Phi(\beta_i)h + (1-\Phi(\beta_i))H - \mu_i)) \tag{37}$$

$$\sigma^2_{C^f} = Var(C^f) = \sum_{k\in K_F}(p^{(k)}(\sigma^2_{C^f,(k)} + (\sum_{i\in F}(vq_i(\pi_i^{(k)} - \mu_i^{(k)})) - \mu_{C^f})^2)) \tag{38}$$

where $\sigma^2_{C^f,(k)} = v^2 \sum_{i\in F}\sum_{j\in F} q_i q_j Cov(Y_i^{(k)}, Y_j^{(k)}),\, k \in K_F$.

3.5. Costs at Downstream Stations

The impact of vehicle holding spreads from the current location to downstream hubs and coordinated transfer operations at these downstream hubs might be disrupted due to the late arrival of the subject vehicle. Theoretically, one disruption can propagate to the ultimate boundary of a well-coordinated network, regardless of the network's scale. In practice, such a consideration can easily render existing methods useless due to the size of impact region, especially when one needs to evaluate the decision in a real-time manner. It is argued that the impact region should be refined due to the reason described below, even if very efficient algorithms are already developed.

Supposing that it takes 30 minutes for the vehicle holding decision at the current location to affect vehicle movements in another region (referred to as Region 1) and it is clearly known that updated forecasts of feeder arrivals are available in the next 5 minutes, should we consider Region 1 at the current decision time (Time 0)? This seems unadvisable because the holding decision should be reevaluated with the new arrival information (probably around Time 5) and the consideration of Region 1 at Time 0 is thus futile. Intuitively, we should anticipate and consider affected region, but not excessively. Similarly, it does not make much sense to consider the delayed feeder vehicles in New Haven when the train is still held at the Baltimore intermodal station although cargos from the delayed vehicles are scheduled to be transferred to the train. Probably by the time this train arrives at Philadelphia, the feeder vehicle already gets back on schedule. Therefore, a guideline can be proposed for determining the size of impact region.

Guideline 1

If it takes more time for the delay to propagate to a downstream hub than to receive updated forecasting information, the downstream hub may be removed from the impact region.

Assume the impact region is denoted as J (a set of affected stations in this study). Only a downstream station $j \in J$ is analyzed in the following discussion.

3.5.1. Cargos on feeder routes D_j^f

If a feeder vehicle $i \in F_j$ with the arrival time t_i, arrives before the arrival time of the held vehicle, i.e., $t_i < s_j + h$, the extra waiting time is $s_j + h - t_i$; if the feeder vehicle arrives after $s_j + h$, the missed connection waiting is $s_j + H - t_i$. F_j is the set of feeder vehicles to j and s_j is the link travel time from the current station to station $j \in J$.

Similarly, the mean and variance of the waiting cost D_j^f can be obtained as follows:

$$\mu_{D_j^f} = E(D_j^f) = \sum_{i \in F_j} (vq_i(\Phi(\alpha_i)h + (1-\Phi(\alpha_i))H + s_j - \mu_i)), j \in J \tag{39}$$

$$\sigma^2_{D_j^f} = Var(D_j^f) = \sum_{k \in K_{F_j}} (p^{(k)}(\sigma^2_{D_j^{f,(k)}} + (\sum_{i \in F_j} (vq_i(\pi_i^{(k)} + s_j - \mu_i^{(k)})) - \mu_{D_j^f})^2)), j \in J \tag{40}$$

where $\alpha_i = (h + s_j - \mu_i) / \sigma_i$, $i \in F_j$.

3.5.2. Loading cargos D_j^l

For loading cargos L_j, the waiting cost is linear with respect to the holding time:

$$D_j^l = vhL_j, \forall j \in J \tag{41}$$

3.5.3. Cargos to be received D_j^r

For simplicity, correlations between departing vehicles are not modeled in this study. Departures might be correlated due to capacity limits in departing resources, e.g., runways or rail tracks. For transferring cargos on the subject vehicle to be picked up by a receiver vehicle $i \in R_j$ with the departure time t_i, if the held vehicle arrives after the departure time of the receiver vehicle i, i.e., $t_i \le s_j + h$, the missed connection waiting is $T_i^r - s_j - h$, where T_i^r is the departure time of the next scheduled receiver run on route i; if the held vehicle arrives before t_i, the extra waiting is $t_i - s_j - h$. Figure 3-5 shows an example where parameters are specified as follows:

$s_j = 40, h = 5, T_i^r = 60, \mu_i = 46, \sigma_i = 1.$

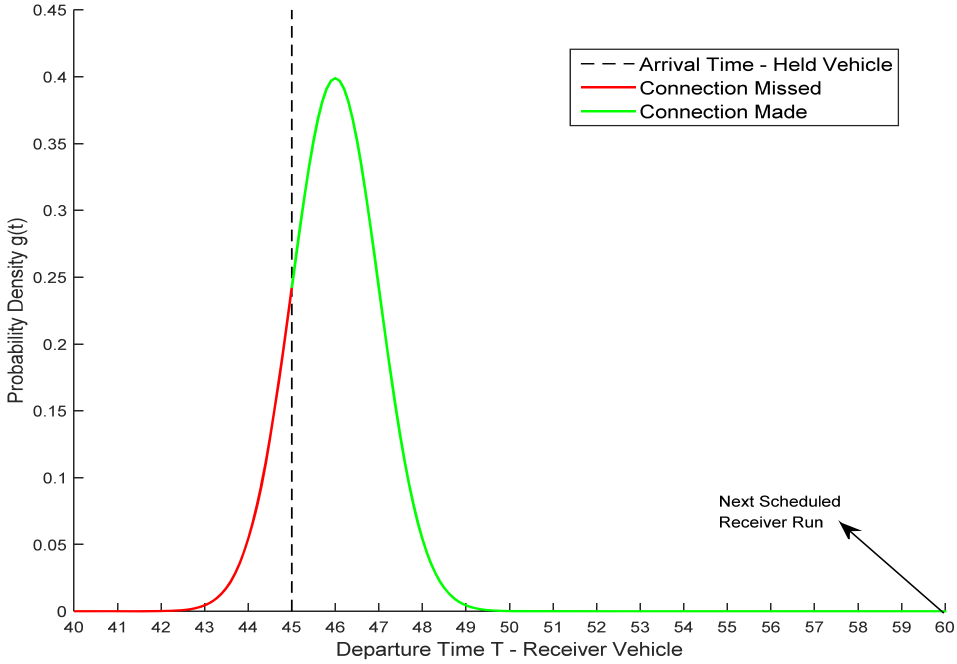


Figure 3-5 Departure time distribution of a receiver vehicle

Denoting the waiting time of these transferring cargos to be received by vehicle i as $Z_i, i \in R_j$, we have:

$$Z_i = \begin{cases} Z_i^{(1)} = T_i^r - s_j - h, & s_j < t_i \le s_j + h \\ Z_i^{(2)} = t_i - s_j - h, & s_j + h < t_i \le T_i^r \end{cases} \tag{42}$$

Conditional means and variances of $Z_i, i \in R_j$ can be derived as follows:

$$\mu_{Z_i^{(1)}} = E(Z_i^{(1)}) = T_i^r - s_j - h \tag{43}$$

$$\sigma_{Z_i^{(1)}}^2 = Var(Z_i^{(1)}) = 0 \tag{44}$$

$$\mu_{Z_i^{(2)}} = E(Z_i^{(2)}) = \mu_i + \sigma_i \frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)} - s_j - h \tag{45}$$

$$\sigma^2_{Z_i^{(2)}} = Var(Z_i^{(2)}) = \sigma_i^2(1+\alpha_i \frac{\phi(\alpha_i)}{1-\Phi(\alpha_i)} - (\frac{\phi(\alpha_i)}{1-\Phi(\alpha_i)})^2) \tag{46}$$

where $\alpha_i = (h + s_j - \mu_i)/\sigma_i$, $i \in R_j$. Note that when the connection is missed, the waiting time no longer depends on the departure time and it is a constant having a zero variance.

The unconditional mean and variance of $Z_i, i \in R_j$ are thus given by:

$$\mu_{Z_i} = E(Z_i) = \Phi(\alpha_i)\mu_{Z_i^{(1)}} + (1-\Phi(\alpha_i))\mu_{Z_i^{(2)}} \tag{47}$$

$$\sigma^2_{Z_i} = Var(Z_i) = \Phi(\beta)(\sigma^2_{Z_i^{(1)}} + (\mu_{Z_i^{(1)}} - \mu_{Z_i})^2) + (1-\Phi(\beta))(\sigma^2_{Z_i^{(2)}} + (\mu_{Z_i^{(2)}} - \mu_{Z_i})^2) \tag{48}$$

The total waiting cost of transferring cargos to be picked up by the set of receiver vehicles R_j is:

$$D_j^r = \sum_{i \in R_j} v q_i^r Z_i \tag{49}$$

where q_i^r is the cargo volume to be picked up by vehicle i, $i \in R_j$.

Due to the assumed independent departures, D_j^r has the following mean and variance:

$$\mu_{D_j^r} = \sum_{i \in R_j} v q_i^r \mu_{Z_i} \tag{50}$$

$$\sigma^2_{D_j^r} = \sum_{i \in R_j} ((v q_i^r)^2 \sigma^2_{Z_i}) \tag{51}$$

3.6. *Mean-standard deviation model*

When costs at the current station are considered, the total cost ξ_1 is defined as:

$$\xi_1 = E_o + C^o + C^f \tag{52}$$

where E_o is the extra operating cost, C^o is the on-board waiting cost and C^f is the waiting cost by cargos on delayed inbound vehicles. As discussed above, E_o and C^o are deterministic and C^f is random.

When downstream costs are considered, the total cost is:

$$\xi_2 = E_o + C^o + C^f + \sum_{j \in J} (D_j^f + D_j^l + D_j^r) \tag{53}$$

The mean-standard deviation optimization problem is as follows:

$$\min \lambda * E(\xi) + (1-\lambda) * Std(\xi) \tag{54}$$

subject to

$$0 \le h \le H \tag{55}$$

The decision maker minimizes the weighted sum of the expectation and standard deviation. By changing the value of λ in the domain $[0,1]$, various types of decision makers can be modelled. When $\lambda = 1$, the decision maker minimizes only the expectation, as in exiting vehicle holding studies. $\lambda = 0$ represents the other extreme case where the decision makers only minimizes the decision risk. When λ takes other values, a general risk-averse decision maker is modelled.

In practice, the holding time might be subject to other constraints, e.g., the holding should not exceed the reserved slot time in cases of limited departure resources (e.g., gate or track). In such a case, the held vehicle might be dispatched earlier than its optimal holding time because another vehicle needs to use the gate/track/runway occupied by the held vehicle.

Although there is only one decision variable, the solution of this optimization problem is difficult due to the non-convexity of the objective. That means any search algorithm might terminate at a local optimum. Hall et al. (2001) proved that the number of local optima is finite for an ideal case. Ting and Schonfeld (2007) proposed checking each local optimum numerically. This study adopts the same method and employ a direct search to find the optimal solution within a region, when the numerical evaluation of the objective is not expensive.

4. Illustrative Example

Figure 4-1 shows a simplified case where two inbound trucks are delayed when a train is ready for departure. The first truck is estimated to arrive at 6 and the second truck is estimated to arrive at 9. The variance/covariance matrix is $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where ρ measures the correlation between two vehicle arrivals. For convenience, it is assumed that variances of two vehicle arrivals are the same, $\sigma_1^2 = \sigma_2^2$; the headway of the held route is $H = 20$; other parameters q_1, q_2, Q, v and b are assumed to be 1.

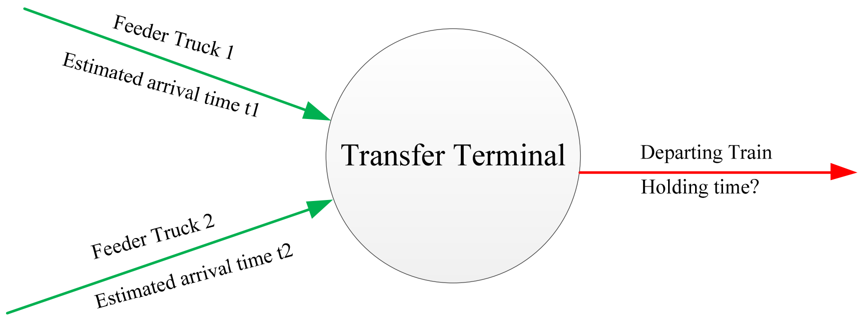


Figure 4-1 An example of two delayed arrivals

The sum of the extra operating cost and the on-board cargos' waiting cost is $E_o + C^o = 2h$. The waiting cost of cargos on two delayed vehicles can be calculated with Equation (37). The changes of the expected total cost with respect to the holding time is plotted in Figure 4-2 under various variances. When the arrival variance is small, implying nearly deterministic arrivals, the curve exhibits a saw-tooth pattern and each sharp angle corresponds to one expected vehicle arrival. As the variance increases, curves become smoother and sharp turns can no longer be observed. Clearly there are already two local optima in the range $[0,14]$, which is a sub-region of the whole solution space. Such a pattern has been reported by Hall et al. (2001) and by Ting and Schonfeld (2007), neither of whom modelled arrival correlations.

While the covariance, i.e., correlation, does not affect the expectation of the total cost, as in Remark 2, effects of arrival correlations on the standard deviation of the total cost are clearly shown in Figure 4-3, in which the arrival covariance $\rho\sigma_1\sigma_2$ changes from -0.3 to 0.6 and the arrival variance is fixed at $\sigma_1^2 = \sigma_2^2 = 0.8$. When $\rho\sigma_1\sigma_2 = -0.3$, vehicle arrivals are negatively correlated; when $\rho\sigma_1\sigma_2 = 0$, there are no correlations between arrivals; when $\rho\sigma_1\sigma_2 = 0.3$ or $\rho\sigma_1\sigma_2 = 0.6$, there are positive correlations.

When h is too large (above 12) or too small (below 2) in Figure 4-3, a larger arrival covariance leads to a larger standard deviation of the total cost. The case $h = 0$ is analyzed as an example. When $h = 0$, there is only one possibility that both vehicles miss their connections. Therefore, the unconditional variance of the total cost equals the conditional variance $\sigma^2_{\xi(4)} = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$. Clearly, the variance of the total cost increases with the arrival covariance $2\rho\sigma_1\sigma_2$. When $\rho\sigma_1\sigma_2 = -0.3$, $\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} = \sqrt{0.8 + 0.8 - 2*0.3} = 1$, which is the lowest standard deviation in Figure 4-3.

When more cases are possible simultaneously, e.g., vehicle 1 (whose expected arrival time is 6) can make the connection and miss it with the same probability when the holding $h = 6$, positive correlations result in a smaller standard deviation.

Remark 3: The decision risk (measured by the standard deviation) is lower if feeder vehicle arrivals are positively correlated when multiple cases can occur at the same time.

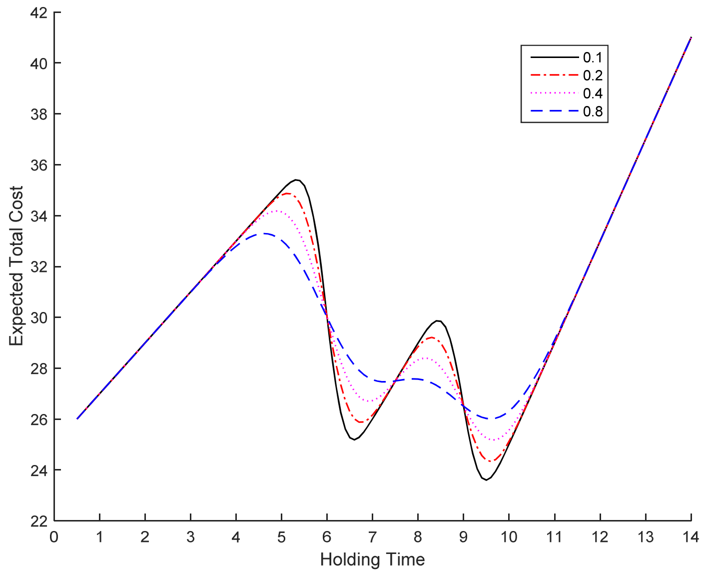


Figure 4-2 Effect of arrival variances on expected total cost

Figure 4-4 shows the expectation and standard deviation of the total cost when the covariance matrix of the joint arrival distribution is given by $\Sigma = \begin{pmatrix} 0.8 & 0.6 \\ 0.6 & 0.8 \end{pmatrix}$. The dashed curve has not been revealed in existing vehicle holding studies which mainly focus on the expected value. The optimal solution obtained by minimizing the expectation is unlikely to be optimal for a risk-averse decision maker, who simultaneously minimizes the decision risk.

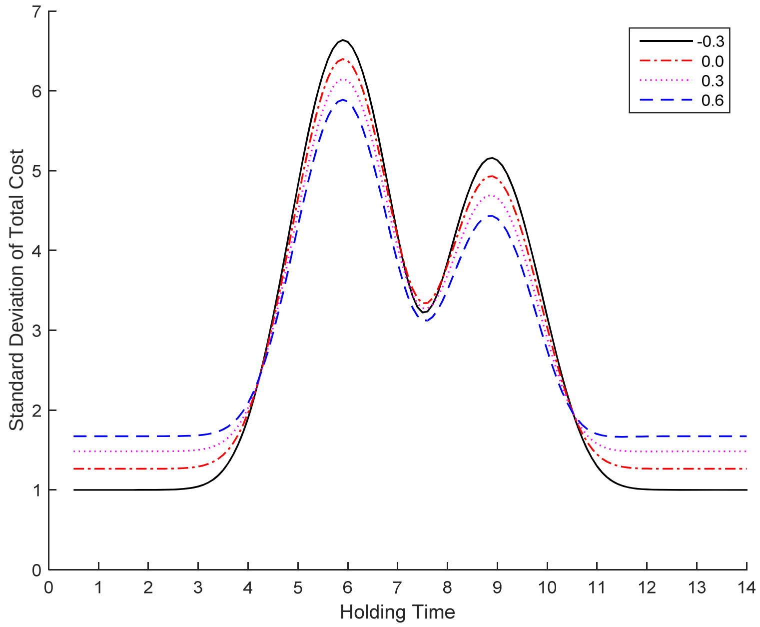


Figure 4-3 Effect of covariances on standard deviation of total cost

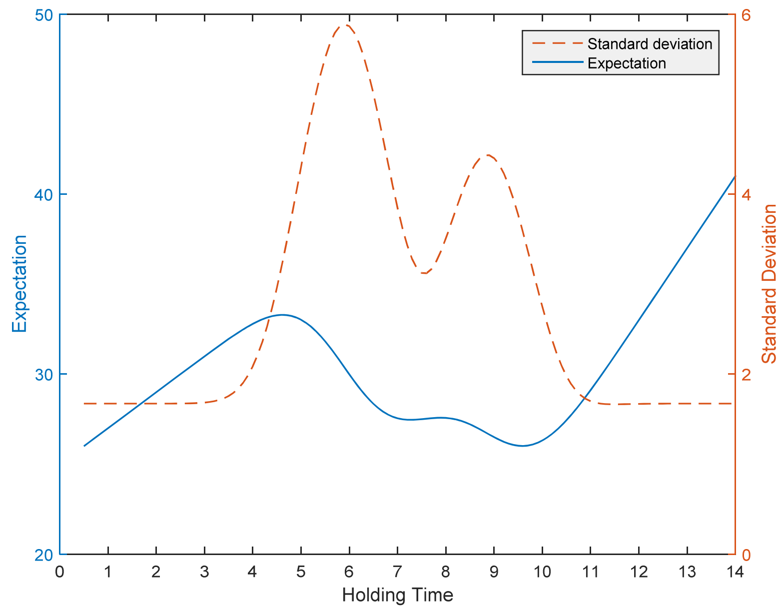


Figure 4-4 Expectation and standard deviation of the total cost vs holding time

5. Case Studies

Figure 5-1 shows a map of the Northeast Corridor located in the United States. Suppose we study one ready vehicle currently at the Baltimore transfer terminal. The vehicle runs on the Washington-New York-Boston route, whose interconnected routes and associated loading/unloading cargos are all shown is Figure 5-1. The headway of this study route is 45 minutes. To make the example general, the terminology of "vehicle" is used throughout this thesis without specifying whether it is a train (a chain of freight cars) or a truck. The origin-destination information of cargos is provided in Table 5-1. All times are measured in minutes.

Table 5-1 Freight origin-destination matrix

OD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	0	0	0	0	0	**5**	**6**	**2**	0	0	0	0	**2**	**4**	**16**
2	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	**4**
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**5**
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**
5	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

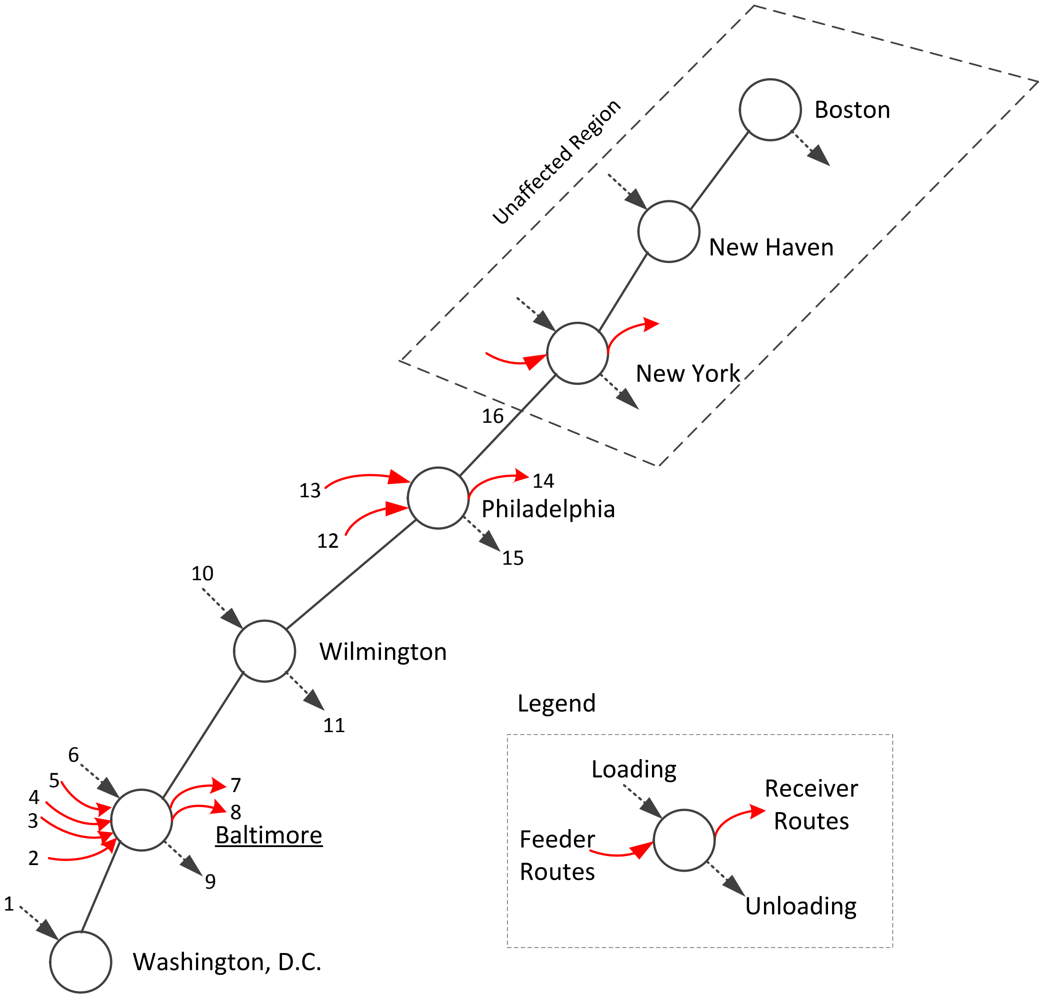


Figure 5-1 Study network

To demonstrate the proposed methods, two numerical examples are presented. In the first study, delay propagations are not considered, i.e., analyzing costs at the current hub only. In the second case, costs at downstream stations in a limited region are further included.

5.1. Case Study 1

When the vehicle is ready for dispatch, cargos from Route 6 have been loaded and goods to Route 9 have been unloaded. These transferring cargos to Routes 7 and 8 have also been transshipped. The amount of on-board cargos is thus obtained by summing up 2, 4, 16, and 5, which is 27. Values of b and v are set to be 12 and 10, respectively.

Feeder vehicles are still on the way. The four vehicle arrivals can be described with a 4-variate normal distribution with a mean vector $\mu = [2, 4, 5, 13]$ and a covariance matrix

$$\Sigma = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.3 & 0.6 & 1 \\ 0 & 0.6 & 2 & 1.6 \\ 0 & 1 & 1.6 & 4 \end{bmatrix} \quad (56)$$

By adopting different values of λ, the weighted objective is plotted as a function of holding time in Figure 5-2.

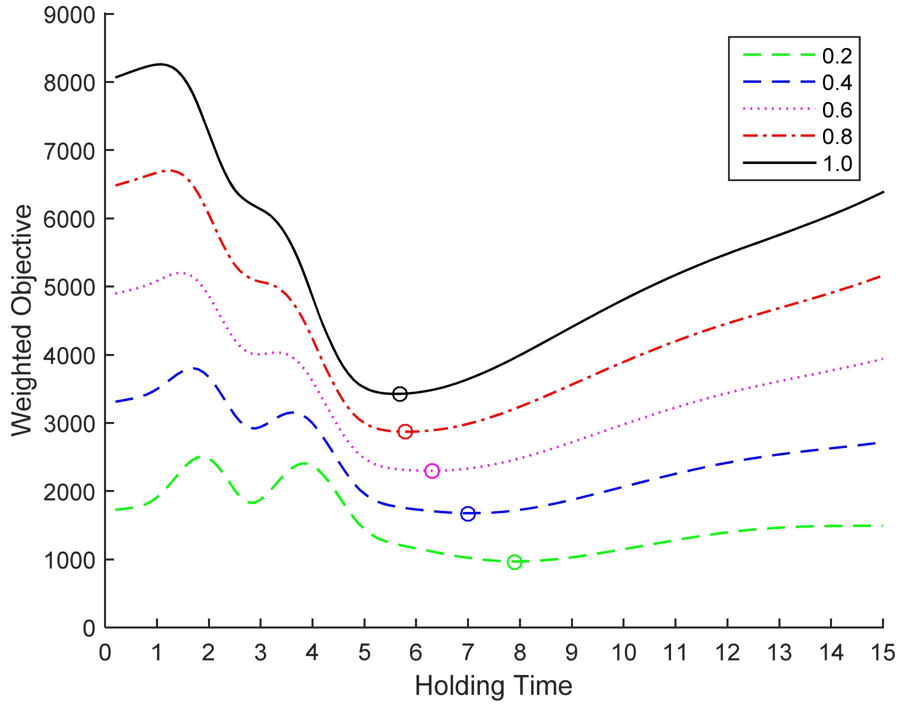


Figure 5-2 Effect of weights on weighted objective (Control station only)

Note that $\lambda = 1$ corresponds to the case where only the expected value is minimized. From Figure 5-2, the optimal vehicle holding is longer as λ decreases, i.e., more weights is given to the minimization of the decision risk. With traditional models which neglect the risk, the optimal holding time would be found through Figure 5-3; however, one is unaware of the decision risk associated with that holding decision, which is shown in Figure 5-4. A risk-averse agent would find that in this example the optimal holding time could be slightly longer to reduce the standard deviation of the total cost at the expense of increasing the expectation. In this way, a more reliable decision is obtained. Thus, the risk-averse formulation can produce more reliable decisions.

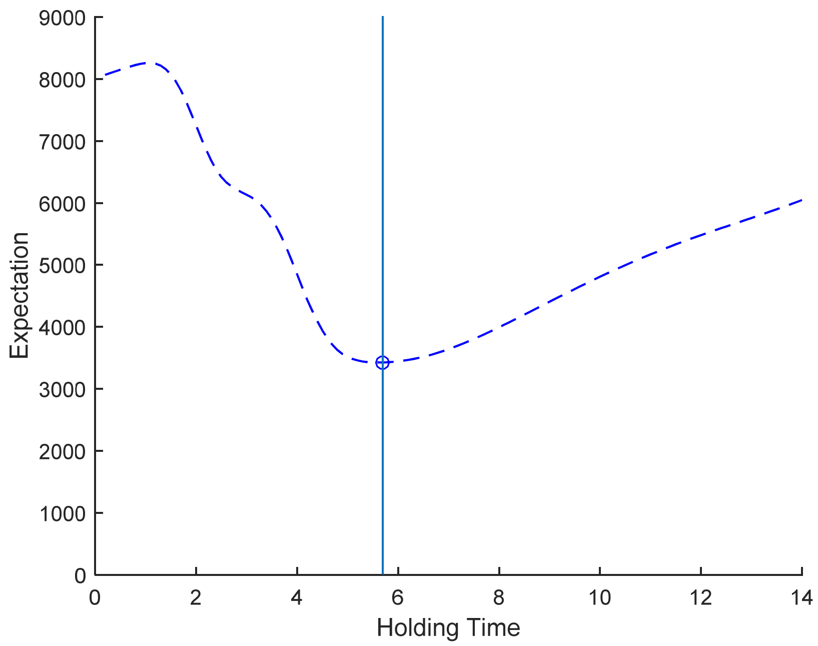


Figure 5-3 Expectation vs holding time

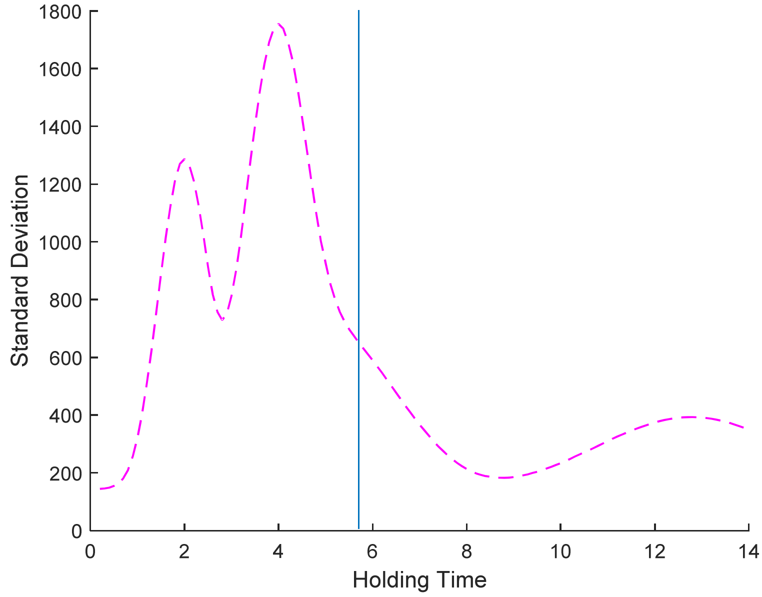


Figure 5-4 Standard deviation vs holding time

5.2. Case Study 2

As discussed in the modeling section, the effect of a holding decision will eventually propa Standard deviation vs holding time Standard deviation vs holding time gate to the whole network if vehicle movements are well coordinated. Due to reasons already provided, it is justified to analyze a temporally and spatially adjacent region, which has a limited size. In Chen and Schonfeld (2011), only the one nearest adjacent downstream station is considered. That approach can be labelled "one step ahead". According to the proposed Guideline 1 for determining impact region, in this example Wilmington is included first into the impact region; Philadelphia is included next. Other downstream stations thus fall into the unaffected region.

The link travel times are 80 minutes from Baltimore to Wilmington and from 40 minutes Wilmington to Philadelphia. The link travel times are assumed to be deterministic because a dedicated (or high) right-of-way is usually provided. For example, in intermodal freight systems, the travel time by rail is more controllable. However, this assumption can be relaxed without much difficulty. If the arrival of the held vehicle is random, the total cost, which becomes a function of the arrival time, can be further integrated over the range of arrival time.

Two delayed feeder arrivals (i.e., Routes 12 and 13) to Philadelphia are described with a bivariate normal distribution with a mean vector $\mu = [124, 126]$ and a covariance matrix $\Sigma = [1.5, 2; 2, 3]$. The departure time distribution of receiver route (i.e., Route 14) is described with a univariate normal distribution with a mean $\mu = 126$ and a variance $\sigma^2 = 1$. If cargos from the held vehicle miss the connection to this run of Route 14, the next scheduled run is at 170.

When the immediately next downstream station, i.e., Wilmington, is considered, the change in the optimal holding time is almost negligible compared to the curve in Figure 5-2. This occurs because only loading cargos (whose amount is quite low) are present at Wilmington and there are no timed transfers at this station; thus there are no missed connection/extra waiting costs. Note only the curve when $\lambda = 0.6$ is shown in Figure 5-5 because the change in the weighted objective with λ has been shown in Figure 5-2.

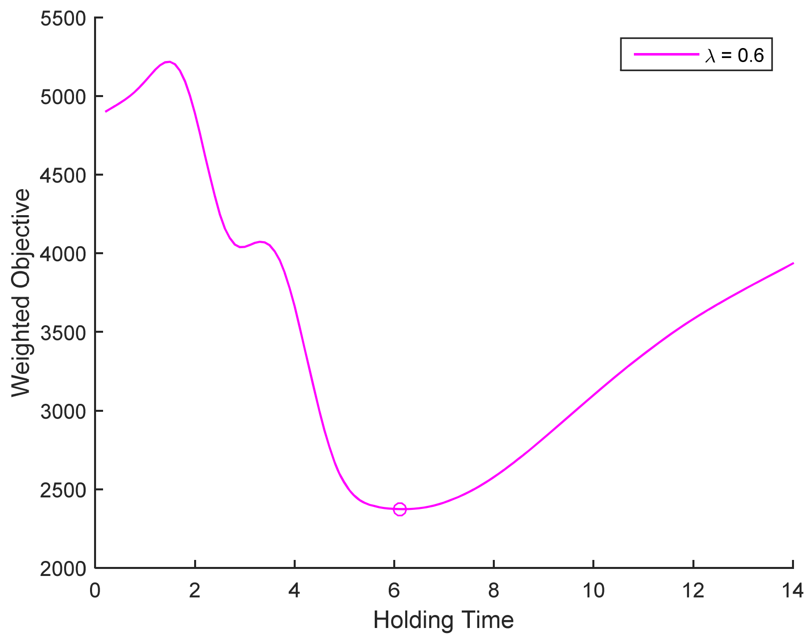


Figure 5-5 Costs at Baltimore (Control station) + Wilmington (Downstream)

If one additional downstream hub, i.e., Philadelphia, is included, the change is significant. The optimal holding time increases to over 8, compared to around 6 in Figure 5-6. The reason is that vehicles inbound to Philadelphia are also behind their schedule, justifying a longer hold. However, it should be noted that the decision is problem-specific. The vehicle might also be released sooner if (1) significant loading cargos are ready, and/or (2) inbound vehicles are on schedule.

One can further add downstream stations into the analysis, up to the boundary station, i.e., Boston, in this case study. Nonetheless, considering that the travel time from the control station to the system boundary is 8 or 9 hours, the reliability of arrival forecasts at Boston seems low and the forecast is almost sure to be updated within some time interval, e.g., 1 or 2 hours. Once the new forecast is available, the decision would be reevaluated, making the previous analysis futile. Based on such an argument, it is suggested here that the impact region should be limited, i.e., only subset of all downstream stations be included at the decision time.

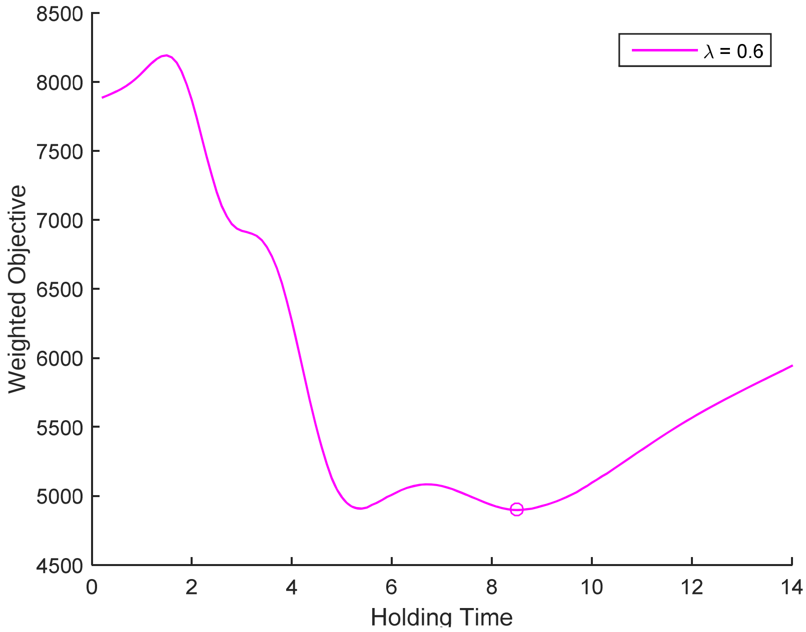


Figure 5-6 Costs at Baltimore (Control station) + Wilmington (Downstream) + Philadelphia (Downstream)

Since a direct search is conducted over a range (from 0 to 14 minutes in this example), the global optimum is not guaranteed. When the numerical evaluation of the objective is inexpensive, one can explore as far as one likes. However, it is unhelpful to search too far if new events (e.g., the feeder vehicle on Route 5 arrives) tend to occur or updated forecast information becomes available frequently. New events or information should trigger reevaluations of the holding decision. Guideline 2 is proposed for determining the extent of the numerical evaluation of the decision variable.

Guideline 2

The extent of the numerical search of the holding time should not be longer than the time to receive updated forecasting information.

In light of the above discussions, a rolling horizon approach (Figure 5-7) can be adopted in the decision making process. The duration of rolling period depends on the frequency of actual vehicle movements or the forecast updates. For example, if vehicle arrivals are updated hourly, it makes little sense to determine a rolling period of 5 minutes. A duration nearer to 60 minutes would be more suitable. The decision maker also needs to know how far into the future forecasts are needed for making the current holding decision. In real time operations, it might take considerable time for the information to be sent to the vehicle, i.e., the communication may not occur in real-time, which also affects the length of rolling period. These are open questions, which are not raised in the existing literature regarding vehicle holding.

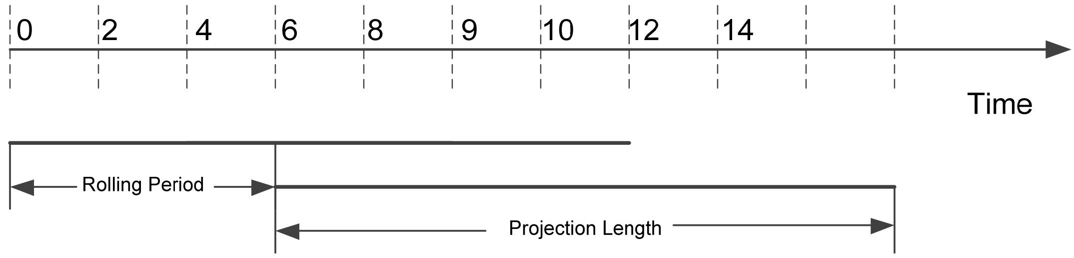


Figure 5-7 An illustration of rolling horizon approach

6. Conclusions

To improve the transfer reliability of coordinated intermodal freight operations, a model is proposed for optimizing the vehicle holding decision in real time. After reviewing both preplanning and real-time control studies for various transportation systems, including air, public transit and intermodal freight transport, it is found that vehicle arrivals are assumed to be independent. To overcome this literature gap, this thesis explores the effect of arrival correlations on the vehicle holding decision. It is proved that *the expected value of the total cost is not affected by the correlations, while the variance can be miscomputed when arrival correlations are neglected*. Specially, it is observed that the decision risk (measured by the standard deviation) is lower if feeder vehicle arrivals are positively correlated when multiple cases are possible simultaneously.

It is also discussed to what extent costs at downstream stations should be considered and suggest the size of the affected region should be limited due to both the computational burden and more importantly the frequency of information/event updates. In other words, even one is equipped with very efficient algorithms for evaluating vehicle holding decisions, it is not justified to include all downstream stations in the analysis because new information would require reevaluations of decisions before the delay propagates to the system's boundary. A rolling horizon framework is briefly discussed as a viable approach to improving current decision making.

An illustrative example is presented, followed by larger case studies, to demonstrate the effectiveness of the proposed method. Numerical results further support the above stated theoretical analyses.

Although significant improvements are made in this study, especially in incorporating correlated vehicle arrivals, this work can be enhanced further in the following aspects:

1) Currently the vehicle holding decision is made independently in the order in which vehicles become ready. This simplifies the problem by allowing us to make separate decisions regarding each ready vehicle. This assumption could be relaxed to reflect the case where vehicle dispatches require shared resources. For example, train departures have to be made on certain tracks of a station and ready flights have to queue up for a certain runway.
2) Although some guidelines are provided, this study does not seek to determine the affected region in this study. A rigorous method for determining it would be very useful.
3) It is generally assumed (including this study) that the vehicle delay due to holding control cannot be recovered. Nonetheless, one can explore the potential of taking delay recovery measures (e.g., speeding up late vehicles) by trading off delay reductions and extra operating costs. This would add one further decision in addition to the holding time optimization.
4) Due to the difficulty of minimizing the nonconvex objective analytically, a simple direct method is used to exhaust all solutions in a region. Obviously, such a brute-force method can be replaced with more advanced searching algorithms, especially when the objective is expensive to evaluate numerically and vehicle arrival updates are frequent.

5) The holding decision is evaluated independently for each ready vehicle in the order of dispatch time. Unless multiple vehicles are ready simultaneously, only one decision at one location is made. However, in some cases where decisions must be made for more than two vehicles, holdings at different locations can be jointly considered and coordinated. Future work should explore the benefit of such a coordinated decision process.
6) A simple risk measurement, i.e., standard deviation, is used in this study. Other risk measures such as the Conditional Value-at-Risk (CVaR) (Rockafellar and Uryasev 2000) can be employed in the future.

Bibliography

Abkowitz, M., Josef, R., Tozzi, J., & Driscoll, M. K. (1987). Operational feasibility of timed transfer in transit systems. *Journal of Transportation Engineering*, *113*(2), 168-177.

Andersen, J., Crainic, T. G., & Christiansen, M. (2009). Service network design with management and coordination of multiple fleets. *European Journal of Operational Research*, 193(2), 377-389.

Bartholdi, J. J., & Eisenstein, D. D. (2012). A self-coordinating bus route to resist bus bunching. *Transportation Research Part B: Methodological*, 46(4), 481-491.

Berrebi, S. J., Watkins, K. E., & Laval, J. A. (2015). A real-time bus dispatching policy to minimize passenger wait on a high frequency route. *Transportation Research Part B: Methodological*, 81, 377-389.

Bertsimas, D., Lulli, G., & Odoni, A. (2011). An integer optimization approach to large-scale air traffic flow management. *Operations Research*, 59(1), 211-227.

Bontekoning, Y. M., Macharis, C., & Trip, J. J. (2004). Is a new applied transportation research field emerging?—A review of intermodal rail–truck freight transport literature. *Transportation Research Part A: Policy and Practice*, *38*(1), 1-34.

Caris, A., Macharis, C., & Janssens, G. K. (2008). Planning problems in intermodal freight transport: accomplishments and prospects. *Transportation Planning and Technology*, 31(3), 277-302.

Chen, C. C. F., & Schonfeld, P. (2010). Modeling and performance assessment of intermodal transfers at cargo terminals. *Transportation Research Record*, No. *2162*, 53-62.

Chen, C. C. F., & Schonfeld, P. (2011). Alleviating schedule disruptions at intermodal freight transfer terminals. *Transportation Research Record*, No. *2238*, 32-43.

Chowdhury, M. S., & Chien, S. I. J. (2001). Dynamic vehicle dispatching at the intermodal transfer station. *Transportation Research Record*, No. *1753*, 61-68.

Chowdhury, S. M., & I-Jy Chien, S. (2002). Intermodal transit system coordination. *Transportation Planning and Technology*, *25*(4), 257-287.

Chuah, K. H., & Yingling, J. C. (2005). Routing for a just-in-time supply pickup and delivery system. *Transportation Science*, *39*(3), 328-339.

Delgado, F., Munoz, J. C., & Giesen, R. (2012). How much can holding and/or limiting boarding improve transit performance?. *Transportation Research Part B: Methodological*, 46(9), 1202-1217.

Dessouky, M., Hall, R., Nowroozi, A., & Mourikas, K. (1999). Bus dispatching at timed transfer transit stations using bus tracking technology. *Transportation Research Part C: Emerging Technologies*, *7*(4), 187-208.

Dessouky, M., Hall, R., Zhang, L., & Singh, A. (2003). Real-time control of buses for schedule coordination at a terminal. *Transportation Research Part A: Policy and Practice*, *37*(2), 145-164.

Greene, W, (2012) Econometric Analysis, 7th edition, Boston, Pearson Education, Inc, ISBN 10: 0-13-139538-6

Hall, R. W. (1985). Vehicle scheduling at a transportation terminal with random delay en route. *Transportation Science*, *19*(3), 308-320.

Hall, R., Dessouky, M., & Lu, Q. (2001). Optimal holding times at transfer stations. *Computers & Industrial Engineering*, *40*(4), 379-397.

Hickman, M. D. (2001). An analytic stochastic model for the transit vehicle holding problem. *Transportation Science*, 35(3), 215-237.

Hoffman, R., & Ball, M. O. (2000). A comparison of formulations for the single-airport ground-holding problem with banking constraints. *Operations Research*, 48(4), 578-590.

Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Muñoz, J. C. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*, 77, 38-75.

Jeong, S. J., Lee, C. G., & Bookbinder, J. H. (2007). The European freight railway system as a hub-and-spoke network. *Transportation Research Part A: Policy and Practice*, 41(6), 523-536.

Kim, M. E., & Schonfeld, P. (2014). Integration of conventional and flexible bus services with timed transfers. *Transportation Research Part B: Methodological*, *68*, 76-97.

Lee, K. T., & Schonfeld, P. (1991). Optimal slack time for timed transfers at a transit terminal. *Journal of Advanced Transportation*, *25*(3), 281-308.

Lee, K. T. & Schonfeld, P. (1994). Real-time dispatching control for coordinated operation in transit terminals. *Transportation Research Record*, No. *1433*, 3-9.

Macharis, C., & Bontekoning, Y. M. (2004). Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research*, *153*(2), 400-416.

Muñoz, J. C., Cortés, C. E., Giesen, R., Sáez, D., Delgado, F., Valencia, F., & Cipriano, A. (2013). Comparison of dynamic control strategies for transit operations. *Transportation Research Part C: Emerging Technologies*, 28, 101-113.

Nesheli, M. M., & Ceder, A. A. (2014). Optimal combinations of selected tactics for public-transport transfer synchronization. *Transportation Research Part C: Emerging Technologies*, 48, 491-504.

Odoni, A. R. (1987). The flow management problem in air traffic control. In Flow Control of Congested Networks (pp. 269-288). Springer Berlin Heidelberg.

Puong, A., & Wilson, N. H. (2008). A train holding model for urban rail transit systems. *In Computer-aided Systems in Public Transport* (pp. 319-337). Springer Berlin Heidelberg.

Richetta, O., & Odoni, A. R. (1993). Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Science*, 27(3), 228-238.

Sánchez-Martínez, G. E., Koutsopoulos, H. N., & Wilson, N. H. M. (2016). Real-time holding control for high-frequency transit with dynamics. *Transportation Research Part B: Methodological*, 83, 1-19.

SteadieSeifi, M., Dellaert, N. P., Nuijten, W., Van Woensel, T., & Raoufi, R. (2014). Multimodal freight transportation planning: A literature review. *European Journal of Operational Research*, 233(1), 1-15.

Ting, C. J., & Schonfeld, P. (2005). Schedule coordination in a multiple hub transit network. *Journal of Urban Planning and Development*, *131*(2), 112-124.

Ting, C. J., & Schonfeld, P. (2007). Dispatching control at transfer stations in multi-hub transit networks. *Journal of Advanced Transportation*, *41*(3), 217-243.

Turnquist, M. A. (1981). Strategies for improving reliability of bus transit service. Transportation Research Record, *Transportation Research Record*, No. 818, 7-13.

Vranas, P. B., Bertsimas, D. J., & Odoni, A. R. (1994). The multi-airport ground-holding problem in air traffic control. *Operations Research*, 42(2), 249-261.

Zhao, J., Dessouky, M., & Bukkapatnam, S. (2006). Optimal slack time for schedule-based transit operations. *Transportation Science*, *40*(4), 529-539.

Zhu, X., Mukhopadhyay, S. K., & Kurata, H. (2012). A review of RFID technology and its managerial applications in different industries. *Journal of Engineering and Technology Management*, 29(1), 152-167.