VALUE OF ITS INFORMATION FOR CONGESTION AVOIDANCE IN INTER-MODAL TRANSPORTATION SYSTEMS[1]

FINAL REPORT

Focus Area: Infrastructure Utilization
Year 2

Principal Investigators:

Wayne State University
Industrial & Manufacturing Engineering
4815 Fourth Street, Detroit, MI 48202, USA

- Alper E. Murat (PI), Assistant Professor
  Tel: 313-577-3872; Fax: 313-577-3388
  E-mail: amurat@wayne.edu
- Ratna Babu Chinnam (Co-PI), Associate Professor
  Tel: 313-577-4846; Fax: 313-578-5902
  E-mail: r_chinnam@wayne.edu

Wayne State University
Civil & Environmental Engineering
5050 Anthony Wayne Drive, Detroit, MI 48202, USA

- Snehamay Khasnabis (Co-PI), Professor
  Tel: 313-577-3915 Fax: 313-577-3881
  E-mail: skhas@eng.wayne.edu

Submitted to:

Richard Martinko
Director, University Transportation Center
University of Toledo

Christine Lonsway
Assistant Director, University Transportation Center
University of Toledo

[1] Throughout this document "Inter-modal Freight" refers to the "Shipment of freight involving more than one mode of transportation (road, rail, air, and sea) during a single, seamless journey".

Table of Contents **Page No**

SUMMARY OF RESULTS

Our project has three major mile-stones for the first year:

- **Mile-stone #1:** Data collection for the pilot studies (inter-modal freight transportation networks, historical data on inter-modal terminal facility disruptions and network incidents, etc.)
- **Mile-stone #2:** Preparing case studies for model and algorithm development and pilot implementation
- **Mile-stone #3:** Developing static optimization algorithms and implementation for pilot studies

The Research Team, made up of Dr. Alper Murat (Project PI), Dr. Ratna Babu Chinnam (Project Co-PI), Dr. Snehamay Khasnabis (Project Co-PI), doctoral student – Farshid Azadian, has made very good progress with respect to all these three milestones over the first year.

In what follows, we summarize our achievement with respect to these milesones.

Mile-stones #1 & #2:

We have concentrated the majority of our efforts on the two main project components:

1. **Dynamic routing of freight on inter-modal networks using real-time ITS information**: We have collected data, developed models and algorithms for the ***air-road inter-modal network*** routing problem application in the states of Ohio and Michigan.

2. **Operational response of supply chains to congestion and incidents at inter-modal terminal facilities**: Working with our collaborator Ford, we collected representative data and developed models/algorithms for the optimal production allocation of scarce on-hand inventory to cope with the delivery tardiness caused by congestion and disruptions on the inter-modal network.

We have approached the data collection and case study preparation from multiple directions as described below. We begin by describing our efforts for the ***air-road inter-modal*** problem. Next, we summarize our data collection and case study achievements in the ***operational response*** problem of supply chains.

For ***dynamic routing on inter-modal networks***, we first considered the ***air-road inter-modal problem*** for freight-forwarders, carriers and shippers. For the air mode, we have collected historical data on the *flight departure delays* by origin-destination airports, carrier, time of the day and month of the year. In addition to departure delays, we gathered data on factors contributing to departure delays such as weather and customs. We have compiled these data into a database and developed a software tool for ease in accessing and modifying the data. These historical data sets allow us to estimate the expected departure delays for most air cargo shipments. Lastly, we also have collected historical data for flight times between national airports. We are now able to estimate the total delay distribution for an air cargo shipment between an origin and destination. The *real-time airport congestion and incident* (departure delay, origin/destination airport congestion, cancellations) information is also available from different sources. In summary, our data sources are

Historical Data on Airport Congestion and Flight Departure Delays:

1. Research and Innovative Technology Administration (RITA)- Bureau of Transportation Statistics, Airline Data and Statistics. (http://www.bts.gov/programs/airline_information/)
2. FAA Operations Network (OPSNET): *OPSNET Delays* provides daily data on reportable delays[2] These OPSNET delays are caused by the application of initiatives by the Traffic Flow Management (TFM) in response to weather conditions, increased traffic volume, runway conditions, equipment outages, and other causes.
3. Flight Stats: FlightStats Analytics - On-time Performance Ratings provides historical flight departure delay data by aggregating across multiple sources (FAA data sources, airport data sources, airline online sources, and published schedule data)[3].

Real-time Data on Airport Congestion and Flight Departure Delays:

1. Flight Stats: Flight Segment Messaging service provides real-time flight status monitoring and notification including alerts on delays, gate changes, weather conditions, schedule changes, etc. [4]

In the air-road inter-modal network, our emphasis has been on the utilization of the Toledo Express airport versus the two Detroit Area airports (DTW and YIP) for shippers and carriers in the OH-MI region. While the road network congestion is not a major concern in the Toledo region, it is an important issue in the Detroit region, especially due to the traffic volume and the location of the area airports. Therefore, our data collection and calibration efforts so far spent on the Detroit region. For the ***road network***, we have leveraged the synergies between our project and another project funded by the MI-OH UTC on the congestion avoidance on road network. From their results, we have gathered and generated such network structure data as the network topology, design parameters, link characteristics for the Detroit. For ***link velocity data*** collection, we are collaborating with the MITS center and Traffic.com for accessing to real-time and historical traffic flow data (such as velocity, occupancy). These datasets play a critical role for incorporating the road-network congestion into our dynamic routing models on the air-road inter-modal network. In addition to recurrent congestion, we also have access to data on the non-recurring congestion (incidents and special events) which improves the accuracy of our models. Especially, Traffic.com has an extensive archive of incident data for developing parametric delay models.

We have developed ***a baseline case scenario*** to test and validate our ***static and dynamic routing models on air-road inter-modal*** networks. Our current base-line model is the allocation of air cargo shipments to different airports in the OH-MI region based on the flight and terminal congestion status (e.g., based on seasonal loads) by considering TOL, DTW, YIP and CLE (Cleveland Hopkins International) airports. The objective has been the validation of the methodology (models, algorithms) rather than a comprehensive assessment of benefits to carriers and shippers. Once we completely validated our algorithmic framework (especially for dynamic models), our next step will be to develop realistic and representative scenarios that account for shipment urgency, freight characteristics (e.g., weight, value, destination), multiple-shipments

---

[2] *"Delays to instrument flight rules (IFR) traffic of 15 minutes or more, which result from the ATC system detaining an aircraft at the gate, short of the runway, on the runway, on a taxiway, or in a holding configuration anywhere en route, must be reported."* - FAA Order 7210.55E

[3] The FlightStats platform of the Conducive Technology Corp. provides real-time and historical flight information. (http://www.flightstats.com/)

[4] Ibid.

with varying origin and destinations, and others important to freight forwarders, carriers and shippers. In this model, we only consider the road network congestion for Detroit. However, Traffic.com's sensor network is also available for Cleveland, so, in our validation, we will represent the road-network and collect link velocity data for Cleveland to better account for recurring/non-recurring congestion on the road network.

For the ***operational response of supply chains*** to congestion and incidents at inter-modal terminal facilities, we have been working with our collaborator Ford. Ford, through its suppliers such as Visteon, imports a significant volume of automotive components (power-train, electrical, chassis and steering) from Asia. For Ford and most other OH-MI supply chains, this sourcing strategy translates to higher pipeline and safety inventory costs. These supply chains therefore feel a constant pressure to reduce their inventories but at the same time improve the delivery reliability of their shipments on the inter-modal network (sea, rail, road, air). In this effort, we collected representative data and developed models/algorithms for optimal production allocation of limited on-hand inventory to cope with the delivery tardiness caused by congestion and disruptions on the inter-modal network. Our representative data contains *product level data* (component, commonality, vehicle configuration, etc.), *plant level data* (production levels, mix-rates, capacity constraints, other manufacturing flexibility characteristics, etc.), and other *supply chain level data*.[5] The current model considers the production allocation as the sole operational response mechanism. An extension of this operational response model is its integration with the dynamic routing models, where we will consider ***break bulk shipments***, i.e., a portion of the containerized freight is shipped via an alternative mode. For this, additional inter-modal network and terminal congestion/incident data from representative routes and facilities are needed.

Mile-stone #3

We have developed models and algorithms for two problems: the ***air-road inter-modal*** problem and the ***operational response*** problem of supply chains. Next, we describe our achievements in respective order.

Majority of our efforts here went toward developing ***dynamic freight routing and mode selection models and algorithms*** on the inter-modal networks. We have tackled the problem specifically for the ***air-road inter-modal network*** as previously described. We have also gone beyond Mile-stone #3 in that our emphasis was not just static but both static and dynamic algorithms.

We first developed ***static models for the air-road freight shipments*** based on expected departure delays, travel times, flight times, and other congestion factors (e.g., security and customs processing) on the road and air modes. These models are large-scale discrete mathematical programs and implemented within the ILOG modeling and optimization platform. Next, we developed compact yet effective *parametric stochastic models* for estimating flight departure delay and airport incidents (e.g. cancellations, diverted flight) based on historical data as well as real-time information. These stochastic models for the air mode are then integrated with the stochastic models developed for the recurring and non-recurring road network congestion. We then have developed *Stochastic Dynamic Programming (SDP)* based ***dynamic models and algorithms for air cargo routing and mode selection*** under real-time traffic and flight-airport information. These SDP algorithms yield optimal routing policies, however they are not computationally efficient for real-world inter-modal networks. This is because, in the presence of multiple modes with different delay intervals and departure schedules, the state space

[5] Our current data sets are representative of the actual operations data and modified to preserve Ford's confidentiality.

representation of SDP becomes prohibitive. Currently, we are testing the algorithmic efficiency of our SDP models and making improvements by investigating ways to decompose the problem by mode so that we could tackle each mode separately with dedicated sub-models. We will use the solutions of SDP models for testing and benchmarking the effectiveness of fast heuristic algorithms that are under development.[6]

We have developed ***static operational response*** models for manufacturers with JIT operations such as our collaborator Ford and many other supply chains in the OH-MI region. These models are used for optimal production allocation of scarce on-hand inventory to cope with the delivery tardiness caused by congestion and disruptions on the inter-modal network. The decisions made by these models are the production schedules, plant idling, production shift decisions, and component substitutions. The static models take into account uncertainty in delivery times and quantities via their expected values, hence are deterministic mathematical programs which can be solved efficiently. We implemented these models within ILOG modeling and optimization platform and are able to solve them in reasonable time. However, solving dynamic models with more realistic size and characteristics (e.g., break bulk) will be a challenge, which we are currently pursuing.

Description of Sections A, B, and C

In Section A we propose a modeling and solution framework for the dynamic air cargo routing on air networks subject to stochastic flight departure delays. After developing a stylized experimental setup, we illustrate the effect of various network factors on the dynamic routing efficiency. In addition, we present a case study using the real data for a dynamic air cargo routing originating from the Cleveland Hopkins International Airport (CLE) and destined to the Seattle-Tacoma International Airport (SEA).

In section B, we extend our approach in Section A to the integrated dynamic routing on the air-road intermodal network. In addition to routing on the air network, we also make alternative access airport selection and dynamic routing decisions on the road network. We illustrate the approach via a case study for a cargo originating from the regions of southeast Michigan and northern Ohio. We consider three main commercial airports in this region Detroit Metropolitan Wayne County Airport (DTW), Toledo Express Airport (TOL) and Cleveland-Hopkins International Airport (CLE) and determine alternative access airport for the cargo under various scenarios.

In section C, we consider the operational response model of an automotive manufacturer facing with a delay in shipments of a component. We consider the case where the manufacturer allocates scarce component inventory among different product lines such that the impact of shipment delay is minimized. We illustrate the modeling and solution methods in a stylized example from a major OEM.

[6] We will first consider the dynamic routing of freight on the inter-modal network using real-time information. However, in the presence of delay, one option is to split the load and ship different portions via various modes. This is especially important for our collaborators (Ford MP&L and C.H. Robinson) to increase the delivery reliability. We will extend the functionality of models for break-bulk shipments.

A. Static and Dynamic Routing using Real-time Information – Air Network

1 Introduction

In the last decades, the US has seen massive growth in air-truck transportation. According to the Bureau of Transportation Statistics (BTS), air-truck intermodal shipments show an increase of 88% in value and 20% in ton-miles from 1993 to 2002 (BTS, 2006). Increasing demand for faster and more reliable shipments of lighter and more expensive goods is the key catalyst for this trend (BTS, 2002). In contrast with the sharp rise of demand, the total number of airports has increased only less than 9% whereas the growth in aircraft fleet has been about 19% from 1995 to 2005 (BTS, 2008). Further, the air travel demand of airline passengers, which share the same assets and infrastructures capacity (e.g., air space and runway) with the air-road shipments, has increased only 51% in the period (BTS, 2005). This steep increase in demand and lagging capacity investments is causing more and more congestion in the air-road multi-modal network. For instance, in 2007, more than 21% of flights had over 15 minute departure delays and the on-time performance of air carriers have been in steady decline from 2002 to the present (BTS, 2008).

There are two alternative solutions to address and mitigate this congestion. The first solution is to increase capacity by constructing more airports or expanding the runway capacity of existing airports. However, these capacity expansions are not only costly and slow but also limited with such constraints as unavailability of land or resistance from the public. An alternative solution is to optimize the utilization of existing capacity through more balanced distribution of the demand load. In this study, we are motivated with the later solution where the users (e.g., freight forwarders or shippers) selectively utilize the air-mode transportation capacity by dynamic routing and allocation based on the congestion state of facilities. Hence, we adopt the perspective of users who, typically, do not own the fleet and use air carrier service non-exclusivity. Clearly, this group does not have control over capacity expansion or utilization of the air network. However, optimizing the routing of time-sensitive shipments to avoid congestion potentially leads to more balanced distribution of system load and thus, increased system utilization.

The ability to choose the optimum route for cargo shipments relies on two factors: ability to change path and proper knowledge to support that decision. The former is possible through flexible contracts with air carriers and the latter is achieved as a result of IT enhancements during the last decade.

As mentioned, air cargo is used whenever shipments are urgent, perishable, or have a particularly high unit value. It follows that the demand for air cargo services is actually often generated at short notice before the actual shipping date. In fact, the majority of bookings are taken within a very short time-span and a large percentage of bookings, primarily high yield bookings, are even made within the last 48 hours before departure (Becker and Dill, 2007). Accordingly, the option of booking or canceling the shipment in the last hours is not a new subject to the air-cargo industry. Moreover, especially in recent years, many carriers are offering more flexibility in contracts as a result of excessive competition in the market. It is even observed that in some markets most of the

freight forwarders tend toward flexible contracts (volume-based) for their business (MECo, 2006).

In this paper, the term “flexibility in contract” is used for any situation which allows the freight forwarder to alter the shipment route through the shipping process. This option can be realized by having contracts with one or more air carriers who offer alternative paths from the origin to destination and lets the shipper make/change the routing decisions almost dynamically. Long-term contracts are one type of business agreement, which allow this flexibility. That is, the freight forwarder has the option of choosing different flights from a carrier and finally pays by the total amount of shipment done based on an agreed measure for a given period.

Another practical method to implement the routing flexibility is replacing cargos in a freight forwarder inventory. Usually, a middle size shipping company has contracts with different carriers for various flights during the day to provide time flexibility for accepting and delivering cargo during the day. However, not all the cargos are time sensitive. Accordingly, the required flexibility can be achieved through replacing the cargos and sending the more time sensitive ones via (expected) faster routes. In this case the committed capacity on the carrier side remains the same.

Having the option of routing, freight forwarders can benefit from dynamic routing decision support tools only if appropriate information is available to justify the decisions. Fortunately, the recent advancement in information technology provides the opportunity to access the almost real time data on air-network status. Nowadays, not only the online flight status is accessible but also the user can be informed about the delay announcement before the delay really happens through delay forecasting services.

Our main objective here is to address the problem of air routing in a dynamic network from the point of view of a freight forwarder. The stochastic dynamic programming approach proposed in this paper uses the available online information to booster the quality of decisions and optimize the routing to improve delivery performance criteria.

The rest of this manuscript is organized as follows: Section 2 reviews relevant current literature. Section 3 formulates the problem statement. Section 3.2 provides details on the proposed routing approach. Section 4 presents the experimental study conducted to analyze the effect of problem parameters. Section 5 details a cargo shipment case study and presents the results of the performance of the approach in comparison to static approaches.

2 Literature Review

The problem of routing and specifically finding the shortest path is a subject of interest for many researchers. A plethora of various approaches is proposed for solving this problem. Some of the classic approaches are collected by Deo and Pang(1984). However, Hall (1986) first studied time-dependent stochastic shortest path and showed that in such problem, adaptive decision policy is more effective than single path selection. Results of his proposed dynamic programming (DP) approach was a set of policies which indicate the optimum action based on both location and time. Wellman et. al. (1990) followed up this algorithm with an optimization of reducing the number of

paths considered on a network that obeys the principle of stochastic consistency. The implementation of the optimization algorithm of Wellman et. al. (1990) significantly reduces the running time of Hall's (1989) algorithm. Wellman et al. (1990) also proved that as long as the Stochastic Consistency is held, the principles of traditional shortest path(s) algorithms can be used with the stochastic networks. Kaufman and Smith (1993) also proposed an optimization on Hall's algorithm by using a heuristic to find upper and lower bounds on the travel time of the final path, so that many paths need not be considered. Furthermore, Wellman et al. (1995) proposed an approximation algorithm in which stochastic consistency and stochastic dominance were used to find approximate shortest paths within continuously tightening upper and lower bounds (Wu and Hartley, 2005).

Various assumptions are made in the literature to define how the realizations of the stochastic network are revealed to the travelers (decision makers). Some assumed that one realized the travel cost on arriving at the node from which the link emanates including Andreatta and Romeo (1988), Polychronopoulos and Tsitsiklis (1996), Cheung (1998), Fu (2001), Waller and Ziliaskopoulos (2002) and Provan (2003). Some like Azaron and Kianfar (2003) assumed that the states of the current arc and immediately adjacent arcs are known on arriving.

No matter when the current state of the network revealed to the decision maker, he always needs to have some evaluation of the expected state of the network based on the available information. In the proposed framework, the main stochastic parameter of the network is departure delay, which directly affects the link travel time. Accordingly, the ability of forecasting the departure delay is essential for improving the networks future state estimation. Predicting the on-time performance of air carriers is a subject of interest to various groups for different purposes. Consequently, there is a notable amount of research in this realm focusing on different factors to satisfy the need of different customers. Traditional linear or nonlinear regression methods have been applied to explain the influence of causal factors on delays by Hansen and Hsiao (2005), Hansen and Zhang (2004) and Vigneau (2003). Micro and macro-level simulation tools have been applied to simulate delays at different levels of detail, for instance, by Hoffman (2001) and Wang and Schaefer (2003). An extensive review of the models and their advantages/disadvantages in forecasting and modeling the flight delays is presented in Xu (2007).

The problem we address in this paper can be distinguished from the existing literature in two ways that make it new research and pioneering work in the field. First, in this problem setting, we have the option of waiting at the nodes for the links(flights) to be available. Actually, the links (flights) in this problem are only accessible for a moment and then will become useless afterward. However, although we wait at the node, we cannot change the path during the waiting period and go through another link unless the link (flight) is canceled. Accordingly, the waiting duration is not part of the link travel time.

Moreover, all the waiting times are stochastic in this problem. The decision maker will never realize the exact state of the network; however, he might get an estimation of the network parameters as real-time information. This information also is not constant and changes stochastically during the routing process. That is, the knowledge of the decision maker may improve as the cargo goes through the network.

3 Dynamic Air Cargo Routing Using Real-time Information

3.1 Routing Model

Let $G \equiv (N, A)$ be a directed graph representing the air network with a finite set of nodes $n \in N$ representing *airports* and a set of arcs $l \in A$ representing connecting *flights* between the airports. There could be multiple arcs between any two nodes designating separate flights. Hence, let i denote the flights and $A_l \subseteq A$ denote the set of flights between airports n' to n'' where $i \in A_l$ and $l = (n', n'')$. A dynamic routing problem on this air network is concerned with departing from the *origin node* n_o and arriving to the *destination node* n_d via a series of flight selection decisions. The goal is to find an optimal routing policy that minimizes the total cost criteria.

The arcs have three parameters affecting the flight selection decision. As in most network models, each arc has a *deterministic travel time* $(\tau_i > 0)$, which corresponds to the flight time. The second parameter is the *scheduled departure time of flight i* (θ_i), which, in essence, results in an unknown waiting time at the nodes. Note that, although the scheduled departure times are exactly known, the arrival time to the airport node is unknown and thus makes the waiting time a stochastic variable. The final parameter, which is absent from most routing models, is a *stochastic departure delay* ($\delta_i \geq 0$) corresponding to an uncontrollable waiting time at the origin node of the arc (flight *i*) before traveling through it. This departure delay is attributable to a multitude of factors including congestion at the origin and destination airports, cargo-processing delays, and the like. Departure delay is non-negative, indicating that the flight departs only after the scheduled departure time. Accordingly, if the flight has not departed past the scheduled departure time, the actual departure time depends on δ_i and is stochastic. Once the flight has departed, the arc becomes unavailable. This temporal change in arc availability is, indeed, another attribute that distinguishes this problem setting from the stochastic shortest path problems.

The flight selection decision is made upon cargo arriving at a node, with no recourse decision at that node meaning the flight decision is permanent. This is a reasonable assumption since the freight forwarders are not at liberty to get cargo loaded and unloaded at a short notice. Usually, the carriers sell their cargo capacity some time in advance (e.g., in hours), which is sufficient for the freight forwarder to commit to a flight while en route in preceding airports.

Departure Delay Distribution

We denote the cumulative distribution function of the departure delay for link i with $\Psi_i(\delta_i)$. The assumption is that the flight will depart during the finite period of time like ζ. Accordingly, after ζ the flight is defiantly departed, i.e. $\Psi_i(\delta_i \geq \zeta) = 1$. On the other hand, since departure before the scheduled departure time is not allowed, δ_i could not be negative. Ergo $\Psi_i(\delta_i) = 0$ when $\delta_i < 0$. The detailed explanation of the departure delay distribution estimation is provided latter.

Departure Delay Announcement

Upon the arrival of cargo at a node at time *t*, the real-time information regarding the distribution of departure delay of all flights in the current and all the future nodes are revealed to the freight forwarding agent. This information is called the *departure delay announcement* and characterized by a 2-tuple, $(\eta_i(t), \upsilon_i(t))$, defining the upper and lower bound on departure delay. It is assumed those announcements are reliable meaning the actual final departure delay is always bounded by the announced bounds at any time $t, 0 \le \eta_i(t) \le \delta_i \le \upsilon_i(t) \le \zeta$, where ζ is the *maximum allowed departure delay*. We further assume that the quality of announcement (e.g. spread between the bounds) will improve over time or at least remain the same, $\forall i \in A, \forall t, t' \mid 0 \le t < t' : \eta_i(t) \le \eta_i(t')$ *and* $\upsilon_i(t') \le \upsilon_i(t)$. Moreover, if the flight has already departed, the announcement would be exact, e.g., $(\eta_i(t), \upsilon_i(t)) = (\theta_i + \hat{\delta}_i, \theta_i + \hat{\delta}_i) \mid \theta_i + \hat{\delta}_i \le t$ where $\hat{\delta}_i$ is the realized departure delay.

Probability of Flight Departure

Let cargo arrive at a node at time t_0. On arrival, the decision maker receives the possible departure time window as $(\eta_i(t_0), \upsilon_i(t_0))$. However, his interest is to estimate the likelihood of the flight being available and the cargo departing the airport at any given time in this window. In other words, the decision maker needs to estimate $\Pr(\delta_i = \hat{\delta}_i, \delta_i \ge t_0 - \theta_i)$. On the other hand, relation (1) holds.

$$\Pr(\delta_i = \hat{\delta}_i, \delta_i \ge t_0 - \theta_i) = \Pr(\delta_i = \hat{\delta}_i \mid \delta_i \ge t_0 - \theta_i) \Pr(\delta_i \ge t_0 - \theta_i) \qquad (1)$$

Relation (1) can be interpreted as the multiplication of the probability of departing with $\hat{\delta}_i$ delay given that the departure time is after the arrival time at the node t_0 (the flight is still available) and the probability of the flight being available at the arrival time.

Knowing that the flight will depart in $(\eta_i(t), \upsilon_i(t))$ window and given the distribution of departure delay $\Psi_i(\delta_i)$, $\Pr(\delta_i \ge t_0 - \theta_i)$ can be calculated as below

$$\Pr(\delta_i \ge t_0 - \theta_i) = \frac{\Psi(\upsilon_i(t_0)) - \Psi(\max\{\eta_i(t_0), t_0\})}{\Psi(\upsilon_i(t_0)) - \Psi(\eta_i(t_0))} \qquad (2)$$

Clearly, if cargo arrives before $\eta_i(t_0)$, the flight is definitely still available, $\Pr(\eta_i(t_0) > \delta_i \ge t_0 - \theta_i) = 1$ and if it arrives after $\upsilon_i(t_0)$, the flight has definitely already left, $\Pr(\delta_i \ge t_0 - \theta_i \ge \upsilon_i(t_0)) = 0$.

On the other hand, the probability of flight departure at a given time after the arrival time at the node, $\Pr(\delta_i = \hat{\delta}_i \mid \delta_i \ge t_0 - \theta_i)$, is calculated by relation (3).

$$\Pr(\delta_i = \hat{\delta}_i \mid \delta_i \geq t_0 - \theta_i) = \frac{\psi(\delta_i = \hat{\delta}_i)}{\Psi(\upsilon_i(t_0)) - \Psi(\max\{\eta_i(t_0), t_0\})} \tag{3}$$

Accordingly, the probability that the flight be available after arriving at the node and depart at a given time of $\theta_i + \hat{\delta}_i$ is as (4).

$$\Pr(\delta_i = \hat{\delta}_i, \delta_i \geq t_0 - \theta_i) = \frac{\psi(\delta_i = \hat{\delta}_i)}{\Psi(\upsilon_i(t_0)) - \Psi(\eta_i(t_0))} \tag{4}$$

From now on, this probability $\Pr(\delta_i = \hat{\delta}_i, \delta_i \geq t_0 - \theta_i)$ will be noted as $p_i(t_0, \delta_0)$.

3.2 Solution Algorithm

If cargo left the airport n_a using flight i at time $\theta_i + \delta_i$, it would be at the next airport at time $\theta_i + \delta_i + \tau_i$. Let the cost-to-go at each node at each time be the minimum expected cumulative cost from that node to the destination. If the cost to go to the airport n_b at time t noted as $H(n_b, t)$, then being at node n_a and deciding to take flight i to go to node n_b, there is the probability of $p_i(t_0, \hat{\delta}_i)$ to experience the cost of $H(n_b, \theta_i + \hat{\delta}_i + \tau_i)$ at the next node. Let the cost of taking a flight i be a fixed cost as G_i. Consequently, the expected cost-to-go of committing to flight i at time t_0 to go from n_a to n_b is noted as $F_i(n_a, t_0)$ and calculated through (5).

$$F_i(n_a, t_0) = \frac{\sum_{\delta=0}^{\zeta} p_i(t_0, \delta) \times H(n_b, \theta_i + \delta + \tau_i)}{q_i(t_0)} + G_i \tag{5}$$

$$q_i(t_0) = \sum_{\delta=t_0}^{\zeta} p_i(t_0, \delta) \tag{6}$$

In relation (5), ζ is the upper bound of departure. That is, it is assumed that the flight will depart at ζ in the latest departure case. As defined by (6), $q_i(t_0)$ is the total probability of departing with flight i after time t_0. It should be noted that $\forall t \leq \theta_i : q_i(t) = 1$ and $\forall t > \theta_i : q_i(t) \leq 1$.

Let set of $\pi(n, t_0)$ be all the flights departing from airport n after time t. Clearly, from all the flights in $\pi(n, t_0)$ the decision maker would like to select the one with the lowest cost-to-go. Let flights in $\pi(n, t_0)$ be re-indexed by ascending order of cost-to-go in a way that for $\forall i, j \in \pi(n, t_0) \mid i \leq j$ we have $F_{(i)}(n, t_0) \leq F_{(j)}(n, t_0)$. As mentioned, the probability of experiencing $F_{(i)}(n, t_0)$ is $q_{(i)}(t)$. It should be noted that $F_{(i)}(n, t_0) \leq F_{(i+1)}(n, t_0)$ but there is no

predefined relation between $q_{(i)}(t)$ and $q_{(i+1)}(t)$. Accordingly, the expected cost-to-go at airport n at time t_0 is calculated by (7).

$$H(n,t_0)=\sum_{i\in\pi(n,t_0)}\left(q_{(i)}(t_0)\prod_{\substack{k=1\\k\in\pi(n,t_0)}}^{i-1}(1-q_{(k)}(t_0))F_{(i)}(n,t_0)\right)+M\prod_{i=1}^{N}(1-q_i(t_0)) \qquad (7)$$

In (7), M is the ultimate cost that a freight forwarder will experience if he fails to deliver the cargo at all. The idea in this relation is that the decision maker will select the cheapest flight when it is available and all the cheaper flights have already left. However, if all the flights have left, cargo is stocked in the airport and the freight forwarder will experience a penalty of M. Clearly, this last scenario is not very likely to happen in the real world since there are always future flights available.

Relation (7) is actually the recursive relation to calculate the minimum expected cost-to-go at each node for each decision time. Now, let cost at the final destination be $Pen(t)$ which measure the difference between the time of delivering the cargo and the predefined due date noted as *T*. By assigning $H_t(n_d)=Pen(t)$, it becomes possible to calculate $H_t(n),\forall n\in N,\forall t\geq 0$.

Accordingly, if $S(n')$ is all the potential nodes to visit after node n', by arriving at node n' and after realizing the network state, the decision maker can calculated the $H_t(n''),\forall n''\in S(n'),\forall t\geq 0$. Removing the flights that already left from $\pi(t_0,n')$, the remaining flights are those which will be available and eventually depart. Now, using $H_t(n'')$, the decision maker can estimate the expected cumulative cost-to-go of each flight $F_k(t_0,n'),\forall k\in\pi(t_0,n')$. Clearly, the best action is to commit to a flight with the minimum expected cumulative cost-to-go. So, if $\pi^*(t_0,n')$ represents the best action from the possible set of actions of committing to a flight, it would be driven from relation (8).

$$\pi^*(n',t_0)=\arg\min_{k\in\pi(n',t_0)}\{F_k(n',t_0)\} \qquad (8)$$

The main difference between this approach and ordinary dynamic programming solutions is in the state space situation and consequently the policy calculation. The state space in this problem includes the location of the decision (airport), the time and all the flights' departure delay. Since time and delay are continuous variables by nature, to be able to handle them in this framework they need to be discretized which also reduces the accuracy. It should be noted that even after sacrificing a great deal of accuracy to reduce the state space, the possible combinations of time and flight status for a medium size network is much bigger than the ability of ordinary computers. Moreover, these states are stochastic and the probability of happening for some of them would be very low. Accordingly, as can be seen in the proposed approach, the calculation is postponed to the latest possible time and only done after the network state is revealed to the decision maker that clearly reduces the time and flight combination. In addition, by using the expected cost-to-go at each airport, the need to calculate the action (flight commitment) cost based on future potential actions is replaced by calculating the cost of action based on the future airport to visit. It other

words, the decision maker, in practice, instead of deciding on the sequence of actions he wants to take in the future, at each step will choose between a group of potential actions by selecting the airports.

As mentioned before, this option of postponing the action selection is an important outcome of having flexible contracts with air carriers. The assumption here is that contracts between the freight forwarder and air carrier, enable the freight forwarder to postpone the commitment to a flight up until the arriving time at the airport. Clearly, after committing to a flight, the decision maker cannot alter his choice no matter what he realized about the network state afterwards.

3.3 Alternative Policy Types

To be able to compare the performance of the proposed model, we first defined two additional alternative models and then compare the results of the dynamic approach against them.

3.3.1 Static Policy vs. Dynamic Policy

The purposed approach relies on the flexibility of changing the path that comes from the contracts between freight forwarders and air carriers. Having this flexibility, the performance can be improved by using real-time information to change the prior decision and select the better path; however, without this option freight forwarders would be forced to make a decision earlier and then commit to it for the rest of the shipment. Clearly, if the decision maker cannot change his decision, receiving and processing the new information would not have any value. The former case in this paper is noted as static policy versus the proposed approach that is noted as dynamic policy.

By definition, in static policy the freight forwarder will buy the capacity for the cargo in advance. He could use all the historical information available, however, after the commitment to the air carrier has taken place, the selected path could not be changed in any circumstance.

3.3.2 Benchmark

Although in the proposed approach, the freight forwarder will receive some information about the actual departure delay, this information is not perfect and sometimes the announced boundary could be practically wide in a way that does not help the decision maker. By having a base to compare the performance of the dynamic policy against the ideal situation, all the actual departure delays are known in advance, and a new model is defined as the benchmark. This model will represent an imaginary situation that the freight forwarder is given representing the ultimate perfect information making it possible to compare the performances.

4 Experimental Study

We have conducted a set of controlled experiments to investigate the effect of such problem parameters as the announced delay accuracy, the departure delay distribution, and the number of

connections on various performance criteria (e.g., expected cost and delivery reliability). A select subset of the results is presented and discussed in the remainder.

The experimental study is based on three network configurations denoted as $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_3$ and illustrated in Figure 1 together with the problem parameters. In all three configurations, the origin airport is *A* (origin), the destination airport is *D* (destination), and there are two alternative intermediate airports (*B* and *C*) accessible from the origin. Furthermore, the expected travel time of going through *B* or *C* the is same for all three networks. The $\mathcal{N}_1$ configuration represents the base case configuration from which the other two network configurations are derived. The $\mathcal{N}_2$ network configuration is identical to the base case except for airport *C*, where we vary the departure delay distribution shared by all outgoing flights. The expected delay at node *C*, however, is kept constant for all distributions. The $\mathcal{N}_3$ network differs from the base case, in that, direct flights from *B* to *D* are replaced with one-stop flights that connect at node *E* with different delay distributions. We assume that, at t_0=95, the cargo is processed and ready for loading onto the next available flight. Further, the due date is set at T=100, e.g., the cargo requires expedited shipment.

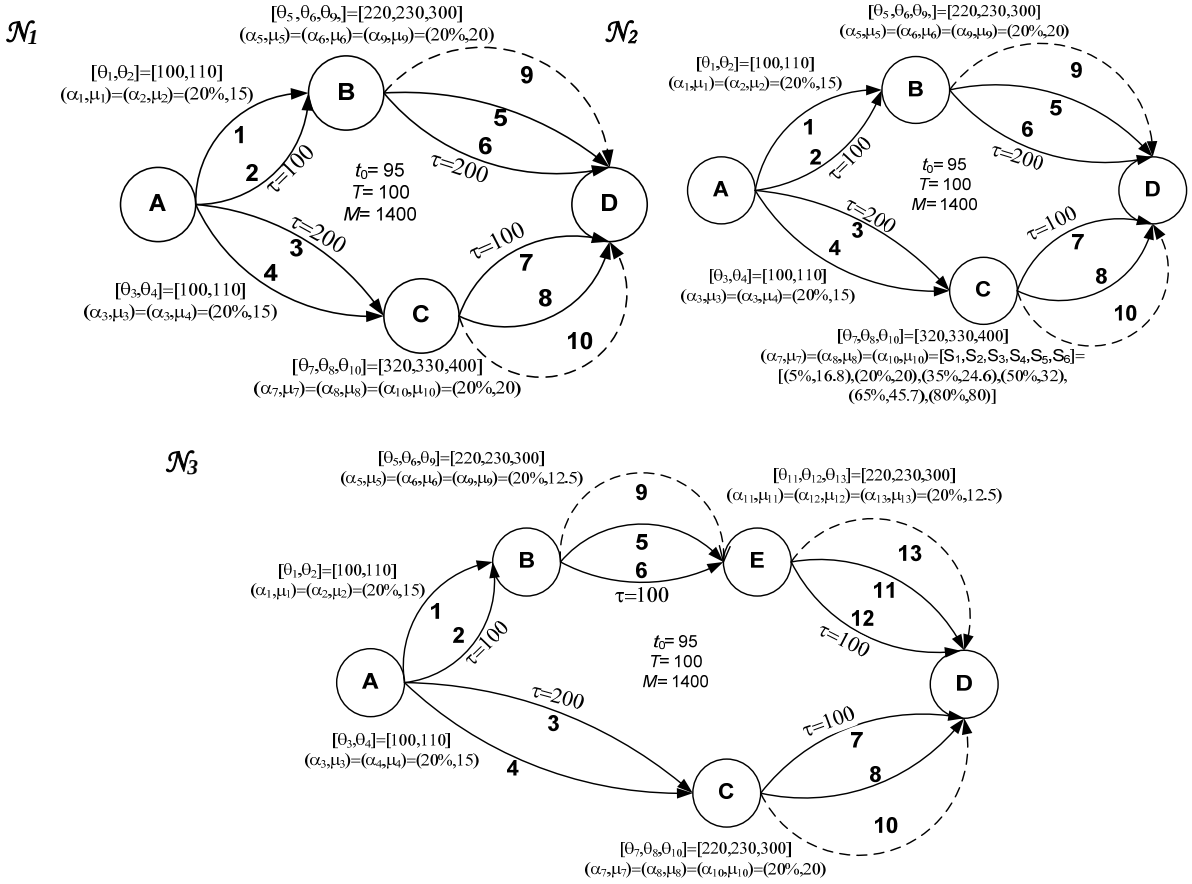


Figure 1. Network structure and parameters for the three network configurations ($\mathcal{N}_1$,$\mathcal{N}_2$,$\mathcal{N}_3$).

In order to better understand the effect of experimentation parameters and without loss of any generality, we consider the travel time, rather than the cost (flight cost, delay penalty), as the performance measure. Accordingly, our penalty function is delivery tardiness. We further assume

that there are no flight cancellations and that all flights depart within 90 minutes of scheduled departure. It is also assumed that the cargo will be accepted so long as it arrives before the flight departure (e.g., no capacity constraints). In Figure 1, the flights with dashed lines (*9*,*10*,*13*) represent the bypass flights which guarantees the delivery of the cargo. The departure times of these bypass flights are separated in such a way that they are not selected unless the connection flight is missed.

In all network configurations, we assume that the departure delay distributions are exponentially distributed. The probabilities of on-time departures and the exponential distribution parameters are provided in Figure 1. In the experimentation, we study the effect of announcement accuracy and departure delay distribution on various performance measures. These effects depend on both the delay distributions selected and the representation of the announced delay information. Figure 2, illustrates the delay bounds for various levels of announcement accuracy and the departure delay distributions at airport *C* in the network configuration $\mathcal{N}_2$.

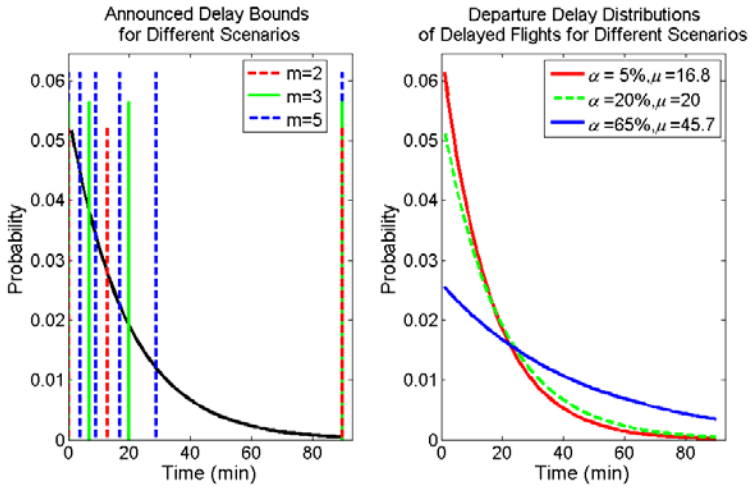


Figure 2. Announced delay bounds for airport C in ($\mathcal{N}_1$, $\mathcal{N}_3$) and various departure delay distributions for airport C in ($\mathcal{N}_2$).

For each network configuration, we first solved for *three routing policies* (dynamic, static and dynamic with perfect information) and then simulated each policy for 20,000 samples. The solution quality of each policy depends on the problem parameters, e.g., flight durations, departure time separations, and delay distributions. To reduce the effect of problem specific parameters in quantifying the comparative performance, we define the equation 9.

$$\rho = \frac{H_{\pi_s}(\Omega_k) - H_{\pi_d}(\Omega_k)}{H_{\pi_s}(\Omega_k) - H_{\pi_d}(n_0, t_0)} \qquad (9)$$

The measure (ρ) is defined as the dynamic policy's improvement over the static policy and expressed as a percentage of total savings possible under perfect information. Note that this is a conservative estimate, since the perfect departure delay information is not readily available in practice.

Figure 3 presents the distribution of flight paths for three different network configurations. Because the dynamic policy can choose any flight path combination, we present only the paths with more than 2.5 per cent occurrence and the rest are categorized as "Others". In the case of $\mathcal{N}_1$ with m=1, the dynamic policy is almost indifferent between the two intermediate airports (B and C) and tends to choose early flights going out of airport A. As the accuracy of the announced delay information increases from m=1 (no real-time information), the dynamic policy begins choosing secondary flights (e.g., flights *2* and *4*). This is attributable to the instances, where the announced delay for the early flight makes the secondary flight desirable. However, at all accuracy levels, the two intermediate airports (B and C) are almost equally visited, albeit through various flight paths. Further, the dynamic policy chooses the secondary flights early in the trip rather than later, i.e., the flight path (*2*,*5*) is selected more frequently than (*1*,*6*). In contrast, the static policy commits to a specific flight path. However, whenever it misses a connecting flight, it chooses the next flight out.

In the case of $\mathcal{N}_3$ with m=1, we observe that the flight path with the lowest number of connections (i.e., passing through C) is preferred the most. Note that the flight delay distribution at connecting airports (B and E) for $\mathcal{N}_3$ has a lower expected delay than that of $\mathcal{N}_1$. Subsequently, the likelihood of getting on the early flight in the connecting airport is lower, which results in longer expected travel time. As the announcement accuracy increases, the dynamic policy begins selecting secondary flights as they become desirable with the delay announcement for earlier flights. In addition, with sufficient announcement accuracy, the dynamic policy sometimes chooses the most preferable path through B, which is all the early flights departing from B and E.

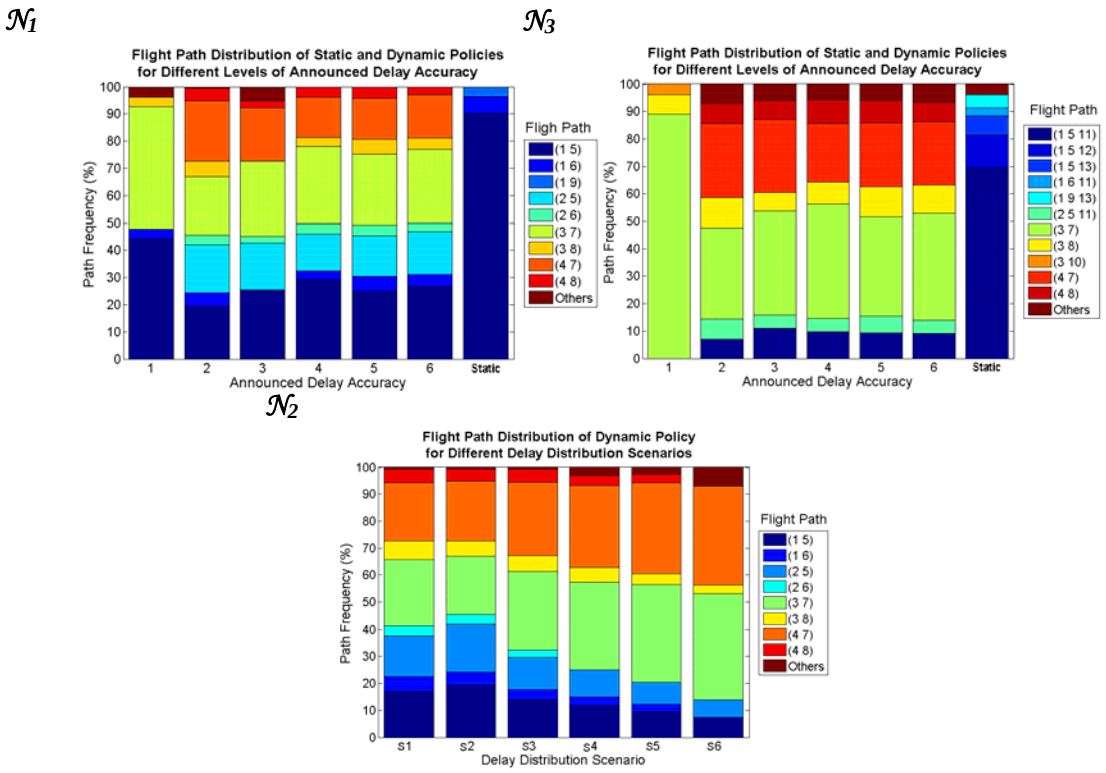


Figure 3. Flight path distributions of static and dynamic policies for different levels of announced delay accuracy ($\mathcal{N}_1$, $\mathcal{N}_3$) and different delay distributions ($\mathcal{N}_2$).

In the case of $\mathcal{N}_2$, Figure 4 illustrates the effect of changing the delay distribution common to all flights departing from airport *C* with *m*=2. These distributions share the same expected delay. Note that the second distribution case (S_2) is identical to the result in the $\mathcal{N}_2$ case with *m*=2. As the conditional expected value of delay distribution increases (or decreases) from that of S_2, then the dynamic policy prefers the flight paths going through *C*. For instance, let us compare the case S_2 (high on-time departure probability) with the case S_6 (high conditional expected delay). In the case S_6, the dynamic policy chooses flights going to airport *C* more, since, in the less likely event that the connecting flight (such as *7*) is delayed, then it can get on the next flight out which is the flight *8*. This alternative flight option, in essence, provides the dynamic policy a *truncation on the delay distribution* at airport *C*.

In summary, the choice of flight paths depends on the policy used. The static policy trades off the tardiness of a fixed path with the risk of missing a connecting flight. In comparison, the dynamic policy exploits both the real-time departure delay information (whenever available) and the multiplicity of flights departing from connecting airports.

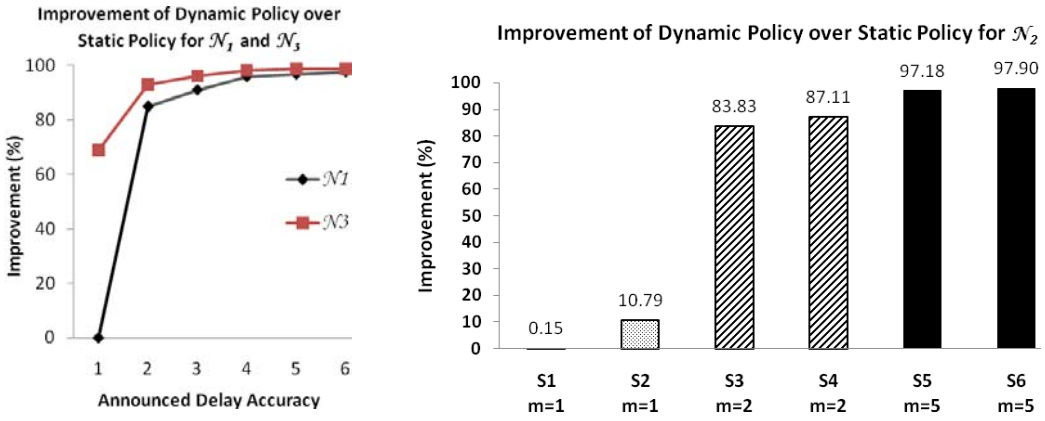


Figure 4. Improvement (ρ) of dynamic policy over static policy for $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$.

In Figure 4, we present the impact of the announced delay accuracy and the delay distribution over the comparative performance of the dynamic policy against the static policy. The comparison for $\mathcal{N}_1$ and $\mathcal{N}_3$ (graph on the left), indicates that the rate of improvement with increased accuracy is diminishing. Hence, we conclude that the dynamic policy can achieve the majority of performance improvement over the static alternative even with some level of real-time delay information.

The effect of different delay distributions on the comparative performance is illustrated in Figure 4 (graph on the right). This effect is most apparent when the accuracy is increased from *m*=1 to *m*=2. As mentioned before, with S_5, the on-time departure probability is very high and there is some level of truncation of the experienced delay at the airport *C*. This is, indeed, the reason as to why case S_5 is better performing than the case S_1. In case S_1, the delay is frequent but has a lower conditional expected duration. With *m*=2, the accuracy of the announced delay information is more effective as the dynamic policy can entertain multiple flights (see Figure 5) for the flight path distribution for *m*=2 and S_1). Similarly, for case S_5, the increased announcement accuracy

leads to a better performance, but the dynamic policy still prefers paths going through C due to the effect of truncation in the experienced delay.

Another important performance measure for the shippers and freight forwarders is the delivery reliability, i.e., the percentage of shipments arriving on time. We use the conditional expected tardiness and the conditional standard deviation to illustrate the delivery reliability differences between static and dynamic policies. Figure 5 shows the conditional expected tardiness for all three networks for different levels of announced delay accuracy and delivery due dates. Note that the tardiness is based on static and dynamic policies obtained for due date T=100, i.e., they are the tail conditional expectations. For $\mathcal{N}_1$ and m=1, the static and dynamic policies share the same tail distribution and thus the tardiness overlap. With increased accuracy, the tardiness is much lower for dynamic policy. Moreover, as we increase the due date, we note that there are some significant late deliveries associated with static policy. The case for $\mathcal{N}_3$ is similar to $\mathcal{N}_1$, but the difference in static and dynamic policy tardiness is more remarkable. In the case of $\mathcal{N}_2$, the conditional expected tardiness of the two policies with no real-time information is similar and insensitive to the delay distribution. Further, with an increased level of announced delay accuracy, the effect of the delay distribution on the conditional expected tardiness diminishes.

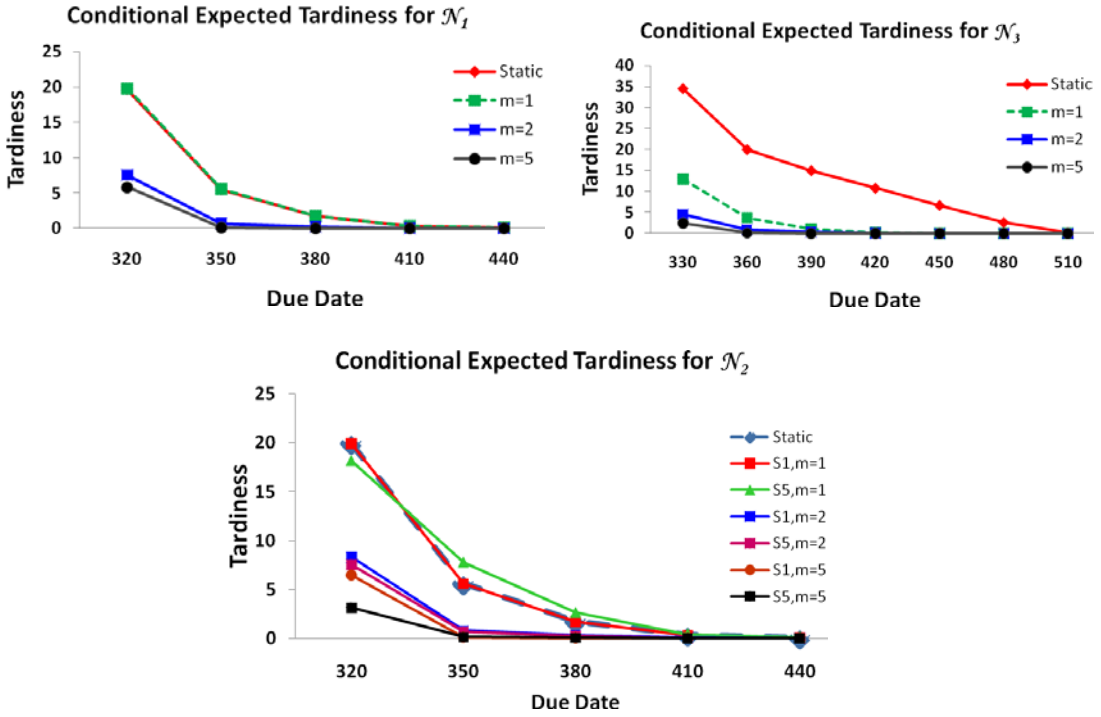


Figure 5. Conditional expected tardiness for different due date levels ($\mathcal{N}_1$,$\mathcal{N}_2$,$\mathcal{N}_3$).

The tail conditional expectation is only an average measure of the delivery reliability and does not represent the degree of variation in the delivery tardiness. Instead, we use tail conditional standard deviation of the tardiness to compare the two policies (Figure 6). For $\mathcal{N}_1$ and $\mathcal{N}_3$, we note that static policy's conditional tardiness variation is significantly higher than that of dynamic policy, except for $\mathcal{N}_1$ with m=1 where dynamic and static policies overlap. Furthermore, as the accuracy

increases, the dynamic policy's conditional tardiness decreases with a diminishing rate. The effect of the delay distribution on the conditional variance of tardiness is insignificant compared to the announcement accuracy.

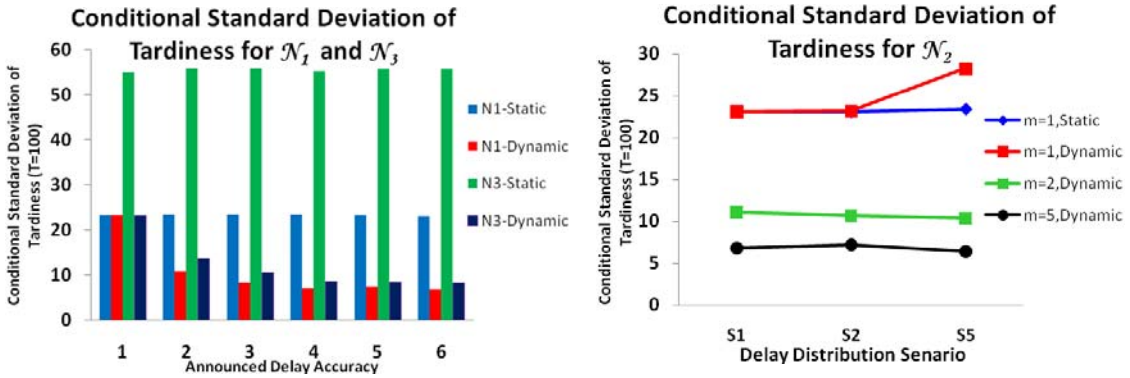


Figure 6. Tail conditional standard deviation of tardiness ($\mathcal{N}_1$,$\mathcal{N}_2$,$\mathcal{N}_3$).

5 Case Study

In this section, we present a case study application of the proposed approach in a real world scenario. The problem in this case study is to route air-cargo originating at Cleveland Hopkins International Airport (CLE) and destined to Seattle-Tacoma International Airport (SEA). We first describe how we estimate the delay distribution model parameters from the real-world data sources. Next, we describe the case study network and then discuss the analysis results.

5.1 Estimation of the Departure Delay Model Parameters

As described in the modeling section, we assume that the departure delay (δ) distribution of flight i has a cumulative density $\Psi_i(\delta)$. The parameters of this function are the on-time departure probability (α_i) and continuous cumulative density $\Phi_i(\delta)$, both of which are estimated using the historical data. The departure delay depends on a number of factors such as time of the day, day of the week, week of the month, month of the year, origin and destination airports, carrier, weather, special days and other non-recurring events.

As mentioned in the literature review, there has been extensive research on how to forecast the departure delay for flights. While some of these existing forecasting models are accurate and based on extensive parametric estimation, they do not satisfy two major requirements of our dynamic routing with real-time information. First, these models are static estimation models, thus they do not allow distribution modification as per the real-time information (e.g., announced delay). Second, they allow early departures (i.e., negative delays) which is not practicable in routing applications since the early departures are only possible once all the cargo is loaded (or passengers have boarded). This can only happen if the capacity is full or if the routed cargo is already loaded on the plane. In the former case, the data is not relevant, since we are estimating

delay for flights with available capacity. The latter case, on the other hand, cannot be considered prior to determining the routing policy and its subsequent implementation.

We used the database of the Bureau of Transportation Statistics (BTS) and FAA's Operations Network (OPSNET) in estimating the parameters of the departure delay model. Both of these sources provide detailed on-time departure and departure delay information on all US domestic flights as well as major US airports. According to the FAA's Air Traffic Organization Policy, all air traffic facilities, except flight service stations, must submit delay reports daily to the OPSNET[7]. However, we note that the data in both the BTS and OPSNET are either aggregated at the facility level or available only for the passenger flights. Since our routing model is applicable to both dedicated carriers as well as passenger carriers, we assume that the departure delay for cargo carrying flights can be approximated with delay data reported for the mix passenger/cargo flights. This assumption can be justified by considering the fact that in both cases the flights are affected by similar factors.

These databases provide multi-year historical data on departure delays. These delay data can be extracted for each combination of the determining factors (i.e., origin–destination airports, time of the day, etc.) to estimate the most accurate non-parametric delay distribution for each flight. However, this results in sample size issues because of the numerosity of independent factors. Accordingly, we follow two strategies to overcome the sample size problems. First strategy relates to the on-time departure performance used to estimate the on-time departure probability α. The BTS reports on-time departure data in two categories: departures with zero delay and early departures. In our estimation, we assume that early departures are equivalent to on-time departures in that they contribute to the on-time departure probability α. Hence, we aggregate the early departure to on-time departure (zero delay). The second strategy relates to reducing the number of independent factors in estimating the delay distribution for a given flight. In our analysis, we have selected the origin airport, destination airport, month of the year and time of the day as the factors to be included as determinants of the departure delay distribution. Departures from the same origin airport are subject to the same structural and transient factors such as runway/gate capacity, air traffic flow, weather conditions, daily backups, etc. Similarly, departures to the same destination are subject to similar conditions at the destination airport. Due to yearly and daily seasonality, the departure in the same month and same times are affected by common factors. While the day of the week and carrier are also important factors, we have performed our estimation by aggregating the delay data across all days of the week and carriers due to the sample size concerns. As described below, this selection approach for delay distribution estimation proved to be acceptable.

Departure Delay Data Processing

As mentioned in section 3, the departure delay distribution is estimated based on the historical data. For this problem, the data is extracted from the BTS database. We limited our data to only one month to avoid seasonality and arbitrarily selected November 2007.

To estimate the departure delay distribution, we needed to estimate the percentage of on-time departure (α_i) and distribution of delayed flights $\beta_i(\delta_i)$. It is usually easy to estimate α_i based on

[7] http://www.faa.gov/airports_airtraffic/air_traffic/publications/at_orders/media/ODR.pdf
http://aspmhelp.faa.gov/index.php/Operations_Network_%28OPSNET%29

the ration of occurrence. However, when it comes to estimating $\beta_i(\delta_i)$, since the number of delayed flights departing in a given time window might be limited, data points are often not adequate for sound estimation. One way to overcome this issue is to aggregate the data as mentioned in section 3. That is, we acknowledge the affecting factor on departure delay and aggregate the data under similar factors to establish estimations, and then extend it to all the contributing groups.

Consequently, after aggregating data for statistically non-influential factors, hieratical clustering was used to cluster departure hours based on their average departure delay. The number of clusters is indicated by the data; however, it was realized that in most cases clustering the departure hours to just two clusters was enough. A sample of this procedure is presented in Figure 7. In this chart, the average departure delay for each hour for ORD-SEA flights in November 2007 is presented and the two clusters are distinguished by color.

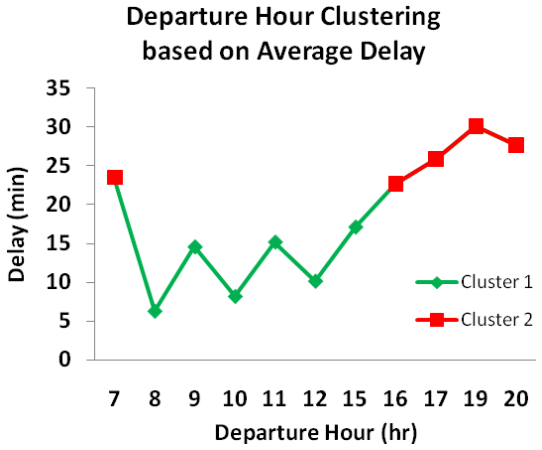


Figure 7. Departure hour clustering based on average departure delay for ORD to SEA

As can be observed in Figure 7, there is a notable difference between the departure delay during the business (8 to 17) and non-business hours. This might rely the fact that the main cause of delay could be because of the excessive demand from passengers to fly in early morning or late evening.

As mentioned, exact forecasting of the departure delay is not in the interest of this paper. However, studying the processed data suggests that the distribution of departure delay for delayed flights is following the exponential distribution. This assumption for the data on hand was confirmed by goodness-of-fit test using the Kolmogorov-Smirnov method.

After estimating the departure delay, we used simulation to generate the needed departure delays for this problem. Moreover, based on the historical data, the upper bound for departure delay (ζ) is selected as 90 minutes. Briefly, departure delay for a given flight in this problem is zero with a probability of α_i or a number between one and ζ based on the exponential distribution of $\beta_i(\delta_i)$ with the probability of $1-\alpha_i$.

Delay Announcement Policy

The information about the expected delay of flights is usually distributed by airports and/or carriers. The policy of when and how to estimate and announce the expected departure delay, varies from carrier to carrier and from airport to airport. In this paper, one of the goals is to analyze the importance of the quality of the real-time information. Accordingly, the problem was solved for different announcement policies.

Announcement of departure delayed is based on simulated delay as mentioned before. In this problem setting, the delay distribution is divided into m segments with equal probabilities. Then, the boundary of the segment that contains the simulated departure delay is announced as $(\eta_i(t), \upsilon_i(t))$ in the beginning of the problem for all flights in all airports. The conducted case study includes different values of m.

5.2 Case Study Flight Network

To demonstrate the performance of the proposed approach in a real world scenario, a sample case study was established. In this case study, the goal was to pickup a load from Cleveland Hopkins International Airport (CLE) and deliver it to Seattle-Tacoma International Airport (SEA). To avoid affecting the result by flight cost, the performance criteria were limited to the delivery time, the sooner the better. Accordingly, a penalty function was developed to penalize late delivery. Although it is possible, early arrival was not penalized in this problem setting.

To reach the destination from the origin, three different paths are possible: flying directly, having one stop at Chicago O'Hare International Airport (ORD), or having one stop at Denver International Airport (DEN). It is assumed that the freight forwarder has contracts with couple of airlines and accordingly there are multiple connecting flights with various departure schedules available between the mentioned airports. The flights and their schedules are presented in Figure 8 and Table 1. The schedules and travel times were extracted from BTS database.

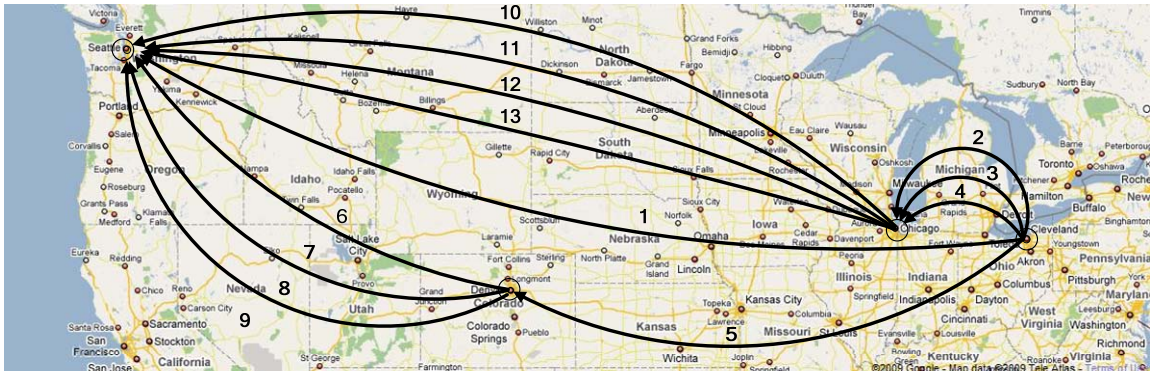


Figure 8. Case Study Air-Cargo Network

Table 1. Flights details

Carrier	*Cont.*	*United*	*American Eagle*	*United*	*Express Jet*	*United*	*South West*	*United*	*United*	*Alaska*	*United*	*AA*	*United*
Label	1	2	3	4	5	6	7	8	9	10	11	12	13
Origin	CLE	CLE	CLE	CLE	CLE	DEN	DEN	DEN	DEN	ORD	ORD	ORD	ORD

Destination	SEA	ORD	ORD	ORD	DEN	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA
Duration(min)	313	90	90	90	205	177	177	177	177	279	279	279	279
Departure *	10:40	8:15	10:15	10:35	9:05	11:20	12:37	14:06	14:15	8:45	9:15	10:55	10:57
On time(%)	43.5	68.8	68.8	68.8	79.6	51.3	51.3	51.3	51.3	46.8	57.6	57.6	57.6
Delay(min)	16.69	18.14	18.14	18.14	11.0	16.19	16.19	32.29	32.29	26.2	10.01	10.01	10.01

* All time are US Easter time

After setting the problem as mentioned, we used Matlab to code the proposed model. The problem then was solved for different departure delay announcement polices (various values of *m*) and simulation was used to generate 20,000 samples for each case. Each of these problems was then solved with the three models (Dynamic, Static and Benchmark) and the results were recorded. In this section, we will first evaluate the performance of the proposed model against other mentioned models and discuss its merits and demerits. Next, we proceed with evaluating the effect of the announcement policy on the performance of the model.

5.3 Results

Performance Comparison

As expected, different approaches gave different performance and although the announcement policy affects the individual case results, we can recognize three general outstanding differences in the outcomes.

First, on average the performance of the dynamic approach is better than static ones and closer to the benchmark as illustrated in Figure 9. This fact indicates that merely by following the dynamic approach when possible, we expected to make more profit (or avoid loss).

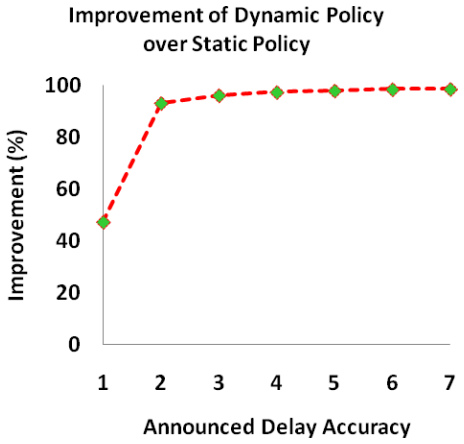


Figure 9. Improvement of Dynamic Policy over Static Policy

This fact can be further investigated by considering the comparative performance measure graph as presented in Figure 9. It can be observed that even without any real time information, the Dynamic approach shows over 47% improvement against the static approach. This improvement rate increased sharply under the availability of real time information. As presented, with *m=2*, we can experience more than 45% improvement. Although the improvement continues after m=2, however, as shown in the experimental study, the improvement rate will decrease with the increment of announced delay accuracy.

Although superior average performance is usually enough for justifying a method, the dynamic approach has other advantages too. Figure 10 presents the distribution of travel time of the samples in the simulation for $m = 2$. It should be noted that although the distribution and values change for various *m*, the pattern is generally the same. As can be seen in the result, since it is possible to freely choose the flights, dynamic policy, unlike the static approach has the chance to be more aggressive when it comes to choosing more risky flights. In other words, the static policy usually goes for the flight with a higher probability of availability when needed; on the other hand, the dynamic policy (when there is a chance) will aim for earlier flights. The latter approach will provide the opportunity for the dynamic method to enjoy the earlier flights when possible. This simply means faster delivery in a competitive way. As can be seen in Figure 10, unlike the static policy that started from 474 minutes, in some cases the dynamic policy enjoyed faster delivery ranging from 384 to 414 minutes.

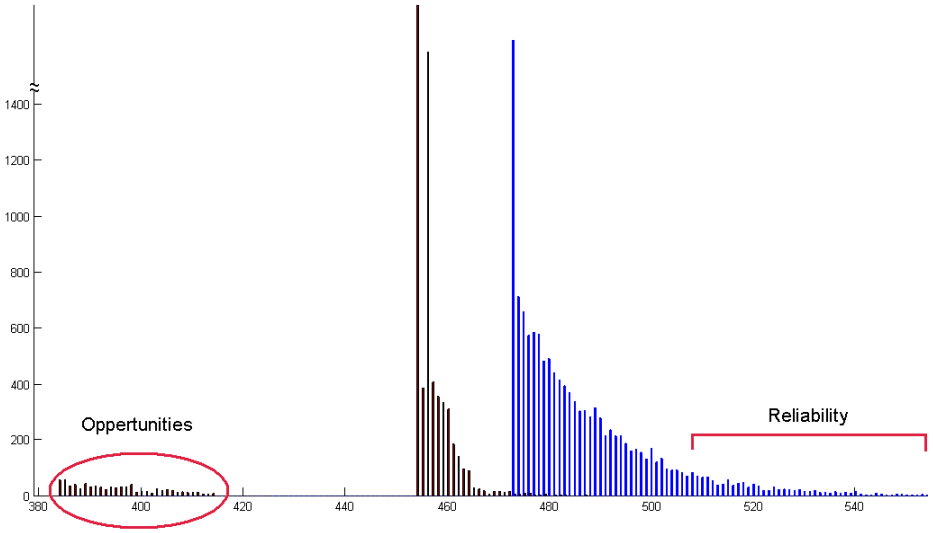


Figure 10. Distribution of travel times for dynamic model (in black) and static model (in gray) for *m*=2.

The last but not the least merit of the dynamic approach over the static policy is the reliability of delivery. As mentioned, one of the key factors in the popularity of the air-cargo system is the increase in demand for the shipment of small, light and expensive cargos in a fast and reliable manner. As presented in Figure 11 and Figure 10, results suggest that the tail of the delivery time distribution is much shorter in the dynamic policy compared to the static one. This can be better realized by comparing the speed of reduction in conditional expected tardiness for *m>1*. In other words, following the dynamic model makes the delivery more reliable by reducing the probability of extreme delays.

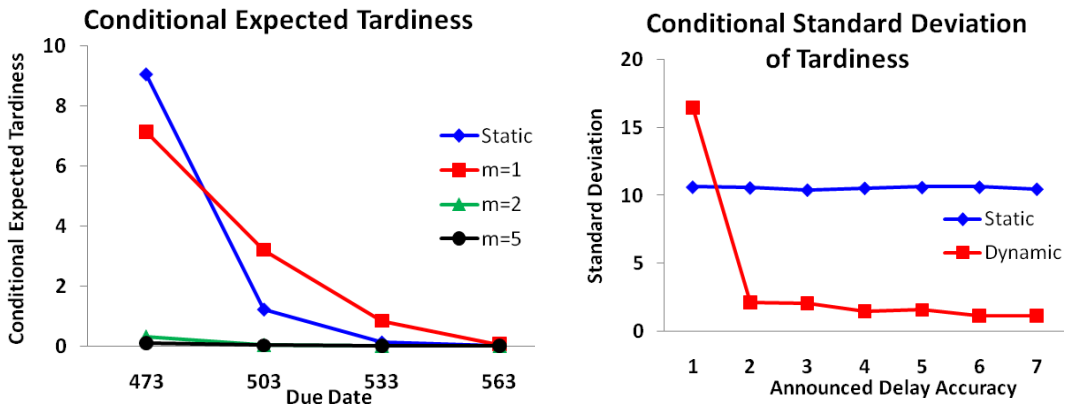


Figure 11. Conditional Expected Tardiness (right) & Conditional Standard Deviation (left),

Sensitivity Analysis on Real-Time Information Accuracy

In this section, we want to evaluate the value of real-time information based on their accuracy. Clearly, when it comes to the announcements, the closer the lower and upper bounds of the announcement are the more accurate the announcement is. That is, bigger m will result in announcements that are more accurate. However, having information that is more accurate is usually costly and practically hard to achieve. Accordingly, it is necessary to be able to estimate the optimal accuracy level that not only provides the reasonable accuracy, but also is practical and economical.

However, the amount of improvement in delivery performance is decreasing as m increased. As a case in point, there is hardly a notable difference between the performances for m=3 and m=30. This fact emphasizes the importance of establishing an optimal level of accuracy for delay announcements to fit the problem.

However, inaccuracy and ambiguity of delay announcements comes with a price for the dynamic model. As mentioned, the dynamic model tries to select the flight more aggressively and this means it will accept more risk in counting on future flights that might not be available when needed. On the other hand, since the static policy does not have a chance of changing the chosen flight which is set, it will go for those that are more reliable. Although in general these facts bring superiority for the dynamic model, inaccurate data can make the dynamic model end up missing the flights that were expected to be available based on the announcements. Since this problem is designed for fast shipment in general, missing a flight is interpreted as a failure of delivery and is thus assigned a huge cost. As a case in point, with m=2 if the actual delay is 80 minutes, the announcement could be (15,90). Clearly, using the delay distribution, expected delay based on the announced boundary is much lower than the real value. This may cause the dynamic model to choose this flight and end up missing it.

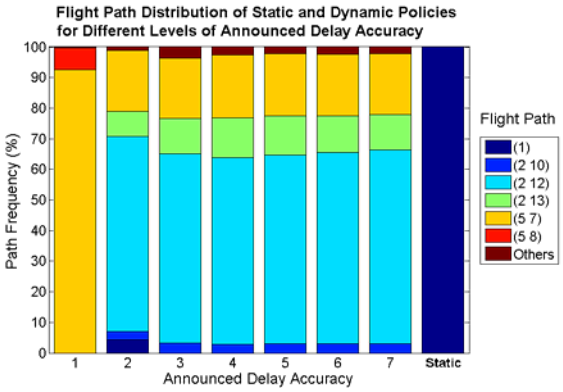


Figure 12. Flight Path Distribution

6 References

Andreatta, G., Romeo, L., 1988. Stochastic shortest paths with recourse, *Networks*, 193–204.

Azaron A., Kianfar, F., 2003. Dynamic shortest path in stochastic dynamic networks: Ship routing problem. *European Journal of Operational Research* 144, 138–156.

Becker, B., Dill, N., 2007. Managing the complexity of air cargo revenue management. *Journal of Revenue and Pricing Management* 6 (3), 175-187.

BTS, 2002. Flight shipments in America. *Bureau of Transportation Statistics.* http://www.bts.gov/programs/freight_transportation/html/freight_and_growth.html

BTS, 2005. Airline Travel Since 9/11, *Issue Briefs* 13. http://www.bts.gov/publications/issue_briefs/number_13/html/figure_01_table.html

BTS, 2006. Flight in America. *Bureau of Transportation Statistics.*

BTS, 2008. Transportation Statistics Annual Report 2007, *Bureau of Transportation Statistics.*

Cheung, R.K., 1998. Iterative methods for dynamic stochastic shortest path problems, *Naval Research Logistics* 45, 769–789.

Deo, N., Pang, C.Y., 1984. Shortest Path Algorithms: Taxonomy and annotation. *Networks* 14, 275-323.

Fu, L., 2001. An adaptive routing algorithm for in-vehicle route guidance systems with real-time information, *Transportation Research Part B* 35(8), 749–765.

Hall, R.W., 1986. The fastest path through a network with random time-dependent travel times, *Transportation Science* 20, 182–188.

Hansen, M., Hsiao, C.Y., 2005. Going South? An Econometric Analysis of US Airline Flight Delays from 2000 to 2004. *Proceedings of TRB 84th Annual Meeting, Washington, D.C.*

Hansen, M., Zhang, Y., 2004. Operational Consequences of Alternative Airport Demand Management Policies: The Case of LaGuardia Airport, *International Journal of Transportation Research Record* 1888, 15-21.

Hoffman, J., 2001. Demand dependence of throughput and delay at New York LaGuardia Airport. *The MITRE Corporation 2001 Technical Paper*.

Kaufman, D. E., Smith, R. L., 1993, Fastest paths in time-dependent networks for intelligent vehicle-highway systems applications. *IVHS Journal* 1(1):1-11.

MECo, 2006. Opportunities within Japan's Consolidating Logistics Industry for 2005, *Morgen, Evan & Company, Inc.*, http://www.morgenevan.com/pages/reports/Japan%20logistics%20Industry%20Report%202005.pdf

Polychronopoulos, G.H., Tsitsiklis, J.N., 1996. Stochastic shortest path problems with recourse, *Networks* 27, 133–143.

Provan, J.S., 2003. A polynomial-time algorithm to find shortest paths with recourse, *Networks* 41(2), 115–125.

Vigneau, W., 2003. Flight Delay Propagation: Synthesis of the Study. *EEC Note No.18/03, Eurocontrol Agency, Brussels, Belgium*.

Waller, S.T., Ziliaskopoulos, A.K., 2002. On the online shortest path problem with limited arc cost dependencies, *Networks* 40(4), 216–227.

Wang, P. T. R., Schaefer, L. A. , Wojcik, L. A. , 2003. Flight Connections and Their Impacts on Delay Propagation. *Proceedings of Digital Avionics Systems Conference, DASC'03*.

Wasserman, L., 2006. All of Nonparametric Statistics. *Springer*.

Wellman, M.P., 1990, Fundamental concepts of qualitative probabilistic networks. *International Journal Artificial Intelligence* 44, 257-303.

Wellman, M.P., Ford, M., Larson, K., 1995, Path planning under time-dependent uncertainty. *In Proceeding of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)*, 532-539.

Wu, Q., Hartley, J., Al-Dabass, D., 2005. Time-dependent stochastic shortest path(s) algorithms for a scheduled transportation network, *International Journal of Simulation Systems, Science & Technology* 6(7-8), 53-60.

Xu, N., 2007. Method for deriving multi factor model for predicting airport delays. *PhD thesis, George Manson University, Fairfax, Virginia, USA*.

B. Integrated Routing on Air-Truck Intermodal System

In this section, we demonstrate the opportunities of the proposed model in an integrated air-truck multi-modal system. The purpose of integrating the air-road system is not just to find the optimum path from the cargo pickup point to the airport, but is also to bring new opportunities for selecting the airport. That is we not only rout the truck to the origin airport optimally but we actually chose the airport and avoid the congested one for less crowded airport. We focused on southeast Michigan and the north Ohio region as our test region. The problem here is to ship a cargo picked up from the aforementioned area, route it to the optimum airport through the road network and finally originating from the selected airport route the cargo through the air network and deliver it the destination airport. Clearly, the problem on hand can be defined as the combination of two dependent sub-problems. The first decision is to choose the best airport. This can only be done by closely investigating the road network and properly estimating the arrival time at each airport in the region. Knowing the arrival time at an airport, we can then realize the expected cost-to-go and policy at the airport calculated via proposed air-routing model.

In the following sections, we first describe the case study configuration in detail and after stating our assumptions, we will explain the approach to solve the integrated air-road multi-model for the case study. Then, the result will be presented and discussed and finally we will finish with the conclusion.

1 Case Study: Intermodal Cargo Routing on the Air-Road Network in OH-MI Region

In the southeast Michigan and north Ohio region, three main commercial airports can be recognized: Detroit Metropolitan Wayne County Airport (DTW), Toledo Express Airport (TOL) and Cleveland-Hopkins International Airport (CLE). We are to select one of these airports to pursue the shipping through the air network after picking up the cargo and routing it to the selected airport. As far as the air network is concerned, we use BTS as our data source to configure the network links (flights) and their attributes. The base configuration is presented in Table 2. We used the actual flights for this case study, however we ignored those flights that fly less frequently than twice a week.

Once again, since cargo carriers, in contrast with passenger carriers, are not obligated to report their performance to BTS, we had access only to mixed passenger-cargo air carriers. Accordingly, we assumed that both cargo flights and mixed flights behave similarly when it comes to departure delay. Moreover, we assumed that the departure delay distribution is exponential. Distribution parameters were estimated from historical data. Departure hours were integrated based on their similarity of hourly average departure delay and then the distribution parameters were estimated from the processed data similar to the method explained in section A.3.

Figure 13 shows the geographical region in which we conducted the case study. To analyze the affect of the decision time over the choice of the airport, we investigated the pickup time window

of 6:00 am until 10:30 am. To evaluate the case study based on travel time only, we defined the due date for shipment delivery at the destination airport (SEA) as 6:00 am and penalized tardiness linearly.

Figure 13. Case Study Geographical Region

Table 2. Case Study Flight Configuration

From	To	Flight Number	Scheduled Departure*	On-time Departure (%)	Average Delay (min)	Flight Time (min)
CLE	ORD	UA675	6:00 AM	76.2	17.79	87
	ORD	MQ4382	6:25 AM	76.2	17.79	87
	ORD	UA226	8:15 AM	76.2	17.79	87
	ORD	MQ4257	10:15 AM	76.2	17.79	87
	ORD	UA477	10:35 AM	76.2	17.79	87
	DEN	XE2467	9:05 AM	86.7	9.00	205
DTW	SEA	NW211	9:30 AM	51.7	11.21	280
	SEA	NW215	12:24 PM	51.7	11.21	280
	ORD	UA199	6:51 AM	84.7	13.44	49
	ORD	NW12357	7:00 AM	84.7	13.44	49
	ORD	AA1615	7:05 AM	84.7	13.44	49
	ORD	UA265	7:55 AM	84.7	13.44	49
	ORD	AA1637	8:35 AM	84.7	13.44	49
	ORD	NW1237	9:02 AM	84.7	13.44	49
	ORD	AA765	10:20 AM	84.7	13.44	49
	ORD	UA793	10:41 AM	84.7	13.44	49
	DEN	F9-628	6:30 AM	88.2	14.00	170
	DEN	F9-622	7:35 AM	88.2	14.00	170
	DEN	UA579	9:06 AM	88.2	14.00	170
	DEN	NW1223	9:13 AM	88.2	14.00	170
TOL	ORD	MQ4312	6:20 AM	93.1	9.00	67
	ORD	MQ4359	10:00 AM	93.1	9.00	67
DEN	SEA	AS537	9:00 AM	51.0	14.76	157
	SEA	WN1235	9:45 AM	51.0	14.76	157
	SEA	F9-847	10:25 AM	51.0	14.76	157
	SEA	UA875	10:27 AM	51.0	14.76	157
	SEA	F9-835	11:55 AM	51.0	14.76	157
	SEA	UA339	13:18 PM	51.0	14.76	157
ORD	SEA	AA1611	8:40 AM	71.1	10.44	251
	SEA	UA755	9:00 AM	71.1	10.44	251
	SEA	AS21	9:05 AM	71.1	10.44	251

	SEA	UA331	10:45 AM	71.1	10.44	251
	SEA	AA643	11:15 AM	71.1	10.44	251
	SEA	UA331	10:57 AM	71.1	10.44	251

*All times are US eastern time

As for the road network, we used the Google Map API (GMA, http://code.google.com/apis/maps/). This platform provides driving directions and estimated travel times and distances for a giver pair of origin-destination. We used this simple system instead of the aforementioned vehicle road routing model we discussed earlier for two reasons. First, the dynamic routing system is based on real-time information that can potentially affect the airport choice because of incident or unusual congestion. Since our goal is mainly to investigate the decision making process over airport selection, we tried to avoid uncontrolled temporary affects. In other words, we mainly desire to demonstrate the expected best airport for different pickup times/locations than the momentarily optimal one. Moreover, the GMA works based on static data and offline which is considered faster than online calculations for expected scenarios as mentioned.

We define the geographical region between latitude [41.2 42.38] and longitude [-81.64 -83.98] degrees with a resolution of 0.02 degree for both latitude and longitude. Then we calculated the expected travel time for each point on the map to the three airports and estimated the arrival time. The expected cost-to-go at each airport can be estimated through the proposed model presented in section A.3. Comparing the expected cost at each airport, the optimum airport would be the one with the cheapest expected cost-to-go and in the case of tie, the closest one based on the estimated GMA distance.

2 Case Study Results and Additional Experiments

Solving the above mentioned problem, we identified the optimum airport choice based on time and location. Figure 14 presents the sample result for the optimum airport choice based on locations at 6:00 am. In this figure, the infeasible points are the regions where no road network exists. As can be observed, Lake Erie and Lake Saint Clair fall in this category. In addition, it seems GMA does not care for boarder crossings. It is well known that travel time from the Canadian region to any of the potential airports includes crossing the boarder and is potentially much longer than what GMA estimated. Moreover, GMA also uses ferry routes over Lake Erie that are impractical to include in our case study. As a result, for evaluating the figure, we shall ignore the Canadian areas and lake routes.

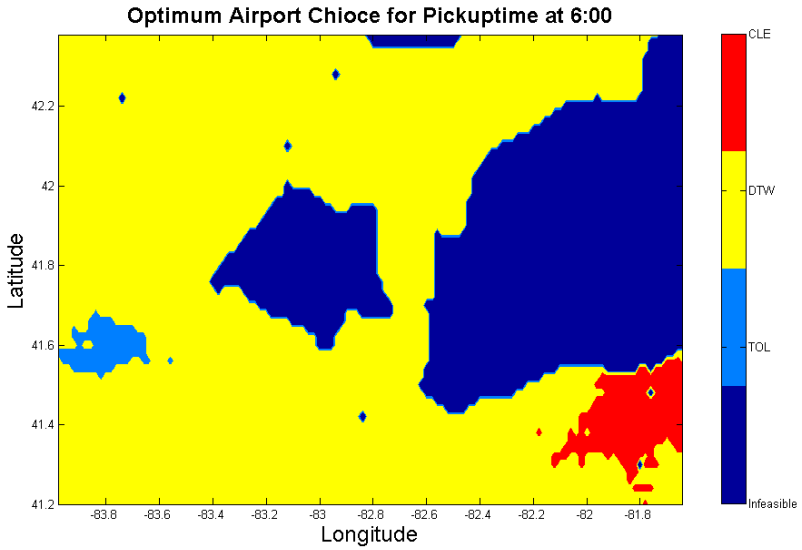


Figure 14. Regional Best Airport for Base Case

It can be clearly perceived that at the investigated time, DTW almost dominates all the other airports and the remaining two can only compete in their closely surrounding areas. This can be better understood by considering Figure 15. In this figure, the expected cost-to-go at different airports for a range of pickup times is presented. As can be seen, DTW provides significantly cheaper costs compared to others. Therefore, until the location in question is far enough away from DTW so that the travel time counters the advantage of DTW over a closer airport, DTW will be the best choice. Consequently, as expected, through the test time window DTW will be the optimum choice for most of the regions. This can be observed in Figure 16. This figure presents the percentage of locations that chose a specific airport. It shows that except for the small timeframe in the early morning in the case of TOL, and some locations between 7 to 9 AM in case of CLE, DTW is the better decision.

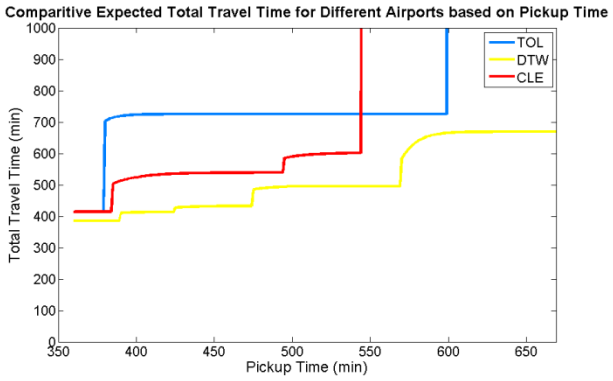


Figure 15. Comparative Expected Cost at Different Airports

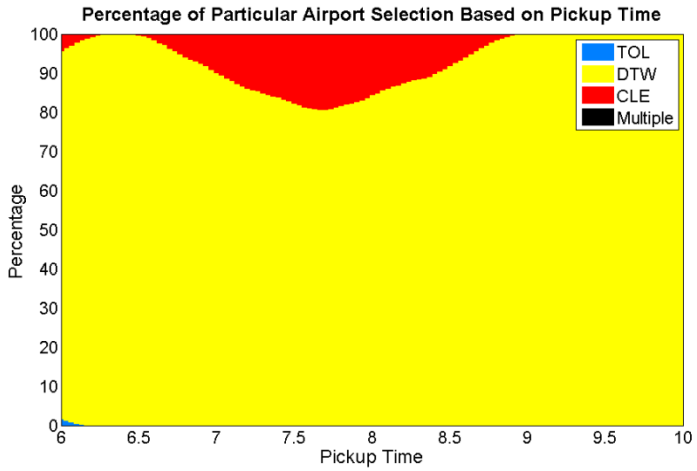


Figure 16. Percentage of Airport Selection

It should be noted that this result is affected by our assumption of merely considering mixed passenger/cargo flights. As a matter of fact, this assumption is highly in favor of DTW and CLE and unfair toward TOL since the first two have a wide variety of flight options where TOL passenger flight data is limited. Actually, to our best knowledge, most of the TOL airport flights are cargo flights. Accordingly, to overcome this issue and do more evaluation, we launched a series of experiments to investigate the effect of potential flight additions on the TOL optimality territory.

For these experiments, we added an extra direct flight to TOL's choice of flights to go directly to SEA. We assumed that the delay distribution of the newly added flight is the same as the passenger flights from TOL to ORD. Moreover, since there was no record of direct flight statistics from TOL to SEA, we estimated the flight time based on the air distance between the two airports compared to the distance and travel time of the direct flights from DTW to SEA. We evaluated the effect of the scheduled departure time on the optimality region. Results of this analysis can be observed in Figure 17 to Figure 20.

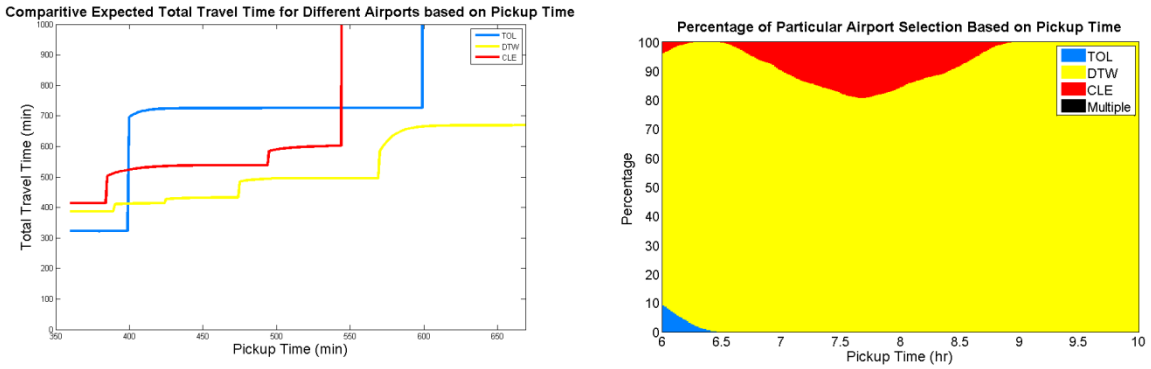


Figure 17. Experiment Result for Departure at 6:40AM

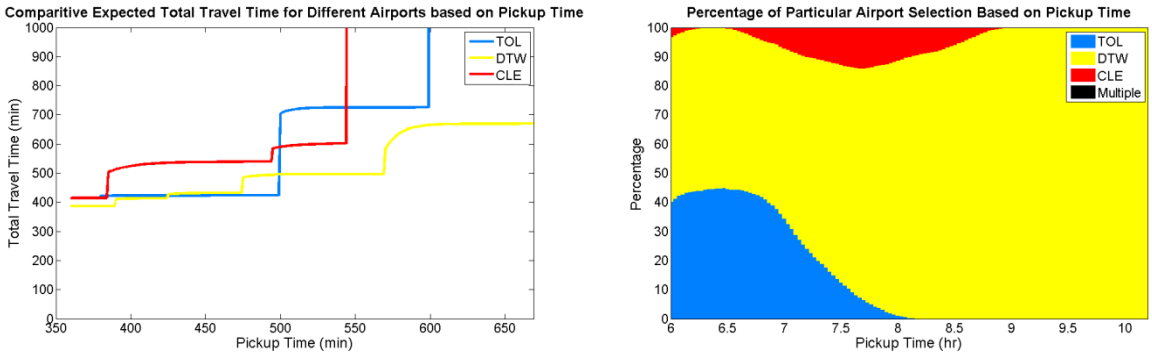


Figure 18. Experiment Result for Departure at 8:20AM

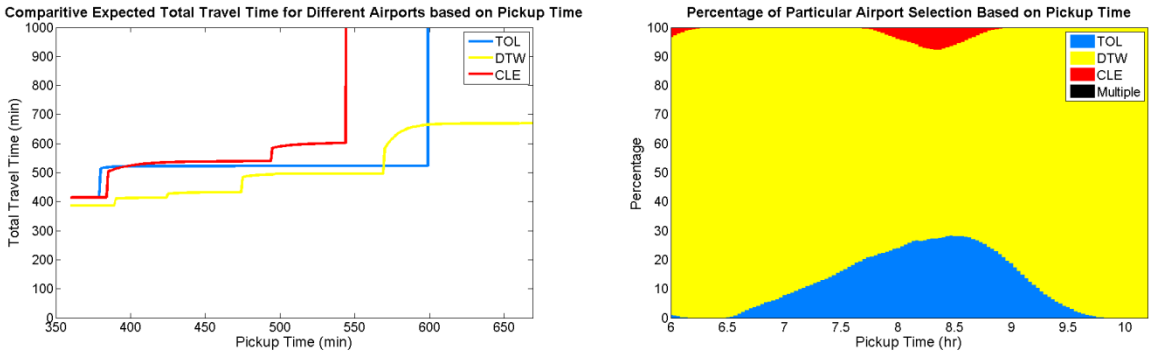


Figure 19. Experiment Result for Departure at 10:00AM

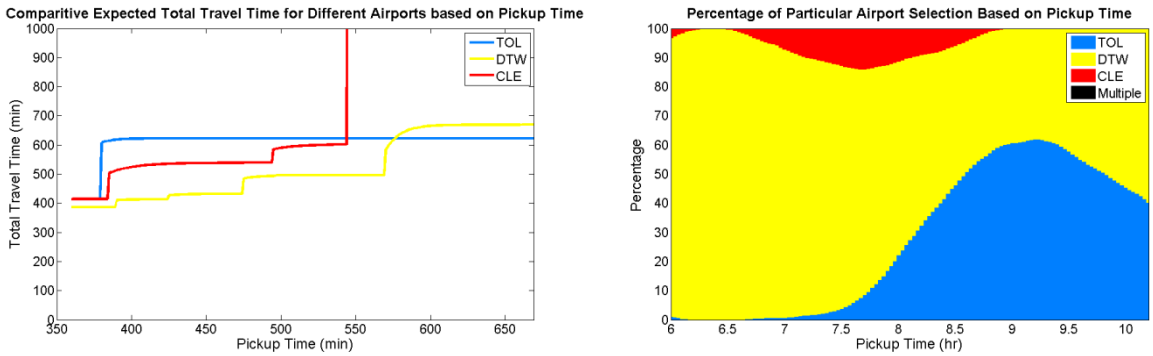


Figure 20. Experiment Result for Departure at 11:40AM

As can be perceived from the results, just adding a single flight can change the situation dramatically in favor of TOL. Figure 17 presents the effect of adding a flight that was scheduled to depart at 6:40 AM. Although the TOL share increased to some extent, it seems the departure time is so early that the effect of the flight addition will not be beneficial since there is not enough time to drive there and catch this flight from many locations.

In Figure 18, the flight in question was scheduled to depart at 8:20 AM. It is obvious that the effect of having this flight is much stronger and long lasting than the previous experiment. This can be explained by considering the time needed to drive to TOL and catch the flight. Clearly, this

latter departure time provided a better opportunity for further to benefit from it. It should be considered that the potential delay even make the expected departure time latter.

Figure 19 shows the lesser advantage for the departure time of 10:00 AM compared to the latter case. Moreover, it seems this flight is more likely to attract the CLE customers than the last scenario. Considering the expected cost graph, it seems TOL's cost is very close to the two other airports and a minor adjustment in scheduled departure time or congestion in road networks can change the situation in one way or another.

Figure 20 suggests that scheduling the flight at 11:40 AM has a significant effect in changing the TOL optimality territory. The effect of this addition can be realized as early as 7:00 AM. The peak effect occurs around 9:00 AM. It can be stated that compared to the previous charts, most of the gain in new optimal territories for TOL are actually coming from DTW.

The above-mentioned interpretations can be better studied by analyzing the geographical optimality map. The series of these maps are provided for 7:30 AM in Figure 21. As can be observed, in the first experiment, TOL does not play a notable role at 7:30AM. In this experiment, the added flight to TOL is scheduled to depart at 6:40AM and considering the road trip time, the effect of the new flight vanishes almost by 6:30 AM (Figure 17). Having the flight at 8:20 AM clearly affects the situation at 7:30 AM since close locations to TOL have the chance to catch the flight, and from the expected cost-to-go at Figure 18 it is known that the TOL cost is less than DTW and CLE even without considering the road trip time. As the flight departure time is postponed to 10:00 AM, a new pattern appears. As shown in Figure 21, the added flight does not affect the airport's surrounding region but attracts the locations that are rather far away from TOL and closer to CLE. This can be attributed to the expected cost-to-go at the airports. That is, because of the comparative expected cost, it is better to drive all the way to TOL and catch the direct flight that eventually has a cheaper cost than go to CLE and catch the 8:15 to ORD or 9:05 to DEN. This clearly highlights the advantage of decision support tools and making a proper decision versus the common sense of going to the closer airport.

Similar phenomenon can be observed when it comes to the 11:40 AM departure time for the added flight at TOL. However, since the departure time is rather late, TOL can attract shipments from locations quite far away. Actually, evaluating the expected cost charts in Figure 20 reveals that in this scenario we will face the situation that other airports flight options are almost exhausted and TOL remains the cheaper option.

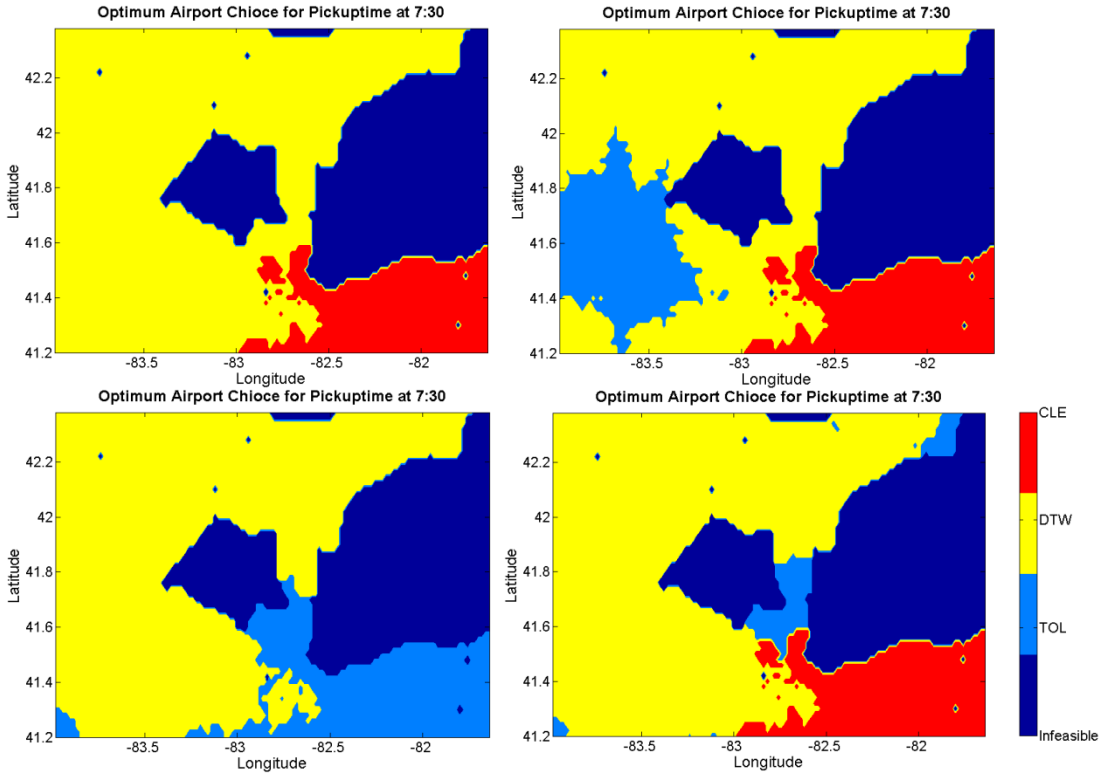


Figure 21. Geographical Airport Choice for Different Experiment 6:40AM (top-left), 8:20AM (top-right), 10:00AM (down-left) and 11:40AM (down-right)

In a nut shell, based on the case study and experiments we can conclude that a freight forwarder can really benefit from a decision support tool that provides him with not only routing on the network but also help him select the network he wants to ship the cargo through. As demonstrated, it is possible to attribute an optimum airport choice to each time-location. By wisely selecting the origin airport, freight forwarders can select the range of options (connecting flights/intermediate airports) that are expected to work more to his favor.

We demonstrated the dynamic nature of airport selection and cargo routing as a result of the airport congestion status. In this study, we only considered the static factors and historical trends. However, as mentioned the proposed model is designed to reroute the cargo based on real time information that includes incidents on the road, severe departure delays, adverse weather conditions and the like. In such situations, dynamic rerouting can dramatically reduce the cost and help the freight forwarder to avoid tardiness penalties.

C. Operational Response Model

1 Introduction

Over the past two decades, supply chains have become more global and competitive in an effort to accommodate dynamically changing consumer preferences and the business environment. The proliferation of product variety, reduced product life-cycles and aggressive outsourcing practices are attributable to these trends in globalization and competition. The effective management of supply chains in such conditions is becoming more and more challenging, primarily, because of the increased supply and demand uncertainty. Furthermore, these conditions lead to supply chain disruptions which differ from regular fluctuations in supply and demand in that they are low-probability and high-impact events (Christopher et al. 2002). Regular fluctuations in supply and demand, such as quality defects, seasonality, and variability in lead-times, are usually well managed by effective inventory management practices. However, these coping mechanisms, such as buffers, are not adequate for effectively managing supply chain disruptions caused by excessive transportation delays and intermodal facility disruptions.

Following the events on September 11, the risk assessment and management of supply chain disruptions has become an important consideration for supply chain planners and managers. Supply chain disruptions can be categorized based on their outcomes (e.g., processes affected) as well as on their sources. Disruptions can affect one or more of the three main supply chain processes, namely supply-side, manufacturing, and demand-side processes. According to their sources, supply chain disruptions can be classified as either man-made or natural disasters. Man-made disruptions are caused by the actions of people. These actions are either intentional such as strikes and terrorist attacks or unintentional such as quality problems. A recent example for intentional man-made actions is the three-month strike at the American Axle plants which slowed or idled as many as 30 General Motors assembly and parts plants with an estimated loss of $2.82 billion in revenue (Automotive news, 27 May 2008). In March 2000, Nokia and Ericsson both faced a shortage of components for a period of six months because of the chip contamination caused by a fire in the Philips plant, in Albuquerque, New Mexico (Tomlin, 2006). In the aftermath of September 11, many international borders were closed forcing production plants to shut-down. For instance, Ford Motor Company had to shut down five of its plants due to parts shortages from its Canadian suppliers. A recent example for the unintentional man-made causes is the leakage problem in the hybrid battery packs for General Motors' three product lines Saturn Vue, Aura and Chevrolet Malibu. This quality problem in the batteries supplied by Cobasys led GM to alter its production plans and recall many vehicles. Examples for natural sources of disruptions are earthquakes, floods and hurricanes. In 2005, two major hurricanes, Katrina and Rita, hit the U.S. Gulf Coast and damaged the oil refining capacity, inventories of lumber and coffee and other fresh produce (Snyder, 2006b). The Taiwan earthquake in September 1999 had caused world-wide supply shortages in the semiconductor market (Kleindorfer, 2005). On 16 July 2007, an earthquake in Japan halted the production at Riken, Japan's largest maker of piston rings and seals, and disrupted Honda's North American manufacturing operations.

The impacts of these supply chain disruptions on the company's performance change according to the vulnerability of the disruptions. Christopher and Peck (2004) and Sheffi and Rice (2005) specified that vulnerability depends on the probability of the occurrence and the severity of the disruptions. Less vulnerability is more likely not to have catastrophic results. Sheffi and Rice (2005) suggested that reducing vulnerability can be achieved by increasing resiliency. Christopher and Peck (2004) define the resilience as "the ability of a system to return to its original state or move to a new, more desirable state after being disturbed." According to Sheffi and Rice (2005), supply chain resiliency is strategic in nature and contributes to a company's competitiveness. They offer two mechanisms to achieve resiliency: increasing flexibility and judiciously installing redundancy. Traditionally redundancy can be achieved by safety stock or working with multiple suppliers. Sheffi and Rice (2005) pointed out that while redundancy is a part of a resilient supply chain, both stocking up and working with multiple suppliers are more costly than establishing a resilient supply chain by installing flexibility.

Flexibility, as a mechanism to cope with disruptions can be in five elements of the supply chain: supply and procurement by working with multiple suppliers or having deep relationships with suppliers, conversion flexibility by having the ability to deal with disruptions in the manufacturing system, distribution and customer-facing by making decisions to respond to customers efficiently right after disruptions, control systems by having the ability to realize the disruptions quickly before they have affected the supply chain, and the right culture by generating a culture that is aware of the disruptions, Sheffi and Rice (2005). Conversion flexibility- manufacturing flexibility, is defined as the ability to cope with changes and uncertainties without anguishing over the loss of performance in the system, Gupta (1996). The most well known manufacturing flexibilities that are studied are machine, process, product, routing, volume, expansion, operation, and production flexibilities. So far the literature has focused on the flexibilities but paid little attention to the product design (e.g., commonality) or the flexibility decision (e.g., manufacturing flexibility). Besides the process and product design flexibility has a huge impact on coping with the uncertainties (disruptions) in supply chains, considering the component is an element needed for assembling end-products. One of the important features of product design flexibility is the substitution in components among end-products. There is no doubt that commonality is an extreme case of substituting components among end-products if there is no cost in substituting. As Balaskrishnan and Geunes (2000) said, substituting gives advantages of cost savings as component commonality does. Also Balakrishnan and Geunes (2000) defined the ability of using the same components among different end-products as an opportunity of bill-of-materials (BOM).

In this paper, we propose an operational response model to cope with the disruptions in the supply side of supply chains. In particular, this operational response model leverages the bill-of-materials flexibility, subject to volume and mix flexibility to find the optimum level of substitution in components during the decision making process in product design to cope with supply-side disruptions. Our work is motivated by a series of shortages experienced by a big automotive company (Ford Motor Company) since they have begun sourcing from far east countries such as China. We consider the application-specific integrated circuits (ASIC) which are increasingly being used as the automobile electronic content proliferates. These chips (ASIC) are being used in many of the vehicle subsystems and have different functionalities. For higher end vehicles, these components are more expensive but they could be substituted for lower end components.

In section 2, we do literature review on disruptions and related types of flexibilities. In section 3 we introduce the algorithm and mathematical model we proposed. In section 4, we illustrate the application of the operational response model through a stylized example.

2 Literature Review

In this section, we first review the earlier studies on supply chain disruption coping mechanisms. Next, we review the key studies on manufacturing flexibility as an important feature of disruption coping mechanisms.

In the literature, supply chain disruptions are mostly classified according to their causes. Kleindorfer and Saad (2005) categorized the risks into two that can affect supply chain management. One is risks that come from normal coordination of supply and demand, regular fluctuations that can happen. Another type of risk comes from natural disasters, terrorist attacks, and strikes etc. Hammant and Braithwaite (2007) categorized risks as external risks and internal risks, then they categorized external risks as supply side risks, demand side risks and environment risks. A similar categorizing has been done by Wagner and Bode (2006); they categorized risks as supply side risks, demand side risks and catastrophic risks. Gaonkar and Viswanadham (2004) separated risks as deviations, disruptions and disasters. Sheffi and Rice (2005) categorized risks as random events, accidents and intentional events. Bringing together all categorizing types supply chain disruptions can be categorized based on their outcomes and on their sources. According to their outcomes, disruptions can be classified by their affects on one or more of the three main supply chain processes, such as supply-side, manufacturing system, and demand-side processes, and according to their sources. Supply chain disruptions can be classified as either man-made or natural disasters. In this paper we focused on the supply-side disruptions and these disruptions can be caused both from man-made or natural disasters.

Many papers have studied coping mechanisms with supply chain disruptions, and not surprisingly the number of these papers increased exponentially after September 11. Pochard (2003) stated there is not just one way to cope with disruptions. The best strategy for companies is should be analyzing different solutions and find the best one for their companies' characteristics. Kleindorfer and Saad, (2005), Chopra and Sodhi (2004), Christopher and Lee (2004) all focused on the importance of the assessing and mitigating the level of risk in a supply chain. Kleindorfer and Saad (2005) graph a conceptual framework on potential loses causes by supply chain disruptions and risk mitigating investments. They studied the ways to mitigate supply chain disruptions and proposed that each supply chain system should have its own strategy because they said each environment has a different culture and management methods. Another proposal to mitigate supply chain disruptions are information sharing, trust and a good coordination among supply chain members. According to Christopher and Lee (2004) the lack of confidence and panic force the stakeholders to make unreasonable supply chain decisions. Chapman et al (2002) counted sources of the disruptions and studied on the impacts of the disruptions on the vulnerability of the supply chain, also Wagner and Bode (2006) did their study on supply chain vulnerability. Their aim was to investigate the relationship of supply chain vulnerability and supply chain risks. According to their survey, which they applied to 760 executives in Germany, companies should be careful of their dependency on customer and suppliers. Bundschuh, Klabjan and Thurston (2003),

focused on reliability and robustness in supply chain systems. Reliability and robustness is defined by them as having a low probability that any supplier fails and that the supplier has the ability to maintain their activity even after disruption. They stated that increasing the number of suppliers makes the system more robust. Sheffi (2001), Kleindorfer and Saad (2005) and Pochard (2003) suggested working with multiple suppliers to cope with supply chain risks. Snyder and Daskin (2005) specified that considering risk diversification, if companies face disruptions the number of suppliers increases. Gaonkar and Viswanadham (2004) established an empirical framework which includes questions on selecting suppliers to decrease loses that can be caused by disruptions. Tomlin (2004) studied the supplier selection under the case of the reliability of suppliers are unknown. Snyder et al. (2006) studied both the cost of constructing a network and the costs that results from disruptions. Two other important techniques that have been used in the literature to cope with disruptions are holding inventory or excess capacity (Chopra and Sodhi 2004). Christopher and Peck (2004) and Sheffi and Rice (2005) suggest resiliency as a coping mechanism to supply chain disruptions and they commented that resiliency can be achieved by redundancy and increasing flexibility. According the expensiveness of the holding extra inventory, augmented capacity and redundancy in supplier, Sheffi and Rice (2005) study another strong technique which is flexibility. Flexibility increases the resiliency of supply chain systems against potential disruptions (Sheffi and Rice, 2005), as well as against the regular variability which occurs now and again. In this paper we study related flexibilities, both at the process development stage and product design stage, to develop more resilient supply chain systems against supply disruptions.

As an important feature of coping mechanisms to disruptions and regular variability, Bertrand (2003) defined flexibility as "the ability to change or react with little penalty in time, effort, cost or performance". Boyer (1996) stated manufacturing flexibility is a very important item which responds to the changes in demand and can be used as a solution to cope with problems that are caused from uncertainties. Also Beamon (1999) expressed flexibility is used to measure the ability of a supply chain system's adaptability to the uncertain environment and he stated that flexibility is a critical item for supply chain systems to be successful in an uncertain environment. Vokurka (2000) specified manufacturing flexibility functions as a critical competition component in providing an advantage in the marketplace to manufacturers. Unfortunately, Slack (1995) stated that in their survey, managers do not understand the real meaning of flexibility, and manufacturers are confused among various flexibilities. One of the important points is to understand which flexibility in a manufacturing system is needed because each plant has a different environment, culture and manufacturing process so each system needs different types of flexibility . Another problem for manufacturers is the measurement of flexibility. Koste and Malhotra (1999) indicated that the ability to measure flexibility is one of the first steps in understanding and improving manufacturing capability. They suggested three dimensions to measure flexibility: Range, Uniformity and Mobility. Bertrand (2003) also stated that the different ways of being flexible are range, uniformity and mobility. The most known types of manufacturing flexibilities suggested to the manufacturers are expansion flexibility, modification flexibility, new product flexibility, volume flexibility, and product mix flexibility (Koste and Malhotra, 1999, Chandra et al., 2005). Also many kinds of flexibilities and definitions and classification have been studied in literature. These can be found in literature review articles such as in Sethi and Sethi (1990), Toni and Tonchia (1998), Beach et al. (2000), Vokurka R.J., and O'Leary-Kelly (2000), Bengtsson (2001). Bertrand (2003) indicated three dimensions in flexibility: volume, mix and new product flexibility. The flexibilities that we study in this paper are mix flexibility that provides a system ability of

producing multiple products and volume flexibility that provides a system to vary its capacity with less cost. Mix flexibility and volume has been famously studied in the literature, Fine and Freund (1990) study the trade-off between flexibility and the acquirement cost of different types of capacity. In detail, they study a firm that has n-products and has to decide on the optimal mix of flexible and dedicated capacities under uncertain demand environments. Jordan and Graves (1995) determine that designing partially flexibility well provides similar benefits to total flexibility. Van Mieghem (1998) studies the trade-off between flexible and dedicated resources while firms have two-products under continuous demand. In each case he finds that flexibility is more beneficial than dedicated resources, Tomlin (2006) suggests product mix flexibility as an alternate coping mechanism to the supply uncertainty. Tomlin and Wang (2005) analyze the profits of mix flexibility and also supply diversification in the case of multi-product circumstance. Kurtoglu (2004) studies the flexibility of two assembly lines by changing the system to produce new products. Suarez, et. al. (1995) examine the relationships between market uncertainties and manufacturing flexibility. Chandra, et. al. (2005) study the relationships between flexibilities such as volume, mix flexibility and new product flexibility.

One other important feature that provides a different form of flexibility and has an impact on coping with disruptions is the product design decisions that are made during the product design process. Hence, to cope with disruptions, companies have to consider their product design system. One of the product design decisions related to the flexibility to cope with disruptions is the level of commonality among components used in the products. In the commonality concepts, firms can use the same component in more than one product without paying more costs. Gerchak (1986), Baker (1986), Gerchak (1988) and Tsubone (1994) study the impact of the commonality on demand uncertainty. Gerchak (1988) stated that using common parts provides less inventory stock than by not using common parts among products. However firms should be more careful when deciding on the number of common parts in products, because commonality decreases the variability of products. Chandra, et. al. (2005) has studied a multi-product manufacturing system to which they applied different levels of flexibility and part commonality in the system to find the effects of the flexibility to the manufacturing performance. They find that increasing product-mix flexibility and component commonality increases the profitability of the system. In general, we can impart that commonality is the specific case of substitutability. Component commonality or compatibility of the substitutions between products and flexibilities in manufacturing and substitution save companies from having to respond to demand in the case of disruptions. Balakrishnan and Geunes (2000) defined the ability of using the same components among different end-products opportunistically as bill-of-materials (BOM) flexibility and they propose a single-level lot-sizing model when considering BOM flexibility. They give examples of their practice from the computer industry where modular architecture is the standard, the continuous processing industry (aluminum tubes) or any product where downgrading is possible (high-strength alloy for low-strength, or high-grade integrated circuit for low-grade).

In this paper, we find the strategic level of substitution among components used in products during the product design process through mix flexibility, volume flexibility and bill-of-material flexibility in order to increase resiliency of the supply chain system against supply disruptions. This paper differentiates form Chandra et al (2005) by considering the commonality of components as parts used in products, not as using the same resources, considering bill-of-material flexibility and generalizing models when considering supply disruptions. This differentiates from

Balakrishnan and Geunes (2000) by including process flexibilities (mix flexibility and volume flexibility), considering supply disruptions and studying substitution between components under stochastic supply conditions.

3 Operational Response Model

We first provide the notation used in the model.

Notation:

r_j : revenue from product j

$c^t_{i,i'}$: cost of substitution component i instead of i' in period t

q_{ij} : usage of component i in product j

d^t_i : number of allocated component i in period t

C^t_1 : capacity of one shift plant in period C^t_1

s^t_i : supply of component i in period t

β_j : product mix flexibility coefficient of product j

α_j : product $j's$ fractional production mix

M : a large number

θ_j : capacity usage rate of product j

e : multiplicative coefficient for overtime/undertime capacity

Sets:

T : number of periods

n : number of products

m : number of components

Indices:

j : product

i : component

t : period

Decision Variables:

x^t_j : number of product j produced in period t

$z^t_{i,i'}$: number of component i substituted with i' in period t

y^t_i : number component i delivered from period t to period $t+1$

u^t : binary variable indicating whether plant is operational in period t

v^t : binary variable indicating whether producing one shift or two shift in period t

Assumptions:

- all production is sold at every period;
- zero vehicle inventories,

- no penalty associated with failing to meet the demand,
- components are relatively inexpensive thus their inventory costs are ignored,
- Supplying chips is uncertain
- for each period supply is changes according to the discrete probability estimates.

The deterministic model formulation is as follows:

$$\textit{Maximize} \qquad \sum_{t=1}^{T}\sum_{j=1}^{n} r_j x_{j,t} - \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{\substack{i'=1 \\ i'\neq i}}^{m} c_{i,i',t} z_{i,i',t}$$

subject to

$$\sum_{j=1}^{n} a_{ij} x_{j,t} \leq d_{i,t} \qquad \forall i,t$$

$$y_{i,t} = y_{i,t-1} + s_{i,t} - d_{i,t} - \sum_{i'\neq i}^{n} z_{i,i',t} + \sum_{i'\neq i}^{n} z_{i',i,t} \qquad \forall i,t$$

$$D_{j,t} \leq x_{j,t} \leq C_{j,t} \qquad \forall i,t$$

$$x_{j,t} \in 0 \cup \mathcal{Z}^{+} \qquad \forall i,t$$

$$d_{i,t}, y_{i,t}, z_{i,i',t} \geq 0 \qquad \forall i,i',t$$

where,

$D_{j,t}$: Demand for model j in period t

$C_{j,t}$: Production capacity for model j in period t

Note that we are assuming that there is no demand and production capacity interaction between models, i.e. they are produced at dedicated facilities and demand for one model is not substitutable with the other one. Thus, resulting constraints are merely simple upper and lower bounds. It is also possible to model the demand substitutability as well as shared production capacity.

3.1 Stochastic Formulation

Stochastic formulation is obtained with two-stage stochastic programming formulation. First stage is the set of time periods where the supply is known with certainty, i.e. $t = 1,\ldots,T_s$. Second stage is the set of periods where supply is uncertain, i.e. $t = T_s + 1,...,T$. Let's assume that we could estimate supply amount in each period with a discrete probability distribution for each chip type in the second stage. For this let's define the following set of definitions

Define:

$s_{i,t}^k$: supply alternative k for chip i in period t for $t = T_s + 1 \ldots T$

$\mathcal{A}_{i,t}$: set of supply alternatives for chip i in period t for $t = T_s + 1 \ldots T$.

$p_{i,t}^k$: probability of supply alternative k for chip i in period $t = T_s + 1 \ldots T$

$$\left(\sum_{\forall k \in \mathcal{A}_{i,t}} p_{i,t}^k = 1 \right)$$

$\mathcal{A}_t$: set of all possible supply scenarios in period **t** ($\mathcal{A}_t = \mathcal{A}_{1,t} \times \mathcal{A}_{2,t} \ldots \times \mathcal{A}_{m,t}$)

$\mathcal{A}$: set of all possible supply scenarios in the **second stage** ($\mathcal{A}_t = \mathcal{A}_{T_s+1} \times \mathcal{A}_{T_s+2} \ldots \times \mathcal{A}_T$)

p^a : probability of supply scenario a

$$\left(p^a = \prod_{t=T_s+1}^{T} \prod_{i=1}^{m} p_{i,t}^k \right)$$

Using above definitions we could formulate deterministic equivalent of the two-stage stochastic program. Let's define the stochastic variables as follows:

$x_{j,t}^a$: production level of model j in period t under supply scenario a

$d_{i,t}^a$: availability of chip type i in period t under supply scenario a [8]

$y_{i,t}^a$: inventory of chip type i in period t under supply scenario a

$z_{i,i',t}^a$: quantity of chip type i used to substitute i' in period t under supply scenario a

[8]Previous definition of $d_{i,t}$ ($d_{i,t}^a$) regard it as the usage level in model production. However this is only true if the corresponding resource constraint is binding, which is often not the case due to the discrete nature of production variables $x_{j,t}$ ($x_{j,t}^a$) .

Maximize

$$\sum_{t=1}^{T_s}\sum_{j=1}^{n} r_j x_{j,t} - \sum_{t=1}^{T_s}\sum_{i=1}^{m}\sum_{\substack{i'=1\\ i'\neq i}}^{m} c_{i,i',t} z_{i,i',t} + \sum_{\forall a\in\mathcal{A}} p^a \left(\sum_{t=T_s+1}^{T}\sum_{j=1}^{n} r_j x^a_{j,t} - \sum_{t=T_s+1}^{T}\sum_{i=1}^{m}\sum_{\substack{i'=1\\ i'\neq i}}^{m} c_{i,i',t} z^a_{i,i',t} \right)$$

subject to

$$\sum_{j=1}^{n} a_{ij} x_{j,t} \leq d_{i,t} \qquad \forall i,t = 1,\ldots,T_s$$

$$y_{i,t} = y_{i,t-1} + s_{i,t} - d_{i,t} - \sum_{i'\neq i}^{n} z_{i,i',t} + \sum_{i'\neq i}^{n} z_{i',i,t} \qquad \forall i,t = 1,\ldots,T_s$$

$$D_{j,t} \leq x_{j,t} \leq C_{j,t} \qquad \forall i,t = 1,\ldots,T_s$$

$$\sum_{j=1}^{n} a_{ij} x^a_{j,t} \leq d^a_{i,t} \qquad \forall i,a \in \mathcal{A}, t = T_s+1,\ldots T$$

$$y^a_{i,T_s+1} = y_{i,T_s} + s^a_{i,T_s+1} - d^a_{i,T_s+1} - \sum_{i'\neq i}^{n} z^a_{i,i',T_s+1} + \sum_{i'\neq i}^{n} z^a_{i',i,T_s+1} \qquad \forall a \in \mathcal{A}, i$$

$$y^a_{i,t} = y^a_{i,t-1} + s^a_{i,t} - d^a_{i,t} - \sum_{i'\neq i}^{n} z^a_{i,i',t} + \sum_{i'\neq i}^{n} z^a_{i',i,t} \qquad \forall a \in \mathcal{A}, i, t = T_s+2,\ldots,T$$

$$D_{j,t} \leq x^a_{j,t} \leq C_{j,t} \qquad \forall a \in \mathcal{A}, i, t = T_s+2,\ldots,T$$

$$x_{j,t}, x^a_{j,t} \in 0 \cup \mathcal{Z}^+ \qquad \forall a \in \mathcal{A}, i, t$$

$$d_{i,t}, y_{i,t}, z_{i,i',t}, d^a_{i,t}, y^a_{i,t}, z^a_{i,i',t} \geq 0 \qquad \forall a \in \mathcal{A}, i, i', t$$

Constraints (1),(2),(3) and (4) represent the deterministic equivalent of the supply scenarios. Constraint (2) is expressed with deterministic previous period inventory (y_{i,T_s}) since we enter the second-stage at $T_s + 1$. To exemplify an instance of the problem, consider the case where we have five deterministic periods in the first stage ($T_s = 5$) and a single stochastic period in the second-stage ($T = 6$). In addition consider, we have $n = 4$ type of chips and $m = 2$ models. If each chip has 3 supply alternatives ($|\mathcal{A}_{i,t}| = 3$ for $\forall i,t = 6$), then in total we will have 3^4 scenarios in the second-stage ($|\mathcal{A}| = 81$). Thus the problem size increases exponentially, with the number of chips and time periods in second-stage. However, it increases polynomially with the number of supply alternatives in period.

3.2 Additional Constraints:

Plant-Model Mixture

Plants are set up to run certain fixed mixtures (α^p_j) of models and options. Assume that there is a certain flexibility associated with these mixtures, i.e., product mix is allowed to vary within

$[(1-\beta_j^p)\alpha_j^p, (1+\beta_j^p)\alpha_j^p]$ for model j at plant p. Whenever the mix falls outside these limits, plant is shutdown and production is deferred for one week. Let's define the parameters as follows:

α_j^p : Model $j's$ fractional production mix(default) at plant $p.(\sum_{j=1}^{n} \alpha_j^p = 1)$

β^p : product mix flexibility coefficient of plant p.

β_j^p : product mix flexibility coefficient of model j at plant p ($\sum_{j=1}^{n} \pm \beta_j^p = \beta^p$)

To give an example, let's say that an ideal product mix for a plant p is $(\alpha_1^p, \alpha_2^p) = (0.40, 0.60)$. It is also estimated that the plant's efficiency and resource constraints will not be affected if there is $\beta^p = 0.2$ variability in these mixes. in addition, let's assume that $\beta_{j=1}^p = \beta_{j=2}^p = 0.1$. Therefore allowable mixes lie within $(0.36, 0.64)$ for Model 1 and $(0.44, 0.56)$ Model 2 respectively. In order to model the plant shutdown decisions and relations with the product mix, we need the following additional decision variables and parameters:

M : A large number for constraint enforcement

$x_{j,t}^p$: Production level of model j in period t at plant p

u_t^p : Binary variable indicating whether plant p is operational in period t, i.e., $u_t^p = 1$ plant is in production and $u_t^p = 0$ plant is shut down in period t.

We first express the product mix constraints using the above definitions. The upper limit for each model mix is given in constraint (product mix upper bound-1). Similarly lower limit for each model mix is given in constraint set (product mix lower bound-1). Note that when $u_t^p = 1$, both of the constraints (product mix lower bound-1) and (product mix upper bound-1) are in effect. Accordingly, they are redundant whenever $u_t^p = 0$. Constraint (product mix plant shutdown constraint-1) ensures that there is no production whenever the plant is shut down.

$$x_{j,t}^p \le (1+\beta_j^p)\alpha_j^p \sum_{l=1}^{n} x_{l,t}^p + M(1-u_t^p) \qquad \forall j,t,p$$

$$x_{j,t}^p \le (1+\beta_j^p)\alpha_j^p \sum_{l=1}^{n} x_{l,t}^p + M(1-u_t^p) \qquad \forall j,t,p$$

$$x_{j,t}^p \le M u_t^p \qquad \forall j,t,p$$

These constraints will affect chip substitution decisions such that whenever the substitution decision leads to a plant shutdown, additional profit from substitution could be forgone to prevent loss of revenue as a result of this shutdown.

Discrete-Capacity

Plants are allowed to operate only at certain discrete capacity levels. On a given week plants are set up for either one, two or three shifts of production. Also, it is not possible to change this arrangement on a daily basis as the production is planned on a weekly schedule. In other words, the supply chain is set up in such a way that decisions regarding production must be made at least one week ahead. Assembly lines only run at certain fixed speeds thus throughput can only be adjusted with production time. If a plant is running on a two shifts schedule, one can choose the length of the shifts (eight, nine, ten hours) without significant disruption of the operating pattern.

We ignore the additional cost of overtime and formulate the constraints of discrete capacity. Let's define the following parameters.

e^p : multiplicative coefficient for overtime/undertime capacity at plant p .
$C^p_{t,1}, C^p_{t,2}$: Production capacity for plant p in period t using 1 and 2 shifts, respectively.
θ^p_j : Capacity usage rate of model j at plant p

Let's define the binary decision variable for choosing number of production shifts at each plant and time period (assuming only existence of 1 and 2 shift options, 3 shift option can be modeled with an additional binary variable).

v^p_t : Binary variable indicating whether plant p is running on a single shift in period t , i.e., $v^p_t = 1$ plant is producing for a single shift and $v^p_t = 0$ production is running in two shifts.

In a single shift option we have the upper and lower bounding constraints (1 shift upper bound-1) and (1 shift lower bound-1). Observe that when $u^p_t = 0,$ plant is shut down in period t , then both of these constraints become redundant and the production for all models at plant p is forced to be null due to constraint (product mix plant shutdown constraint-1).

$$\sum_{l=1}^{n} \theta^p_l x^p_{l,t} \le (1 + e^p) C^p_{t,1} + M(1 - v^p_t) + M(1 - u^p_t) \qquad \forall t, p$$

$$\sum_{l=1}^{n} \theta^p_l x^p_{l,t} \ge (1 - e^p) C^p_{t,1} - M(1 - v^p_t) - M(1 - u^p_t) \qquad \forall t, p$$

In double shift option we have the upper and lower bounding constraints (2 shift upper bound-1) and (2 shift lower bound-1).

$$\sum_{l=1}^{n} \theta_l^p x_{l,t}^p \le (1+e^p) C_{t,2}^p + M v_t^p + M(1-u_t^p) \qquad \forall t,p$$

$$\sum_{l=1}^{n} \theta_l^p x_{l,t}^p \le (1+e^p) C_{t,2}^p + M v_t^p + M(1-u_t^p) \qquad \forall t,p$$

With the additional constraints, the revised deterministic formulation would be as follows

$$\textit{Maximize} \qquad \sum_{p=1}^{P}\sum_{t=1}^{T}\sum_{j=1}^{n} r_j x_{j,t}^p - \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{\substack{i'=1 \\ i' \ne i}}^{m} c_{i,i',t} z_{i,i',t}$$

subject to

$$\sum_{p=1}^{P}\sum_{j=1}^{n} a_{ij} x_{j,t}^p \le d_{i,t} \qquad \forall i,t$$

$$y_{i,t} = y_{i,t-1} + s_{i,t} - d_{i,t} - \sum_{i' \ne i}^{n} z_{i,i',t} + \sum_{i' \ne i}^{n} z_{i',i,t} \qquad \forall i,t$$

$$x_{j,t}^p \le (1+\beta_j^p)\alpha_j^p \sum_{l=1}^{n} x_{l,t}^p + M(1-u_t^p) \qquad \forall j,t,p$$

$$x_{j,t}^p \ge (1-\beta_j^p)\alpha_j^p \sum_{l=1}^{n} x_{l,t}^p - M(1-u_t^p) \qquad \forall j,t,p$$

$$x_{j,t}^p \le M u_t^p \qquad \forall j,t,p$$

$$\sum_{l=1}^{n} \theta_l^p x_{l,t}^p \le (1+e^p) C_{t,1}^p + M(1-v_t^p) + M(1-u_t^p) \qquad \forall t,p$$

$$\sum_{l=1}^{n} \theta_l^p x_{l,t}^p \ge (1-e^p) C_{t,1}^p - M(1-v_t^p) - M(1-u_t^p) \qquad \forall t,p$$

$$\sum_{l=1}^{n} \theta_l^p x_{l,t}^p \le (1+e^p) C_{t,2}^p + M v_t^p + M(1-u_t^p) \qquad \forall t,p$$

$$\sum_{l=1}^{n} \theta_l^p x_{l,t}^p \ge (1-e^p) C_{t,2}^p - M v_t^p - M(1-u_t^p) \qquad \forall t,p$$

$$u_t^p, v_t^p \in \{0,1\} \qquad \forall i,t,p$$

$$x_{j,t}^p \in 0 \cup \mathcal{Z}^+ \qquad \forall i,t,p$$

$$d_{i,t}, y_{i,t}, z_{i,i',t} \ge 0 \qquad \forall i,i',t$$

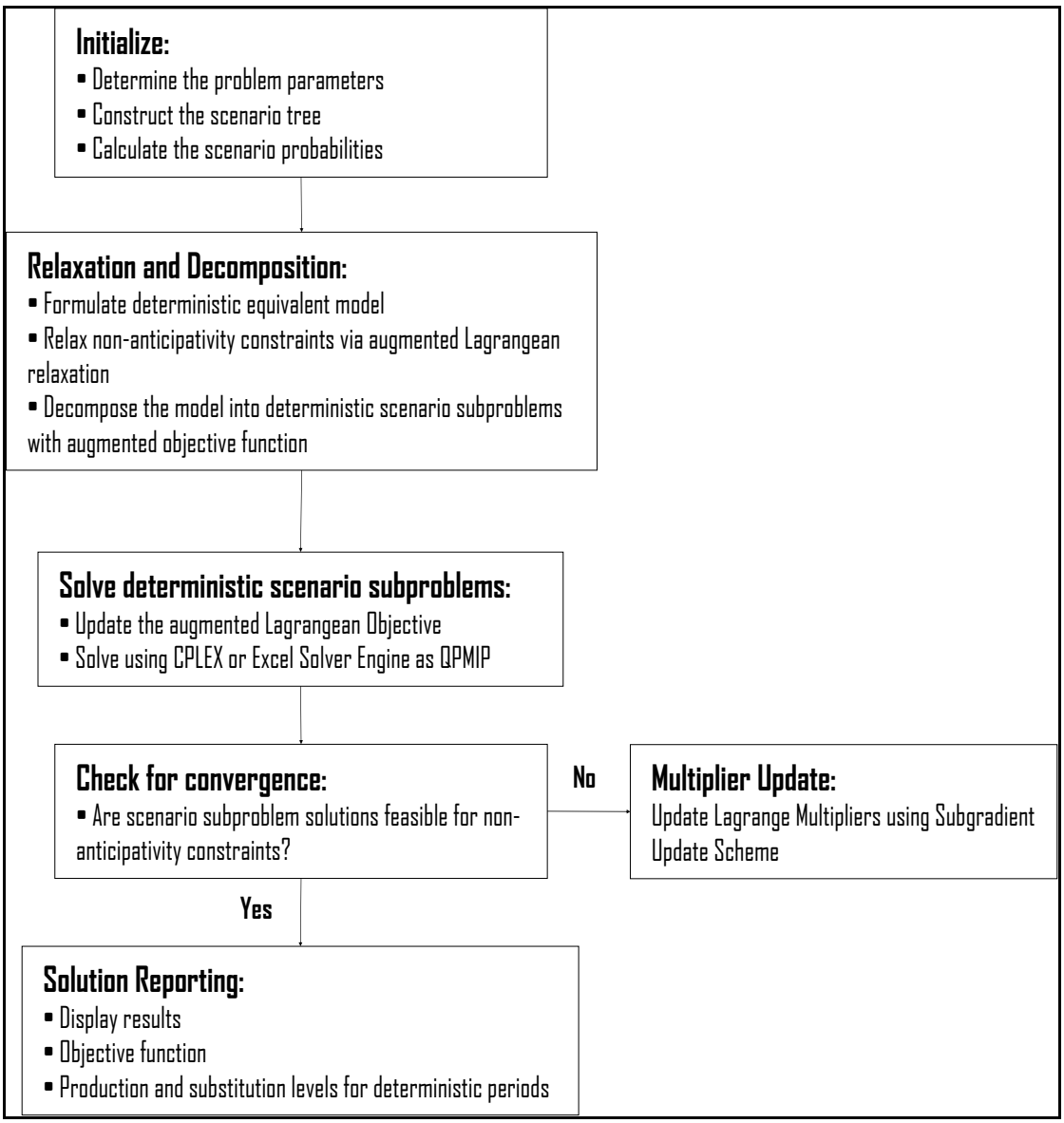
4 Solution Approach - Progressive Hedging Algorithm

The stochastic formulation is a multi-stage stochastic program with recourse. In each stage, we first make such decisions as production, allocation/substitution, shut-down, shifts. In the next stage, we observe the realization of stochastic chip supply and make further decisions. Therefore the decisions in the next stage depend upon realization of chip supply. Associated with the stochastic formulation is the non-anticipativity constraints which states that we cannot anticipate the future and, equivalently, chip supply scenarios with a common history must have the same set of decisions.

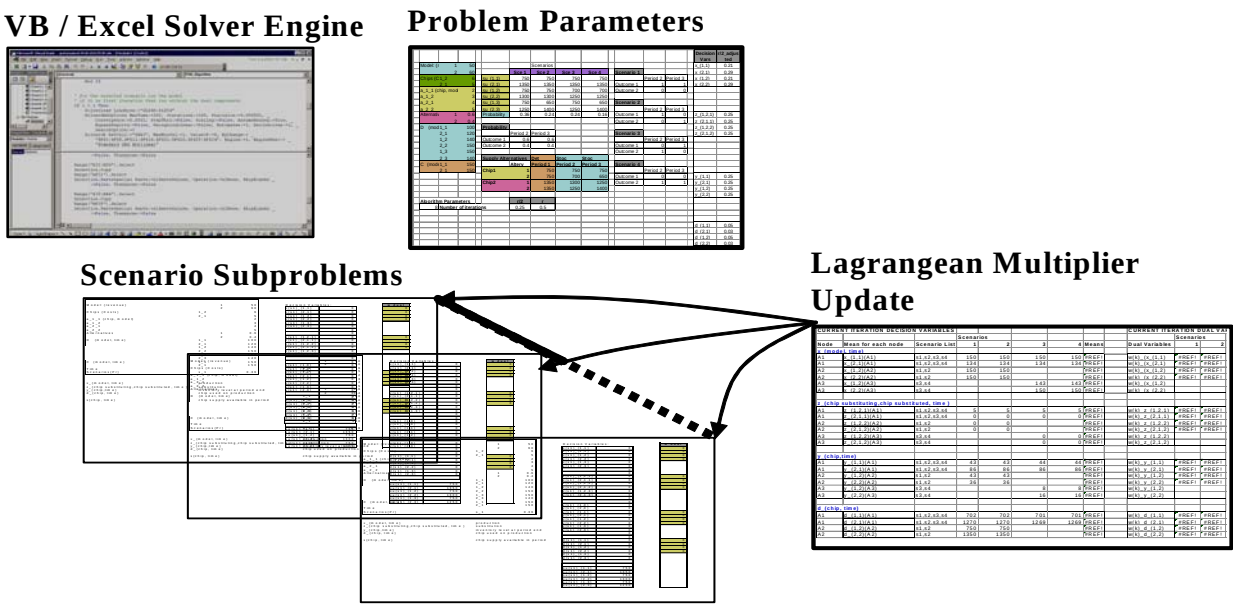
The stochastic formulation presented in the previous section is the stochastic programming model formulation obtained by first developing deterministic equivalent via split variables. In this process, the extended form of deterministic models takes into account the probability distributions. The split variables are merely copies of each variable for each scenario and enforce non-anticipativity. We define the scenario as an instance of realization of stochastic chip supply over the planning horizon. Note that, with the split variable technique, the size of the DE model grows exponentially. For example, assuming a constant number of realizations in each stage/period, the scenario tree grows exponentially. If we consider 2 realizations for each of 6 chip family over 8 weeks then the total number of scenarios is $2^{48} \approx 3x10^{14}$.

To cope with the size of this problem, we propose following the Progressive Hedging Method (PHA). This method is in fact a scenario-based (dual) decomposition where the problem is decomposed into deterministic sub-problems for each scenario. The objective of the sub-problem is an augmented Lagrangian objective for penalizing any lack in the ability to implement -----. The PHA method enforces the implementation feasibility (non-anticipativity) constraints algorithmically and each sub-problem is nonlinear in the sense that there are linear and quadratic terms in the objective. As a result these problems are Quadratic Mixed Integer Programming (QMIP) sub-problems. PHA has the advantage of global convergence in convex problems, and scenario sub-problems are easier to solve. On the other hand, PHA has linear convergence and is locally convergent for non-convex problems thus obtaining an implementable solution requires convergence or a heuristic step.

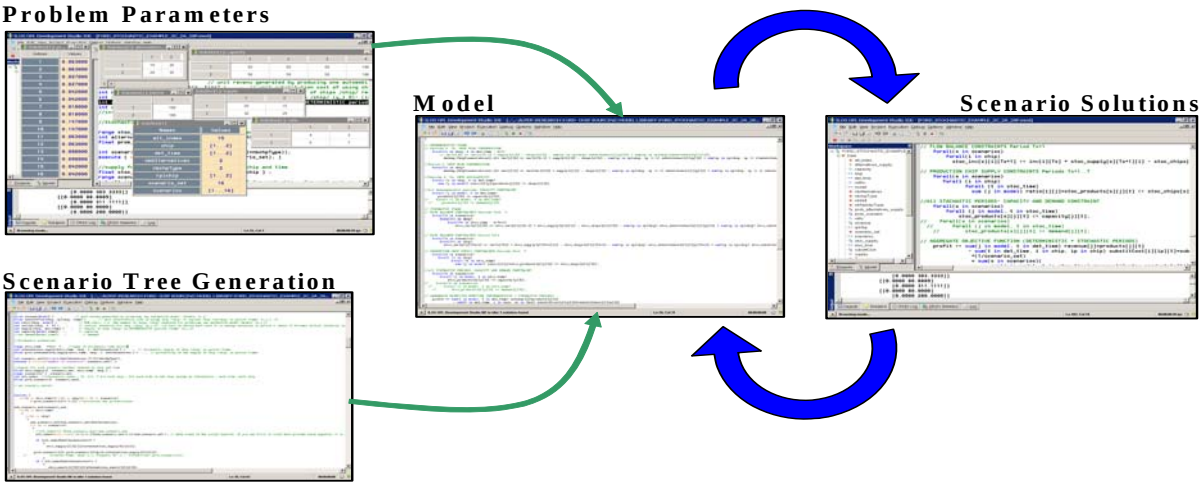
The PHA algorithmic implementation framework is illustrated in the following Figure.



We have implemented the PHA in two platforms. The Excel VBA based implementation is efficient for small problem instances (e.g., <100 scenarios).



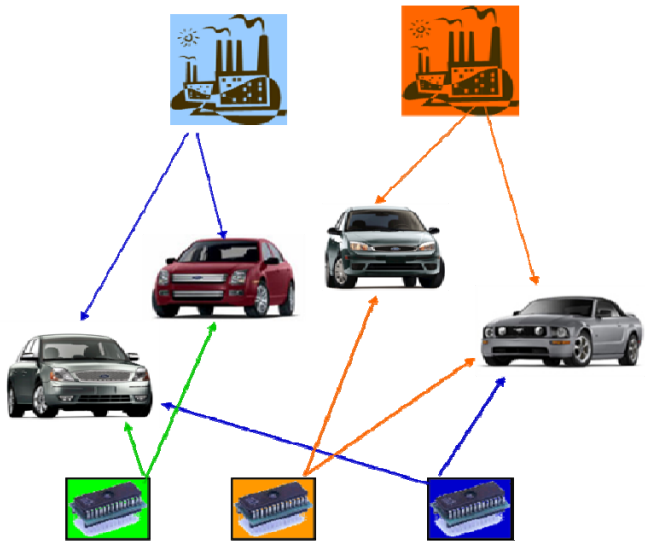
The second implementation platform is the ILOG OPL where we use Cplex 11's Barrier Method for solving QMIP scenario sub-problems. This implementation has proved to be efficient for medium problem instances (e.g., <100,000 scenarios)



Stylized Example

To illustrate the application of the operational response model, we have developed a stylized example. In this example we have 2 plants, 4 vehicles with 2 vehicles assignment for each plant, 3 chips subject to disruption. This example is illustrated in the Figure below.

The green and orange chips are regular chips going into two middle-class vehicles and the blue chip is the premium chips going into two higher end vehicles.



In our disruption scenario, the planning time horizon is 6 weeks long and has 3 periods. The first period is the deterministic periods where we know the supply quantities deterministically. The remaining two periods have stochastic supply. The supplier estimates that the chip supply of these 2 periods (4 weeks) will be affected such that the worst case has 60% and the best case has 40% chance. The outcomes of all three chips are listed in the following table.

	Period #2		Period #3	
	Worst	Best	Worst	Best
Chip 1	0	500	0	400
Chip 2	0	500	100	200
Chip 3	100	300	0	500

We are interested in the decision we have to make for the current period (for 2 weeks from now). Future decisions will also be identified in the form of a policy which tells us exactly what to do as a result of realized supply scenarios. We do not need to update the policy unless the realized scenario differs from what we knew in the preceding period for the future. Let's assume chips arrive in the beginning of period 2 and we received a supply of chip 1 =200. Since this is not one of our outcomes for chip #1, we then have to resolve our problem for period 2 as well as period 3.

The margins for the products are listed in the following table:

		VEHICLES			
					
		$ 2,500	$ 2,000	$ 1,000	$ 1,500
CHIPS		6	10	-	-
		8	-	6	-
		-	2	-	6

The substitution costs and constraints are presented in the following table.

		CHIPS		
				
CHIPS		-	$30	$30
		n/a	-	$25
		n/a	$25	-

The solution for this problem instance is obtained from the PHA algorithm as such that, in the first period (weeks 1 and 2), we operate Plant #1 at 85 per cent utilization and idle the Plant #2 and use about 45% of the available supply of chip #2 inventory to substitute for chip #3. In terms of vehicle production, we allocate 25 per cent of capacity of Plant #1 to produce the silver vehicle (Ford Five Hundred) and 75 per cent of capacity of Plant #1 to produce the red vehicle (Ford Fusion).

5 References

1. Balakrishnan, A., Geunes, J., 2000. Requirements planning with substitutions: exploiting bill-of-materials flexibility in production planning. Manufacturing and Service Operations Management 2 (2), 166–185.

2. Beamon B. M, "Measuring supply chain performance". International Journal of Operations & Production Management. Bradford: 1999. Vol. 19, Iss. 3; pg. 275).

3. Bertrand, J.W.M (2003), Supply Chain Design: Flexibility Considerations. A. G. de Kok and S. C. Graves, Eds., Handbooks in OR and MS, Vol. 11

4. Bundschuh, M., Klabjan, D. and Thurston, D. L. Modeling Robust and Reliable Supply Chain. University of Illinois at Urbana-Champaign June 2003.

5. Chandra, C., Everson, M., and Grabis, J. (2005). Evaluation of enterprise-level benefits of manufacturing flexibility, The Intl. J. of Manufacturing Science, 33, 17-31.

6. Christopher, M. and Peck, H. Building the Resilient Supply Chain. Cranfield School of Management, Volume 15, Number 2 2004.

7. Chryssolouris, G. (1996), Flexibility and Its Measurement. Annals of the CIRP vol. 45/2/1996.

8. Gaonkar, R., Viswanadham, N., A conceptual and analytical Framework for the Management of Risk in Supply Chains. International Conference on Robotics and Automation New Orleans, LA, April 2004.

9. Gupta, D., and Buzacott, J. A. (1996). "A "goodness test" for operational measures of manufacturing flexibility." International Journal of Flexible Manufacturing Systems, 8(3), 233-245.

10. Jordan W. C., and Graves S. C., 'Principles on The Benefits of Manufacturing Process Flexibility'. Management Science; Apr 1995; 41, 4; ABI/ INFORM Global Pg. 577

11. Kleindorfer, P. R. and Saad G. H., Managing Disruptions Risks in Supply Chains. Production and Operations Management. Vol.00, No:0, Month 2005, pp.000-000.

12. Kurtoglu, A. (2004), Flexibility Analysis of Two Assembly Lines. Robotics and Computer-Integrated Manufacturing 20 (2004) 247-253.

13. Nagarur N., Azeem A., 'Impact of Commonality and Flexibility on Manufacturing Performance: A simulation Study'. Int. J. Production Economics 60-61 (1999) 125-134

14. Norrman A., and Jansson U., “Ericsson’s Proactive Supply Chain Risk Management Approach after a Serious Sub-Supplier Accident,” International Journal of Physical Distribution & Logistics Management 34, no. 5 (2004): 434-456.

15. Slack, N. (2005), The Flexibility of Manufacturing Systems. International Journal Of Operations and production Management; 2005; 25, 12; ABI/INFORM Global pg.1190

16. Snyder, L(a)., ‘ Supply Chain Management under the Threat of Disruptions’. U.S. Frontiers of Engineering Symposium, Dearborn, MI, September 23, 2006.

17. Snyder, L.(b), Scaparra, M. P., Daskin, M. S., Church, R. L., Planning for Disruptions in Supply Chain Networks. Tutorials in Operations research, INFORMS 2006.

18. Tomlin B., On the Value of Mitigation and Contingency Strategies for Managing Supply Chain Disruption Risks. Management Science, Vol. 52 No. 5, May 2006, pp.639-657.

19. Tsubone, H., Matsuura H., Satoh S.,(2001), Component part commonality and process flexibility effects on manufacturing performance, International Journal of Production Research; Oct94, Vol. 32 Issue 10, p2479, 15p.

20. Pochard, S., Managing Risks of Supply-Chain Disruptions: Dual Sourcing as a Real Option. Massachusetts Institute of Technology, 2003.

21. Vokurka R.J., and O’Leary-Kelly S. W. ‘A review of empirical research on manufacturing flexibility’. Journal of Operations Management 18 (2000) 485 – 501.

22. Wagner S. M., and Bode C.,’ An Empirical Investigation into Supply Chain Vulnerability’. Journal of Purchasing and Supply Management 12 (2006) 301 – 312.

23. Wu T., Blackhurst J., and O’Grady P. ‘Methodology For Supply Chain Disruption Analysis’. International Journal of Production Research, 45:7, 1665 - 1682.

24. Automotive News, June 2, GM: Batter glitch cuts hybrid sales.

D. Results Dissemination

1. Conference Activity

Conference Presentations:

1. Murat, A., Chinnam, R.B., and Azadian, F., "Dynamic Routing under ATIS for Congestion Avoidance," *Research Issues in Freight Transportation -- Congestion and System Performance Conference,* Seattle (Oct 22-23, 2007).

2. Azadian, F., Murat, A., and Chinnam, R.B., "Enabling Congestion Avoidance in Stochastic Transportation Networks under ATIS," *2nd Annual National Urban Freight Conference*, Long Beach, CA (December 5-7, 2007).

3. Murat, A. , "Managing Supply Disruptions," OEM Global Supply Chain, *SAE World Congress* in Detroit (April 14-17, 2008).

4. Azadian, F., Murat, A., and Chinnam, R.B., "Dynamic Freight Routing on Air- Network Using Real-Time Congestion Information," *INFORMS Annual Meeting*, Washington, D.C. (October 12-15, 2008).

Conference Sessions Organized:

1. We have organized a special session titled "Urban Transportation Planning Models: Dynamic Routing with Real-time ITS Information" at the *INFORMS 2007 Annual Meeting* in Seattle (Nov 3-7, 2007) under the Cluster: Transportation Science & Logistics. The session is Chaired by our PI – Dr. Alper Murat.

2. We have organized a special session titled "Dynamic Routing and Logistics under Real-Time ITS Information" at the *INFORMS 2008 Annual Meeting* in Washington DC (Oct 12-15, 2008) under the Cluster: Real-Time Systems. The session is Chaired by our PI – Dr. Alper Murat.

Conferences Planning to Attend:

1. Murat, A., Azadian, F., and Chinnam, R.B., "Dynamic Freight Routing on Air-Road Intermodal Network using Real-Time Congestion Information" - *POMS Annual Conference 2009 - Global Challenges and Opportunities*, Orlando, 1-4 May 2009.

2. Journal Publications

A journal manuscript is being dispatched this week to the *Transportation Research Part B: Methodological* journal that reports our dynamic routing of air cargo models, algorithms and their performance (Section A of this report). A second manuscript based on dynamic routing on the Air-

Road network is currently under preparation (Section B of this report) and will be submitted for review in this semester. Similarly, a third manuscript based on the operational response model and algorithm (Section C of this report) will be submitted within this semester.